



BrickStor User Guide

Version 21

RackTop Systems, Inc.

RackTop Systems, Inc. provides this document “as is” without representation or warranty of any kind, express or implied, including without limitation any warranty concerning the accuracy, adequacy, or completeness of such information contained herein. RackTop Systems, Inc. does not assume responsibility for the use or inability to use the product as a result of providing this information.

Copyright ©2019 RackTop Systems, Inc. All Rights Reserved. BrickStor is a registered trademark of RackTop Systems, Inc.

Release Date: 19 November 2019

Table of Contents

- Purpose..... 7
- Introduction to BrickStor 8
 - Basic Components of BrickStor 8
 - Physical Components of BrickStor 8
 - Logical Components of BrickStor 9
 - Adaptive Replacement Cache (ARC)..... 10
 - Read Cache..... 10
 - Write Cache..... 10
 - Data Protection Schemes 11
 - Resilvering 11
 - Pool Hierarchy and Containers 12
 - BP (Boot Pool)..... 12
- Initial Out of the Box Configuration 13
 - Default Accounts 13
 - Default Passwords 13
 - RMM (Remote Terminal) IP Address..... 13
 - Configure node name 14
 - Configure Administrative Network Interface Admin0 14
 - Configure Storage Network Interface Data0..... 14
 - Configure Aggregate over network interfaces 14
 - Configure High Availability Heartbeat Interface hb0..... 14
 - Configure Default Gateway 14
 - NTP Setup..... 14
 - DNS Setup 14
 - Hosts Entries 14
 - Local Key Management Configuration 15
 - Configure Time Zone 15
 - System Information and Administration (SIA) 15
 - Check Active Directory under SIA..... 16
 - Joining Active Directory under SIA 16
 - Setup Fault Email Notifications under SIA 16

- Syslog Receiver under SIA..... 16
- Additional Command Line Configurations 16
 - Adding and removing e-mail addresses from Report Notification List..... 16
- myRack Manager 17
 - General GUI Layout and Conventions..... 17
 - Dataset Creation and Manipulation 20
 - Tab Navigation 21
 - Appliance Level Menu Tabs 21
 - General..... 21
 - Sharing..... 22
 - Data Protection 22
 - Encryption..... 28
 - Metrics 30
 - Audit..... 31
 - Network 32
 - System 32
 - High Availability (HA)..... 34
 - Appliance Level Links 38
 - Rack View..... 38
 - Compliance Reports 38
- Pool Level Only Menu Tabs 39
 - Pool 40
 - Sharing..... 41
 - Settings 41
 - Enabling User Behavior..... 41
- Pool Level Only Links 42
 - Pool Performance..... 42
- Pool and Dataset Level Menu Tabs 43
 - General..... 43
 - User Behavior 43
 - Overview..... 43
 - Viewing the User Behavior Audit..... 43
 - Forwarding User Behavior 46

- Sharing 47
- Permissions 50
- Auto Snapshot Data Protection..... 51
- Settings 53
- Storage Utilization 54
- Pool and Dataset Links..... 55
 - Snapshots 55
- Rack View 57
 - Accessing Rack View 58
 - The Rack View Interface 58
 - Self-Encrypting Drive Management 59
- Other Self Encrypting Drive Operations 60
- Exporting and Backing Up Keys 61
- Cryptographically Erasing SEDs..... 61
- SED Protection on the Main Pane 62
 - Creating a Pool within the Rack View 62
 - Modifying an Existing Pool 63
 - Scanning and Repairing a Pool..... 72
- High Availability (HA) Cluster Setup and Management 74
 - HA Cluster Architecture 75
 - HA Scenarios 76
 - Loss of Network Connectivity 76
 - Normal Status Checking: 76
 - Loss of Network Connectivity..... 77
 - Initiating a Manual Failover 77
 - Automatic Failovers 77
 - Performing Maintenance..... 77
 - Witness Configuration..... 77
 - Windows Witness 77
 - Linux Witness..... 78
 - Finishing the HA Cluster Setup..... 79
- Data Protection Best Practices 80
- Encryption Best Practices..... 81

High Availability (HA) Best Practices.....	82
Command Line Operations.....	82
Configuring Ethernet Address on Physical Interfaces	82
VLAN Tagging.....	84
Configuring Default Gateway	86
BSRAPID Configuration	86
Time Zone Setup	87
NTP Setup.....	87
Preparing to Setup and Sync Time.....	87
Hosts Entries	87
Setting up hosts entries	87
RMM (Remote Terminal) IP Address.....	88
Creating Local Accounts	88
Add Local Accounts to Bsradmins Group.....	89
Adding and removing e-mail addresses from Notification List	89
Joining Active Directory	89
iSCSI Share Configuration	91
Creating a Default Target and Target Portal Group.....	91
Configuration & Performance Implications	92
RAID Performance	92
RAIDZ.....	92
Performance of RAIDZ	92
Performance of Mirrors	92
Compression.....	93
Deduplication	93
Clones.....	93
Imbalance of vdev Capacity	94
Performance Monitoring.....	94
Default System Service Ports and Protocols	96

Purpose

The information provided in this document is intended for anyone who wants to configure or administer the BrickStor. It is written for individuals familiar with network attached storage terminology.

If you find incorrect information within this manual, please email info@racktopsystems.com with the subject "Documentation Errata".

Introduction to BrickStor

BrickStor is a CyberConverged™ network attached storage (NAS) solution that fuses scalable capacity and performance with advanced data security and compliance capabilities. BrickStor eliminates attack vectors present in traditional storage systems while automatically ensuring continuous compliance through storage-based data profiles.

Basic Components of BrickStor

The topics that follow describe some of the basic components of BrickStor:

- Physical Components of BrickStor
- Logical Components of BrickStor

Physical Components of BrickStor

Controller / Head / Node

A controller is sometimes referred to as the head or the compute part of a BrickStor system. It contains the operating system and is the gateway interface into managing your BrickStor, as well as the part of the system exposing storage services, providing content management, security, auditing, etc. A typical controller is equipped with multi-core Intel CPUs and 256GB or larger memory. This memory is used for caching discussed in greater detail later in this document. Networking is provided via onboard interfaces with typical system containing two 10GbE Ethernet interfaces onboard, and two or more 10GbE or faster Ethernet interfaces as add-on components for data access. Component redundancy is provided wherever possible, including power, cooling, storage used by operating system, etc. While controllers are field serviceable, a lot of effort is dedicated to eliminating the need for this service in the first place.

Storage Enclosure / Disk Shelf

A storage enclosure, often referred to as a shelf is at its essence a box with redundant components, which just like the controller is engineered to be fault-tolerant and keep functioning in various degraded states. A JBOD is either fully or partially populated with mechanical and/or solid-state drives. A typical configuration is what we refer to as a *'hybrid storage'* system, concept discussed in some detail later in this document. These drives are the primary storage for your BrickStor and organized into logical groupings referred to as pools, another concept discussed in more detail later in this document. Special cache and write optimized *'journal'* devices are frequently also installed in this enclosure. Their purpose is discussed later in the document. A typical configuration consists of at least a single JBOD and a single controller, with some number of drives in the JBOD. There are high availability options available in addition to this 'basic standard' configuration. High availability is a configuration which includes two controllers and one or more JBODs with shared access between these controllers. The basic premise is high availability to some degree protects from catastrophic physical failure, or failure in operating system on a controller. Because storage is common between the controllers, high availability configuration is not meant to provide increased protection for storage, instead storage is protected through mirroring or a parity scheme such as RAID. This is discussed later in the document.

Enclosures are attached to controller(s) via dual SAS host controllers, and utilize SAS drives, which permit dual pathing throughout the system. This is another feature which adds to redundancy of the system. Loss of path to storage may cause a pause, while system recovers from the loss and continues operating with a single remaining path. Whenever possible, RackTop recommends having dual pathing throughout. Diagrams provided at installation time have necessary detail about recommended configuration.

Drives

While in some instances special purpose drives used for caching or journaling are installed in the controller, in a typical configuration mechanical and solid-state drives are installed in the enclosure. Both types of drives use SAS interface, which possesses dual-ported capability and enables dual pathing as described in the last section. Enterprise grade drives are a standard feature in all systems and selected to fit a specific configuration both in terms of capacity and parity scheme or mirroring.

Logical Components of BrickStor

BrickStorOS

BrickStorOS is the Operating System running on your BrickStor appliance. It is not a general-purpose operating system, on the contrary it is for all intents and purposes part of an embedded system, which in combination with RackTop computer hardware becomes a BrickStor storage appliance.

Like most appliances, there is a console mode, and there is shell access, restricted as well as unrestricted, but these exist for supporting very low-level functionality such as configuration of certain things, optimization, troubleshooting and other diagnostic functions. Take caution when attempting to perform actions within the OS that are not documented or recommended by RackTop as it may result in system instability, loss of data and violation of the terms of the system's maintenance contract.

VDEV

A *'vdev'* is a virtual device which can be a single disk, two or more disks that are mirrored, or a group of disks with a parity scheme such as RAID-5. The idea of a vdev is something that abstracts away some unit of storage, which may or may not have any redundancy. One can think of a vdev as a building block in pools, a concept that we address next. Usually, when you hear this term from someone of RackTop it is used to mean a group of disks, and could usually be replaced with word stripe, which will have roughly same meaning in terms of how BrickStorOS implements redundancy in the storage.

Hybrid Pool

A *'hybrid pool'* is the name for a collection of drives, optionally with dedicated read-optimized cache devices and/or write optimized journal devices. All pools are hybrid pools because they are a combination of in-memory read cache as well as actual high capacity persistent storage and optionally read and write cache devices. The high capacity data drives are organized in virtual devices frequently referred to as vdevs. Pools are groups of virtual devices usually with some data protection scheme, such as RAID or mirroring, on top of which filesystems and raw block devices are provisioned. A typical hybrid pool is what RackTop refers to as a hybrid pool is a mix of mechanical drives and solid-state drives. In such a pool data is redundantly stored on large capacity, slower, typically mechanical devices, arranged

into a parity scheme that satisfies data protection as well as capacity and IOPS requirements, while high bandwidth, low latency solid state drives are used for the purposes of caching to accelerate reads and for the purposes of handling synchronous writes, enabling a much better cost to performance ratio over traditional purely mechanical, or purely solid state configurations. RackTop also configures all flash pools which continue to leverage RAM for cache solid state disks instead of mechanical disks to provide consistently lower latency and higher IOPS.

One or more data pools must exist on a system in order to present storage to consumers via AFP, NFS, SMB, etc. While there is no hard limit on number of pools a system could have, usually fewer than 4 pools are configured on any given system. Under normal circumstances the burden of designing and configuring pools is not on the customer, but in the instances where a system is no longer satisfying previously prescribed requirements, RackTop strongly recommends that customer contacts support before any changes are made to configuration of any pool.

From a systems administrator's point of view a pool is a logical organization of independent drives and contains all information about the devices comprising it, structure, filesystems, raw volumes, replication target if any, etc., encoded within its metadata, which makes it possible to easily migrate pools between systems. Critically, this property means that loss of the controller does not in any way compromise data. A replacement controller is all that's necessary to return to normal operations. This feature also enables RackTop's high availability product, which moves pools as well as related network configuration between nodes in the cluster.

Adaptive Replacement Cache (ARC)

The 'ARC' is a portion of memory in the controller dedicated to caching recently accessed data. The ARC caches both recently written data, with assumption that this data may be read soon after being written as well as recently read data, with assumption that this data is potentially going to be read again. Depending on popularity of data it may remain in the cache for a long time, or be evicted in favor of other data, based on criteria which both the user as well as system can optimize for.

Read Cache

Optional SSD Cache device that can be used to extend the amount of data that is cached for Reads. When data is evicted from the ARC it will potentially move to the L2ARC (based up on user configuration settings). Data read from L2ARC will be moved back into ARC.

Write Cache

RackTop uses a journal methodology for its write cache and is implemented in most systems as a mirrored SSDs. A journal is both a software concept and a core physical component, a write ahead log that is used to reduce latency on storage when synchronous writes are issued by clients. RackTop frequently refers to journal as a ZIL, an intent log or a log device. In synchronous write cases, writes are committed to this journal and periodically pushed to primary storage. Journal guarantees that data is protected from loss on power failure due to being in cache before cache is flushed to stable storage.

A log device and is normally only ever written to and never read from. A log device i.e. journal is present to protect the system from unexpected interruptions, such as power loss, a system crash, loss of storage

connectivity, etc. In rare instances where due to power loss or other catastrophe, recovery is necessary, journal is read from in order to recreate a consistent state of the pool, which may require rolling back some transactions, but results in restoring pool to a consistent state, unlike traditional storage systems where only best effort is promised. RackTop recommends mirroring journal devices as a means of preventing loss of journal device, which has performance and potential availability impact. All pools configured at the factory prior to system shipping, the journal, if present, will be mirrored.

Data Protection Schemes

BrickStor is not a traditional RAID system and should not be compared to one. Unlike a traditional system where a RAID controller is a piece of hardware with severely restricted processing power and caching abilities, specifically designed to support one of a number of possible RAID schemes, a BrickStor implements this in software and benefits from the full power and capability of a purpose-engineered operating system, massively powerful processors and huge cache, which in combination allow for things such as encryption, data reduction by means of compression and in certain situations deduplication, end-to-end data integrity by means of check-summing and storing multiple copies of checksums of each data and metadata block elsewhere within a given pool. This is also integrated with a notion of snapshots, which leverage the same underlying building blocks, and made even more useful by read/write snapshots referred to as clones.

This in software implementation allows for various parity schemes as well as mirroring configurations. The following are schemes currently supported by RackTop:

- No Parity - fast, but with only minimal protection, and total loss if any single device is lost, useful for scratch-only data
- Mirrored - Equivalent to RAID 10 / RAID 1+0, aka a stripe of mirrors, where two or more drives in a mirror are possible, offers highest availability with a capacity trade-off
- RAIDZ1 (single parity) - Equivalent to RAID 50 / RAID 5+0, which allows for loss of a single drive in each group (vdev)
- RAIDZ2 (double parity) - Equivalent to RAID 60 / RAID 6+0, which allows for loss of two drives in each group (vdev)
- RAIDZ3 (triple parity) - like RAIDZ2, but with even more parity protection, allowing for loss of three drives in each group (vdev)

See the section about performance to understand the implications of each RAID scheme.

Resilvering

Resilvering is the process of rebuilding a disk within a vdev after a replaced. BrickStor OS does not have fsck repair tool equivalent, common on Unix filesystems. Instead, the filesystem has a repair tool called "scrub" which examines and repairs silent corruption and other problems. Scrub can run while the volume is online; scrub checks everything, including metadata and the data. This process works from the top down and only writes data to the disk that is needed. If a disk was temporarily offline it would only have to rebuild the data that was missed while the device was offline.

Pool Hierarchy and Containers

Within a pool, special containers exist. These Special Containers are used for organizing dataset and volumes so that they are always in the same location on a pool.

- Global – Contains all the datasets and other containers except for the tenant containers on a Pool
- Volume Container – Contains all virtual block devices which are special datasets exposed over iSCSI
- Replication – Top level container for all incoming replication streams from other pools within the same BrickStor or other BrickStor's
- Meta – Contains all of the user behavior audit data and the snapshot index data

Tenant Container (future use) – The tenant container has the same hierarchy as global but is designed to support future multi-tenancy capabilities. The tenant will have potential access to data in their tenant container and global.BP (Boot Pool)

The Boot Pool consists of two mirrored SSDs and contains the operating system and is a mirrored pool used to boot the appliance. This should be untouched during normal operations. Logs that are stored on the boot pool are set to auto rotate and expire to prevent any partition or directory from becoming full.

Initial Out of the Box Configuration

Default Accounts

BrickStor ships with a default administrative account for configuring the system. Similar to Unix, the root account has system wide superuser permissions within BrickStor.

Default Passwords

The default password for root accounts is “**racktop**”. This password is well known and should be changed immediately.

Initial Setup Tool

BrickStorOS comes pre-loaded with a command line program to use for initial setup. To use it, type ‘setup.sh’ from the command line. The following activities are available to set up via the script.

```
BrickStor Initial Setup Utility

Main Menu

1. Configure RMM interface.
2. Configure nodename.
3. Configure administrative (admin0) network interface.
4. Configure storage (data0) network interface.
5. Configure storage aggregate (aggr0) network interface.
6. Configure high availability (hb0) network interface.
7. Configure default gateway settings.
8. Configure NTP settings.
9. Configure DNS settings.
10. Configure appliance for none Internet based configuration.
11. Configure Local Key Manager.
12. Configure TimeZone.
13. Restart appliance.
14. System Information and Administration.
15. Import/Export Configuration.
16. Exit Setup Utility.

Please select menu option and press enter or press enter to exit.
Use CTRL-C to exit at anytime.
```

RMM (Remote Terminal) IP Address

Configuring the IP address for out of band management is required for full support and required to setup an HA Cluster. This out of band management has full control of the box including remote console and power control. It has its own physical network port with a dedicated IP and default gateway.

Configure node name

Configuring the host name to something other than the default is optional.

Configure Administrative Network Interface Admin0

Admin0 is the port required for management function and is the default port to be used for node management and to provide the ability to manage the node and is a static address. Out of the box this interface is enabled with a DHCP address. In an HA cluster the resource groups move between nodes and the IPs travel with the resource group, so it is important for management to have a static unchanging IP.

Configure Storage Network Interface Data0

This should be configured in non-HA clusters but is not used in an HA cluster setup. The Data0 vnic is the first interface to be created over the physical interface to serve network shares to clients. Resource groups handle this vnic within HA clusters.

Configure Aggregate over network interfaces

Use this to create an aggregate over multiple physical network interfaces for load balancing and higher network availability.

Configure High Availability Heartbeat Interface hb0

HA clusters require a direct network connection between the two nodes. This interface is called the heartbeat interface and should be named hb0 on both nodes. Once the physical interface for this direct connect is chosen and defined HA setup will finish the setup with a non-routable IP.

Configure Default Gateway

BrickStor only supports one default gateway. This is normally configured to make traffic to admin0 routable since most storage networks are not routable. However, customers should review their architecture with the BrickStor installation engineer.

NTP Setup

By default, NTP is set to use pool.ntp.org. It is most important that the time is synchronized with the organization's LDAP/Active Directory time because if there is greater than a 5-minute drift BrickStor will fall out of the domain and users will be unable to access their data.

DNS Setup

DNS is required for all environments.

Hosts Entries

Certain environments benefit from having local host entries to deal with situations when DNS is not available.

Local Key Management Configuration

If you are going to use drive or dataset encryption you will need to configure the local key manager or an external key manager. Use this option to configure the internal local key manager. See external documentation to configure a KMIP compliant external key manager. RackTop has specific instructions depending on the key manager and version for configuring external key managers.

You will be required to provide a password to protect the local key database. If you lose this password, you will not be able recover the database later if the configuration file is lost or changed. You should export and backup keys from the local key manager.

Configure Time Zone

The system can be configured to report time in the desired locale or UTC. Although all times are stored as Coordinated Universal Time (UTC), the time can be reported in whatever time zone is desired.

System Information and Administration (SIA)

Under this menu option are additional commands to join active directory, add licenses and add/remove local user accounts.

```
System Information and Administration Menu

1. BrickStorOS Version.
2. Hardware list.
3. Additional System Information
4. License Information
5. Show interface links.
6. Change local password on BrickStor appliance .
7. Add User to BrickStor appliance.
8. Remove User from BrickStor appliance.
9. Review current state of services.
10. Enable or disable service.
11. Add BrickStor appliance to Active Directory.
12. Check Active Directory.
13. IO Status Check.
14. Setup Syslog Receiver.
15. Add email to system fault notifications.
16. Remove email from system fault notifications.
17. Configure POSTFIX for mail relay.
18. Test POSTFIX for mail relay.
19. Add a license key to BrickStor.
20. Upgrade BrickStorOS.
21. Clickwrap BrickStorOS.
22. Support Bundle.
23. Reset BrickStor Appliance.
24. Register BrickStor Appliance.

Please select menu option and press enter or press enter to return to main menu.
█
```

Check Active Directory under SIA

This will verify everything is correctly configured and all required services are enabled to join active directory. If SMB/Server is not on you may need to create a data pool with an SMB share before you can join active directory.

Joining Active Directory under SIA

Joining active directory requires a Domain Admin account to join the domain one time. After that the system uses a certificate to authenticate to the Domain. An admin should run this command, enter their password and receive confirmation of a successful domain join.

Setup Fault Email Notifications under SIA

Setup the node to email fault alerts to an alias or email address. This is different than the system reports emails and is part of the systems fault management system. You can check this configuration by testing the postfix mail relay.

Syslog Receiver under SIA

Configure syslog forwarding so that logs are sent to a log centralization repository.

Additional Command Line Configurations

Adding and removing e-mail addresses from Report Notification List

To add e-mail addresses to receive notifications from the BrickStor appliance, use the following command format at the terminal:

```
# bsradm notify add <email address> --all
```

Other options besides the “all” notifications options are:

```
--system    Add to system notification list
```

```
--reports   Add to reports notification list
```

```
--faults    Add to faults notification list
```

To list users and their notification types, use:

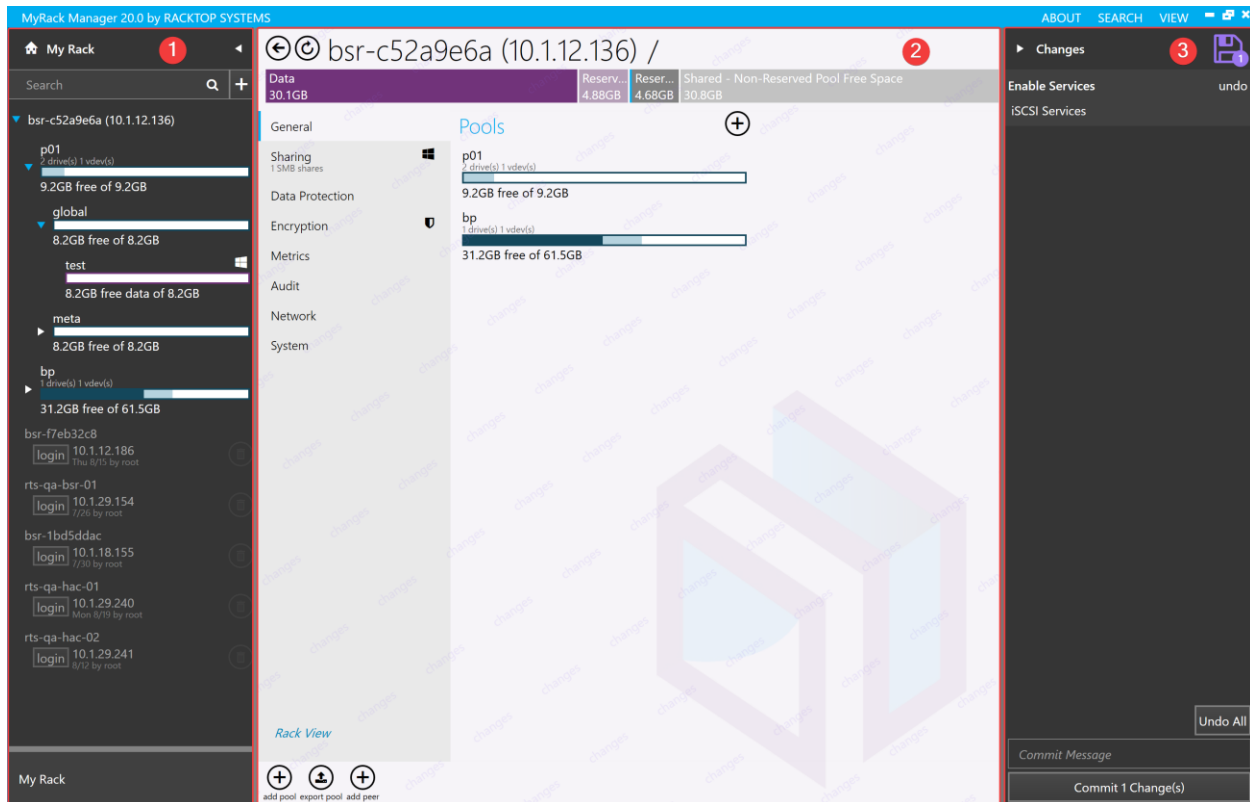
```
# bsradm notify show
```

And to remove users from their notification, use:

```
# bsradm notify remove <email address> --all
```


myRack Manager

General GUI Layout and Conventions

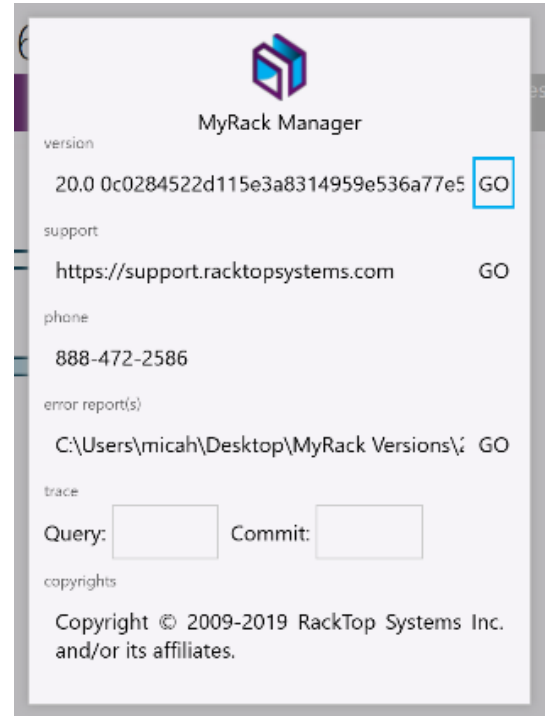


1. 'MyRack' navigation pane. Connect to BrickStor appliances, navigate their pools and datasets.
2. Main details page. Information for the currently selected appliance, pool, dataset, or feature.
3. Changes pane. If any changes have been made, they are listed here. Click 'Commit Changes' to apply the listed changes. The GUI does not actually make changes to BrickStor until they have been committed. Changes that make data unavailable or destroy data require the admin to acknowledge the possible negative affects before the commit button is activated. Note that HA changes and resource group movements are not processed through the commit queue.

At the top right, 'About,' 'Search,' and 'View' options are available.

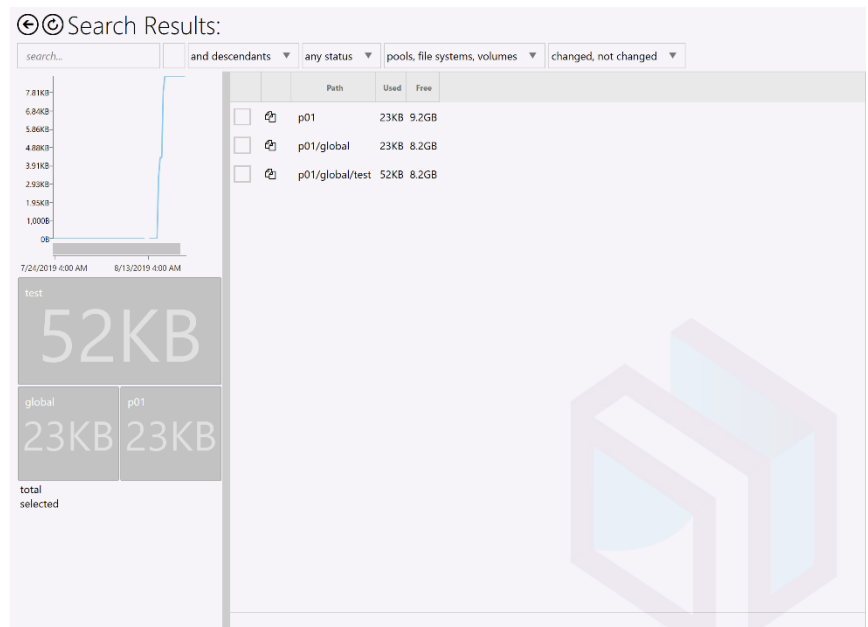
About: Shows MyRack Manager information.

By setting a value for example 5GB in the Trace Query and Commit box will create a local log on the machine running myRack manager with all of the GUI requests and responses.



Search:

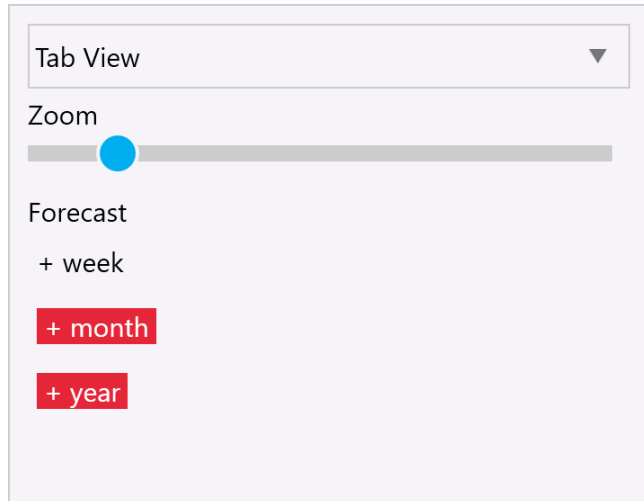
Brings up the Search Results page in the Main Details section of the GUI. Used to search through the appliance for pools, datasets, etc.



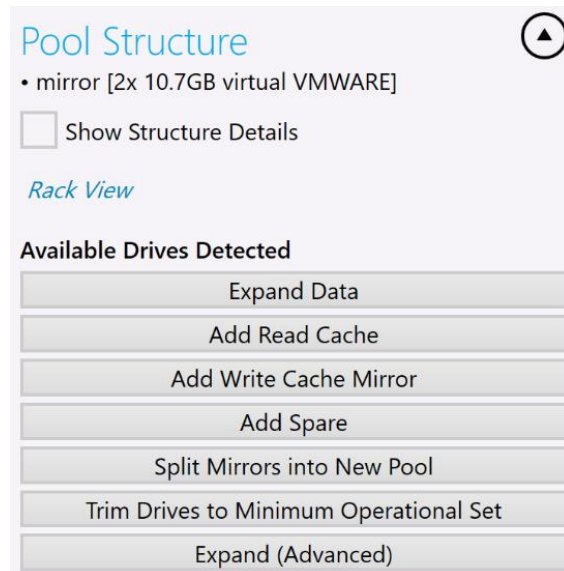
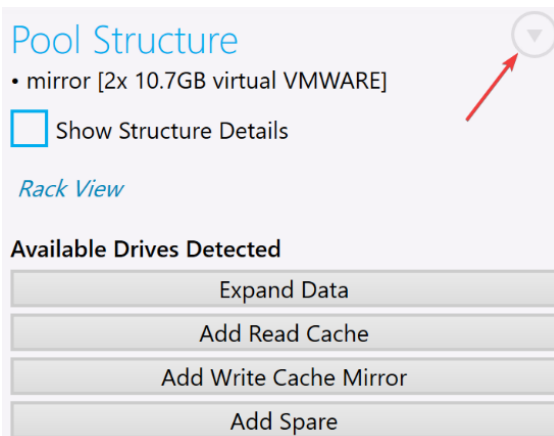
View:

Allows the user to change the layout between Tab View (separates sections into tabs, default) and Flow view (displays all sections next to each other), as well as viewing forecast data for the system. Tab view is the default view as of version 20 and is recommended for normal administration on small screens.

Zoom changes the width of columns in all views.

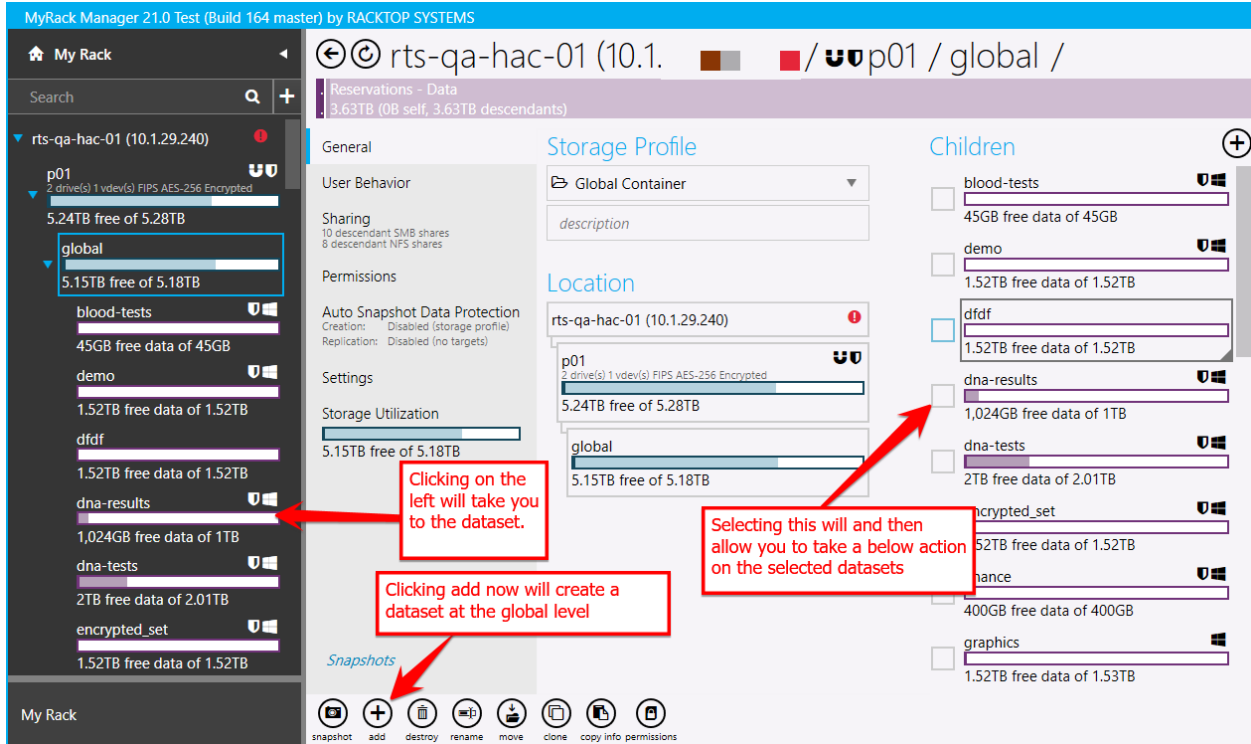


Some sections of MyRack Manager have arrow expansion buttons that reveal more options when clicked.



Dataset Creation and Manipulation

Once you have chosen a pool in the appliance in the left-hand navigation or you are on an existing share you can create dataset in a position relative to your location.



Create Dataset

Name(s)
sampledataset 1

Type - Storage Profile
General File System

Dataset Encryption
 On

Data Quota
X

Data Reservation
0B

NFS Share
 Off

SMB Share
 Off

AFP Share
 Off

Create 1 dataset(s) Cancel

Provide a dataset name and then choose the Storage Profile from the drop down menu based upon the proposed workload. Once you click create you will have an opportunity to continue to modify all the settings displayed in the initial create dataset window as well additional settings.

It is important to note that you cannot enable or disable dataset encryption after you have created the dataset by committing the changes. Similarly, you cannot disable deduplication for any dataset that has had it enabled without moving the data to a new dataset and destroying the old dataset. Most other operations are reversible however the changes only apply to new blocks and files as data in the dataset is modified and created.

More details about these advanced settings is provided in the dataset tab navigation menu section.

Tab Navigation

The tabs and menus available are based on the selection made in the navigation pane. When the top-level Appliance/Node is selected the user will be presented with different menu tabs than when a pool or dataset is selected for example. Also, certain tabs such as user behavior for example will not be visible if it is not enabled. The hierarchy of the navigation and tabs is Appliance, then Pool and then Dataset. If a menu such as user behavior is selected at the pool level the user will see all related activity to the pool. However, if they select it at the dataset level the scope will be narrowed to the dataset. Menus and tabs are relative to position within the GUI.

Appliance Level Menu Tabs

When you select an appliance or node in the left hand myRack navigation pane you will view Appliance level menu tabs.

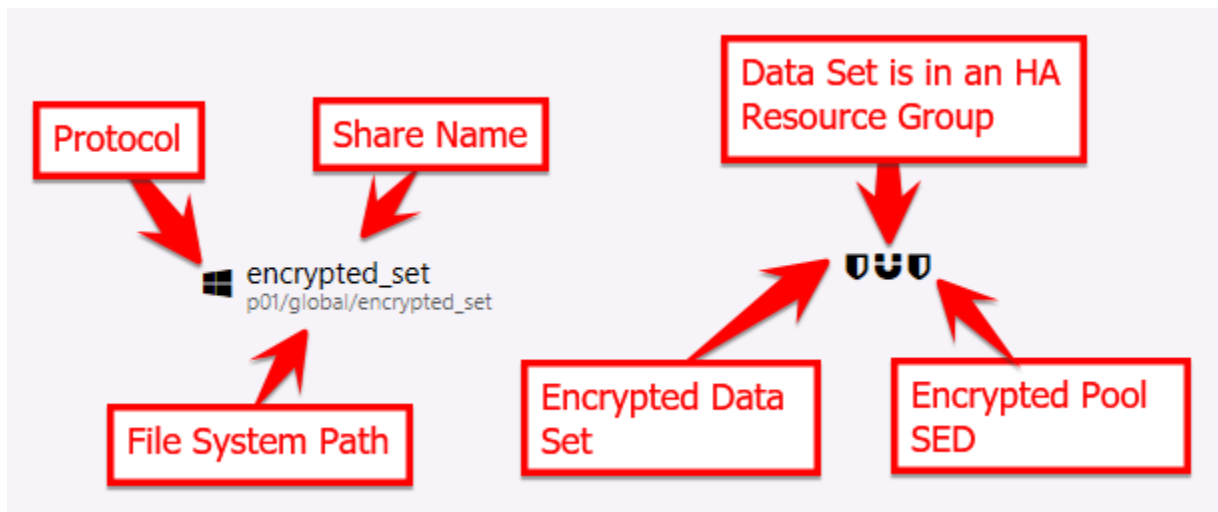
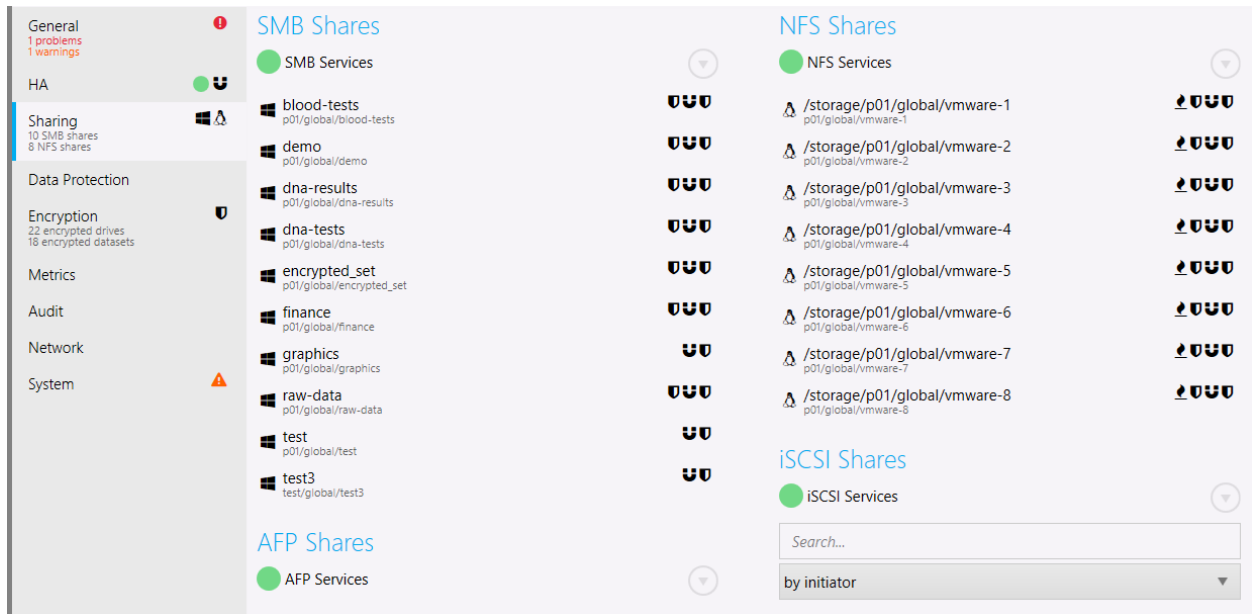
General

This tab lists all current problems and warnings with the node and its imported pools. From this view you can see which pool are currently imported an exported on the selected node.

The screenshot displays the 'General' tab in the myRack Manager interface. On the left, a navigation pane lists various system components: General (with 1 problem and 1 warning), HA, Sharing (10 SMB shares), Data Protection, Encryption (22 encrypted drives, 10 encrypted datasets), Metrics, Audit, Network, and System. The main content area is titled 'Notifications' and shows a red box for '1 Problem(s)' (10.1.19.2: Peer has one or more problem(s)) and an orange box for '1 Warning(s)' (bp (system): One or more warnings detected). Below this, the 'Pools' section lists 'p01' (2 drive(s), 1 vdev(s), FIPS AES-256 Encrypted, 5.24TB free of 5.28TB) and 'test' (8 drive(s), 4 vdev(s), FIPS AES-256 Encrypted, 21.1TB free of 21.1TB). At the bottom, 'Exported Pools' includes 'test2' (Imported by other HA node, 4 drive(s), 2 vdev(s), FIPS AES-256 Encrypted) with a 'move' button.

Sharing

Displays all shares currently on the node by protocol and displays if the datasets are encrypted and on self-encrypting drives. This view also provides a status of the protocol services and health.



Data Protection

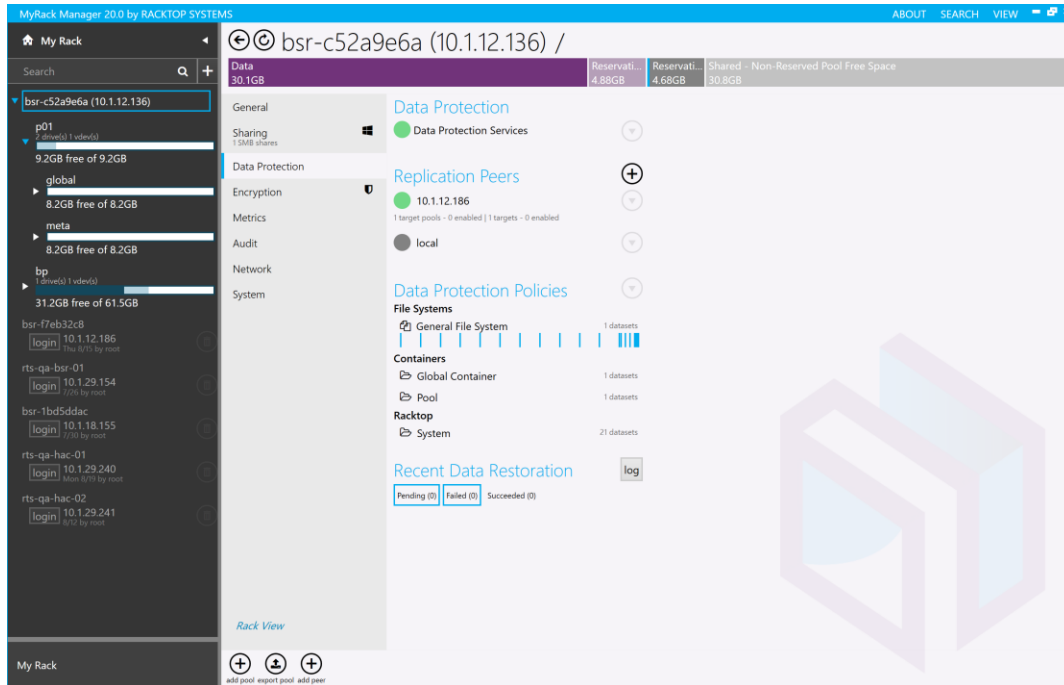
Data protection encompassed snapshots and snapshot replication. From this tab the admin can monitor data protection health and status for the node as well as configure replication and policies.

This tab shows the status of data protections services, peers, policies, and recent restoration.

On the Data Protection screen, you can:

- View the status of Data Protection and its services

- View and drill down into Replication Peers
- View the current status of Replication Tasks



Data Replication

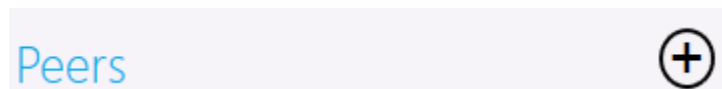
BrickStor supports block replication between two or more pools within the same system or across systems. In order to set up replication between two systems you must establish a peer relationship with the target system from the origin system. Once the peer relationship is created you can set up replication between pools on a per data set basis.

Configuring a Peer Relationship



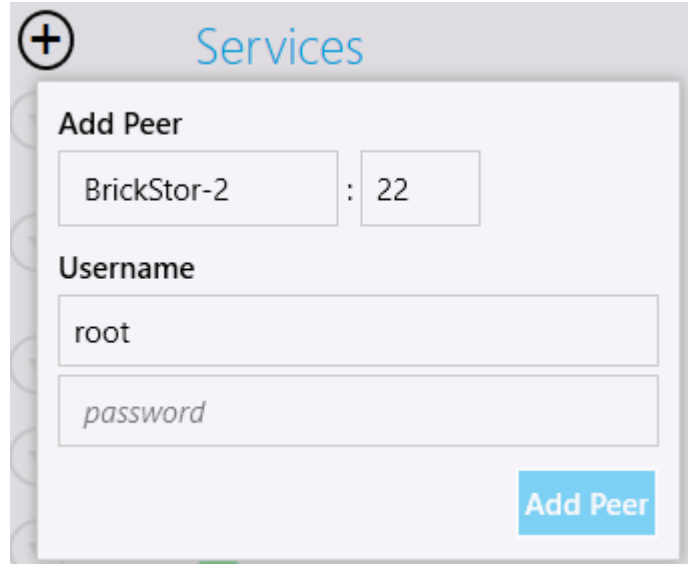
Click on the Add Peer Button at the bottom left of the main panel

Or the Plus Button next to Peers

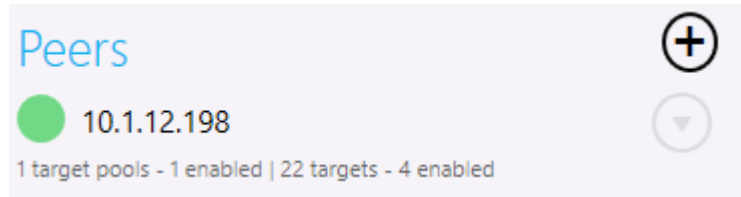


In the next dialogue simply enter the IP address and host name of the BrickStor you wish to add and then click the Add Peer button.

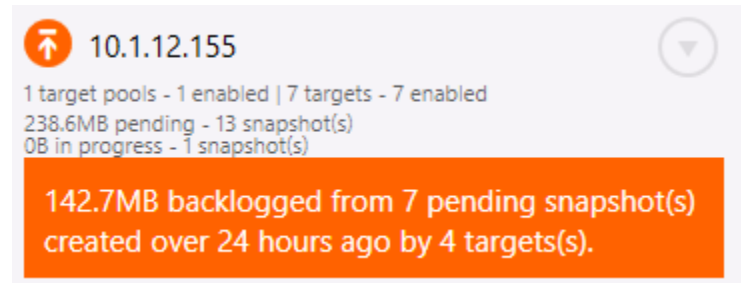
Now the Peer will appear in the list of Peers on the main screen. The Peer will be grey until you have added a target to that peer. You must repeat this process in order to replicate in the reverse direction on the other host.



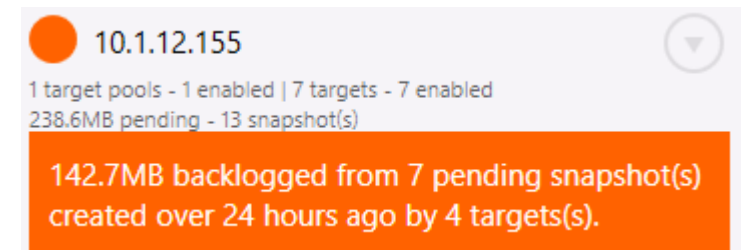
Peer Status Symbols
Healthy No Backlog



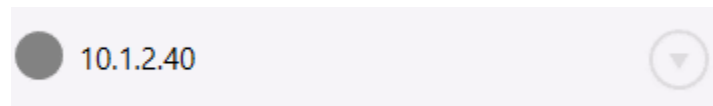
Backlogged with Transfer in Progress



Backlogged No Transfer in Progress



Peer Configured without replication targets enabled for Peer



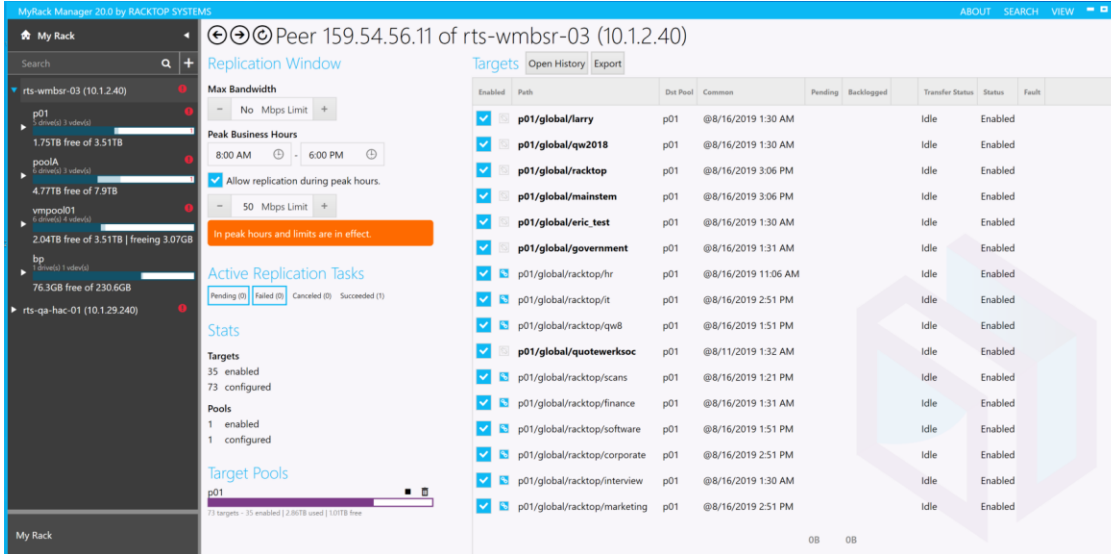
Peer has a Problem

The screenshot shows a list of peers under the heading "Peers". The first peer is 10.1.2.109, marked with a red circle. Below it, a red banner contains the error: "Error querying pools. cannot connect to 10.1.2.109 (dial tcp 10.1.2.109:22: i/o timeout)". A second red banner below that says "1 target pool(s) have a problem." The second peer is 10.1.12.106, also marked with a red circle. Below it, a red banner says "1 target pool(s) have a problem." Both peers have dropdown arrows to their right.

The reason for a Red Peer symbol is that the Peer is unreachable, the target pool is not imported and will show up as [unk] or the target pool is out of space.

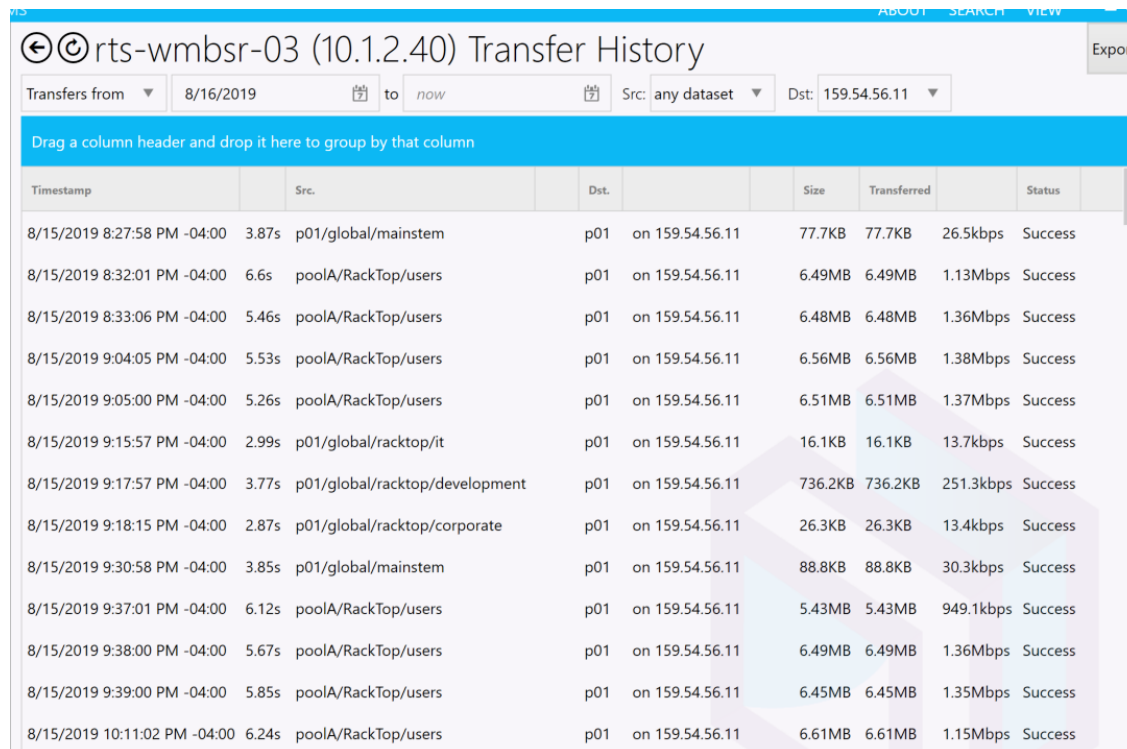
The screenshot shows the details for peer 10.1.12.106, which has a red status indicator. It lists "2 target pools - 2 enabled | 4 targets - 4 enabled". A red banner at the top says "1 target pool(s) have a problem." Below this, three target pools are listed: "bp" with a purple progress bar and a warning "Less than 20% free space available. 0 targets - 0 enabled | 378.1GB used | 52GB free"; "p01" with a purple progress bar and "2 targets - 2 enabled | 100.3GB used | 5.27TB free"; and "[unk]" with a warning "Missing from peer. 2 targets - 2 enabled | used | free [not verified]". Each target pool has a trash icon to its right.

Clicking on a Peer will take you to the replication details page for that peer.



This screen allows you to:

- Set replication window settings for bandwidth throttling and peak business hours
- View and configure replication targets
- Enable/Disable targets
- Set inheritance (whether to inherit replication parameters from the parent)
- View timing and transfer status
- Export a replication report
- Show the history of replication jobs by clicking the Open History button



This screen shows the details of transfers and can be filtered and exported. Details include:

- Time
- Duration
- Source / Destination
- Size
- Speed
- Success Status

Data Protection Replication

Data will be replicated to the target pool under the Replication Container. Through the GUI the source Hostname and IP will be visible along with the original dataset name. However, this information is stored in file system metadata on the replication target so it will not match the exact path name if an admin is browsing the file system on the pool.

Data Replication Hierarchy on filesystem

<Pool Name>

- global
- - replication
 - o <Serial Number of Source BrickStor>
 - Data Set GUID of Source Data Set

Data Protection Policy Configurations

The screenshot shows the 'Data Protection Policies' configuration page. It is organized into several sections: File Systems, Server Storage, Containers, Racktop, and Replication. Each section lists various storage profiles and their associated dataset counts. Red arrows and boxes provide annotations:

- An arrow points to a dropdown arrow icon with the text: "Expand to view storage profiles not currently in use".
- An arrow points to the 'General File System' entry (21 datasets) with the text: "Number of datasets provisioned with that storage profile".
- An arrow points to the 'Oracle Volume' entry (1 datasets) with the text: "Click on storage profile name to configure data protection policy".
- An arrow points to the 'Global Container' entry (2 datasets) with the text: "Graphical representation of snapshot schedule".

Category	Item	Count
File Systems	General File System	21 datasets
Server Storage	Oracle Volume	1 datasets
Containers	Global Container	2 datasets
	Pool	1 datasets
	Volume Container	1 datasets
Racktop	System	31 datasets
Replication	Replication Container	1 datasets

Configure the Data Protection Policy for a Storage Profile

← ↻ General File System profile on rts-qa-hac-01 (10.1..

Default Snapshot Policy Reset

On

Frequency	Retention			
Every 5 min(s) ▼	-	30	day(s)	+
Daily consolidation	-	no	day(s)	+
Weekly	-	3	week(s)	+
Monthly	-	12	month(s)	+
Yearly	-	1	year(s)	+

These settings only apply to new snapshots. Existing snapshots will expire based on the settings at the time of snapshot creation. Sub-daily snapshots will be skipped when no change occurs.

Auto Replicated Snapshots

Have alternate retention ▼

Every 5 min(s) ▼	-	60	day(s)	+
Daily consolidation	-	no	day(s)	+
Weekly	-	4	week(s)	+
Monthly	-	12	month(s)	+
Yearly	-	1	year(s)	+

Additional snapshots may be retained locally until replicated. The replica retention is captured at snapshot creation and will not affect existing snapshots.

Set snapshot retention policy on local storage pool.

Set alternate retention policy on the remote storage pool.

Set snapshot retention policy on the remote storage pool.

Encryption

This tab shows the status and options relating to Self-Encrypting Drives (SEDs) and the Key Manager used for individual dataset encryption. Note that SED management requires a valid TCG license.

For the Drives you can view which drives are SED capable. The boot pool is typically not SED capable or enabled.

SED Pool Status Meanings

- Not encrypted
- FIPS AES-256 encrypted
- FIPS AES-256 encrypted (data only) – Cache drives aren't SED
- FIPS AES-256 encrypted (partial) – Some data drives aren't SED
- FIPS AES-256 encrypted (partial enrolled) – Some drives have not been enrolled but are SED Capable

The screenshot shows the myRack Manager interface for Drive Encryption (SED). On the left, a sidebar lists system components: General (1 problem, 1 warning), HA (green status), Sharing (10 SMB shares, 8 NFS shares), Data Protection, Encryption (22 encrypted drives, 18 encrypted datasets), Metrics, Audit, Network, and System (orange warning). The main area is titled 'Drive Encryption (SED)' and shows 22 drives enrolled in FIPS AES-256 encryption, all 22 are unlocked and ready for auto-lock, and 3 are not supported. Below this are buttons for 'Verify Keys', 'Rekey', 'Export SED Keys', 'Unenroll', and 'Config (Advanced)'. An 'Encrypted Pools' table lists three pools: 'p01' (2 drives, 1 vdev), 'test' (8 drives, 4 vdevs), and 'test2' (imported by other HA node, 4 drives, 2 vdevs). On the right, 'Dataset Encryption' shows 18 datasets are AES-256 encrypted, 18 are unlocked and accessible, and 4 are not encrypted. Below that, a 'Key Manager' section has buttons for 'Export All Encryption Keys' and 'Import Encryption Keys'. At the bottom right, there is a green 'Encryption Services' indicator.

Drive Encryption Related Buttons

Verify Keys – Checks that the node has access to all the appropriate data drive unlock keys through the configured key manager.

Rekey – Changes the data drive unlock key for the data drives by requesting a new key from the key manager and applying it to the SED drive.

Export SED Keys – Exports SED keys to a password protected file that will be saved to the machine running the myRack Manager GUI. This feature must be enabled in the secured service configuration.

Unenroll – Unenroll takes the drive out of the FIPS compliant configuration, sets the drive not to auto lock when power is removed and sets the data drive lock key back to a known default. This feature must be enabled in the secured service configuration. This can be used if you want to transfer the disk to another system without having to share the key. However, the drive will not be protected in transit. It is also a safe way to change from one key manager to another and not have to worry about managing keys through the transition.

Config Advanced – This is only for modifying how often the secured service is performing low level functions.

Key Manager Buttons

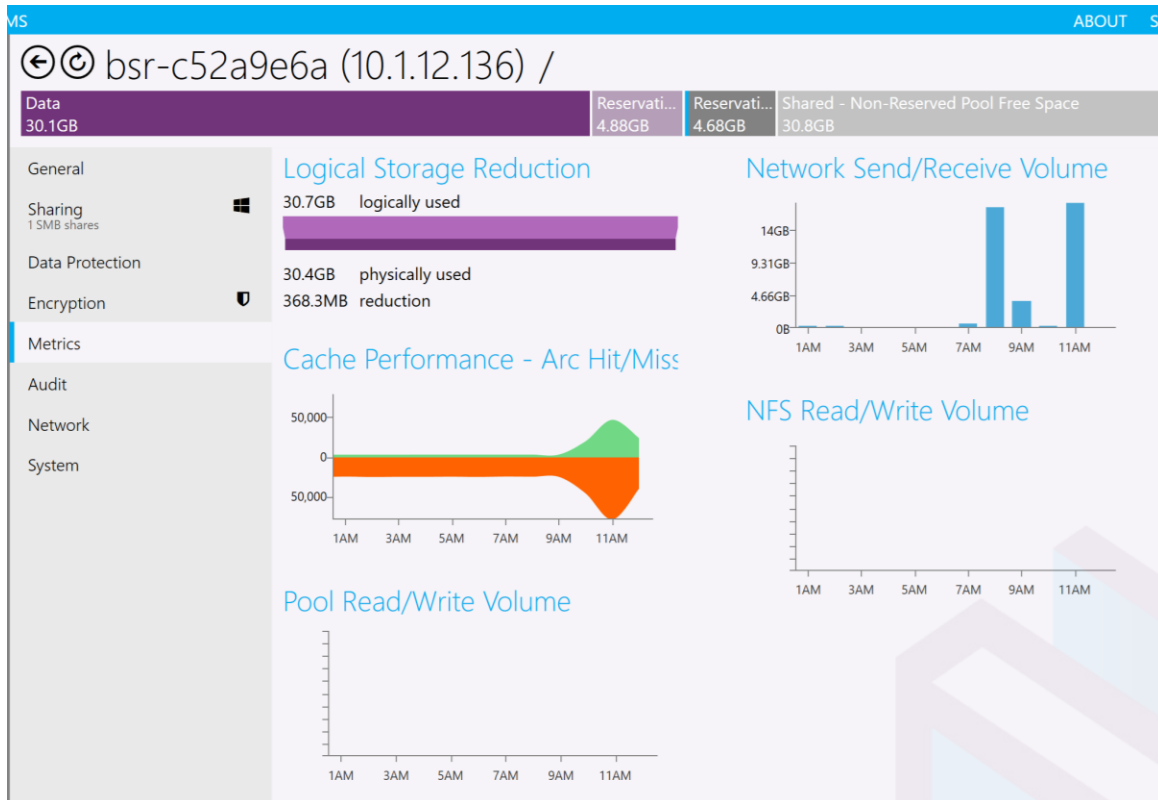
Export All Encryption Keys – Exports SED and dataset keys to a password protected file that will be saved to the machine running the myRack Manager interface.

Import Encryption Keys – Imports keys from a password protected file created by myRack Manager.

Until Version 21 if you are using dataset encryption you will need to manually export and import dataset encryption keys to the replication target and the other node in the HA cluster. It is not a problem to import the same key more than once. It will not create duplicate entries.

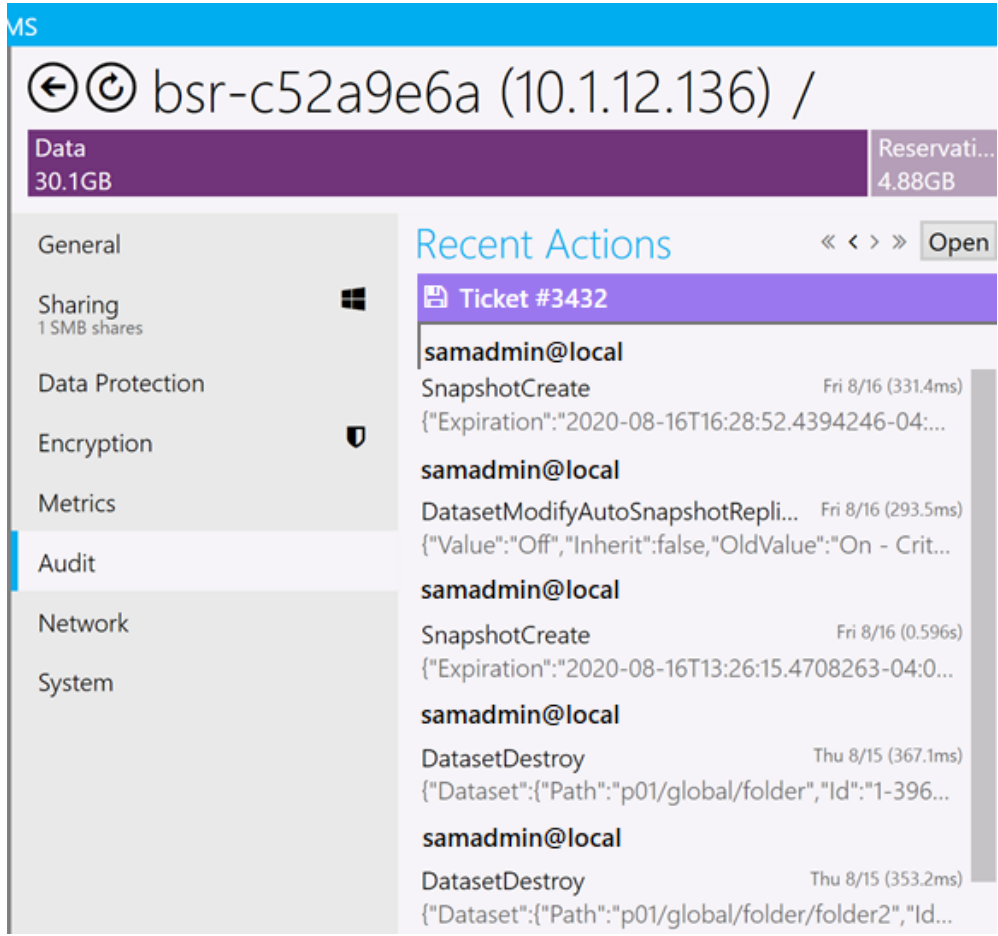
Metrics

This tab contains various charts and graphs relating to storage capacity, cache performance, bandwidth utilization and metrics per sharing protocol.

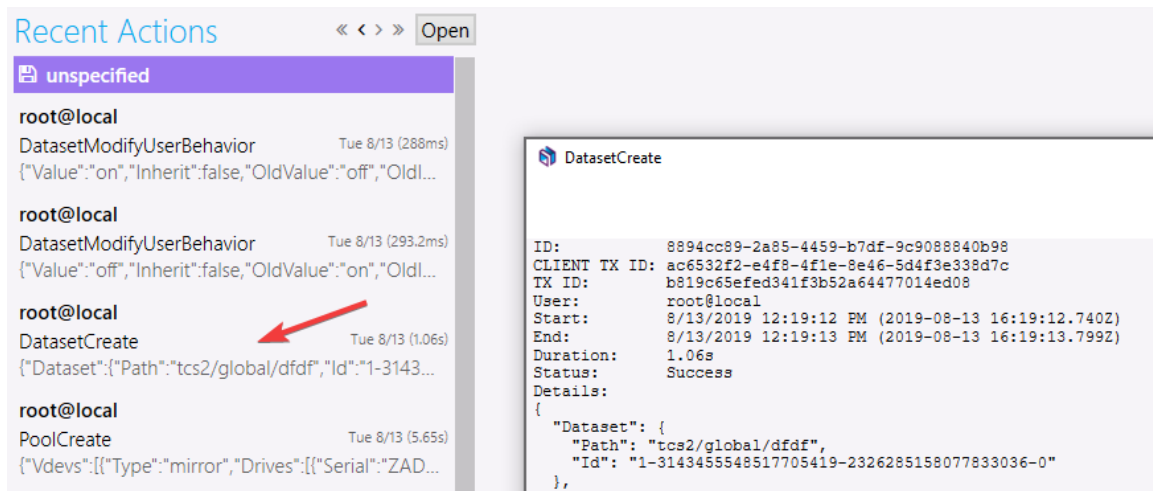


Audit

This screen shows a list of the administrator actions performed through the API and GUI and associated with the user ID of the admin and any the optional commit message if entered when the changes were committed.



Hovering on any of the actions display all of the API messages posted for the change.



Network

The network tab displays all of the interfaces and should have a green status indicator for all vnics. Each interface shows the IP, interface name, physical interface or aggregate the vnic is on, MTU size and port speed.

System

This tab contains system information, service status, and the BrickStor operating systems available for download and installation.

Admins can look on the service tab to find their customer ID, Serial Number and the running version of the OS when calling support. From this admin can all power off and reboot the node as well as access compliance reports. It is from this tab that the admin configures the HA Cluster once the command line steps have been completed. See HA Cluster Configuration for cluster setup details.

The screenshot shows the 'Hardware' tab in the myRack Manager interface. On the left is a navigation menu with categories: General (1 problem, 2 warnings), HA, Sharing (10 SMB shares, 8 NFS shares), Data Protection, Encryption (22 encrypted drives, 18 encrypted datasets), Metrics, Audit, Network, and System (1 warning). The main content area displays the 'BRICKSTOR' logo and the following information: Customer ID: CN000001, Manufacturer: RackTop Systems, Product: None, OS: BrickStorOS 20.0.1, and Serial Number: ZZ0000SS. Below this is a storage bar for 'bp (system)' showing 77GB free of 230.6GB. There are two buttons: 'Setup HA Cluster' and 'Export Encryption Keys'. At the bottom, there are checkboxes for 'System: Reboot' and 'Shutdown', and three links: 'Rack View', 'RMM Console', and 'Compliance Reports'.

The screenshot shows the 'Services' tab in the myRack Manager interface. The left navigation menu is identical to the previous screenshot. The main content area is divided into two sections. The top section, 'Hardware', repeats the system information: BRICKSTOR logo, Customer ID: CN000001, Manufacturer: RackTop Systems, Product: None, OS: BrickStorOS 20.0.1, Serial Number: ZZ0000SS, and storage bar for 'bp (system)' (77GB free of 230.6GB). Below this are the 'Setup HA Cluster' and 'Export Encryption Keys' buttons, and the 'System: Reboot' and 'Shutdown' checkboxes. The bottom section, 'Services', lists various services with their status: System Services (Running), HA Services (Running), Encryption Services (Running), Data Protection Services (Running), SMB Services (Available), NFS Services (Running), AFP Services (Running), and iSCSI Services (Running). The 'SMB Services' entry is expanded to show sub-services: network/ntp, network/security/ldkit_warn, network/smb/client, system/dmap, system/filesystem/local, and network/smb/server, all of which are shown as running.

High Availability (HA)

BrickStor enables admins to manage data availability through resource groups. The cluster settings and resource and resource groups are managed through the HA Cluster tab.

General

HA
1 temp resource group assignments

Sharing
38 SMB shares
708 NFS shares

Data Protection

Encryption
22 encrypted drives
719 encrypted datasets

Metrics

Audit

Network

System

HA Cluster

1 temp resource group assignments

Witness

- Peer rts-qa-hac-01
- Peer rts-qa-hac-02

rts-qa-hac-02 (241)

- NTP
- Peer
- Peer Crossover
- Peer Power
- RMM
- Witness

rts-qa-hac-01 (240)

- NTP
- Peer
- Peer Crossover
- Peer Power
- RMM
- Witness

test	→	●
🏠 192.255.1.240/24		●
🏠 1.2.3.4/24		●
📀 p01		●
📀 test		●

test with 2 vnics	temp →	●
🏠 192.255.1.241/24		●
📀 test2		●

Auto Fail-over

Move resources and disable node ▼

Automatically move HA resources when a node is degraded or faulted.

3:00 PM ⌚

Automatically move HA resources groups back to non-disabled preferred nodes.

Cluster Management

The screenshot shows the HA Cluster management interface. At the top, there are several icons: a plus sign with 'R', a plus sign with 'P', a gear, a square, and a scale. Red callout boxes point to these icons with the following labels: 'Add Resource Groups' (pointing to the plus sign with 'R'), 'Add Pools to HA Cluster' (pointing to the plus sign with 'P'), 'Configure polling and timeout settings' (pointing to the gear icon), 'Disable Cluster' (pointing to the square icon), and 'Rebalance Cluster' (pointing to the scale icon). Below these icons, there are sections for 'Witness', 'rts-qa-hac-01 (240)', and 'test'. Each section lists various components like NTP, Peer, Peer Crossover, Peer Power, RMM, and Witness, each with a green status indicator. A table below shows resource groups with their components and status indicators. Red callout boxes point to these status indicators with labels: 'Health of resource group overall' (pointing to a green circle with a right arrow), 'vnics in resource group' (pointing to a green circle), and 'Pools in resource group' (pointing to a green circle). At the bottom, there is an 'Auto Fail-over' section with a dropdown menu set to 'Move resources and disable node' (labeled 'Auto Failover Action') and a time field set to '3:00 PM' (labeled 'Rebalance Time').

Capabilities


- Add Resource Groups – Clicking on this allows the admin to create a new resource group.
- Add Pools to HA Cluster – Adding a pool
- Configure Time out Settings
- Disable Cluster
- Rebalance Cluster Manually
- Monitor Cluster and Component Health
- Configure Fail-over response
- Configure an automatic rebalance time

Creating and Configuring Resource Groups

Create a new resource group by hovering over one of the nodes on the HA Cluster tab

rts-qa-hac-01 (240)

- NTP
- Peer
- Peer Crossover
- Peer Power
- RMM
- Witness



By default a simple resource group configuration window will appear that will allow the admin to create a resource group with one vnic on the default interface. The admin must specify:

- Resource group name
- vnic IP and subnet
- Member pools
- The node to create the resource group on initially
- The preferred node where it will be placed on rebalance

Create HA Resource Group Advanced

Description

VNIC

CIDR address

VNIC address required (example 1.2.3.4/24).

Add VNIC

Pools Show All

- p01 on test on rts-qa-hac-01 (240)
- test on test on rts-qa-hac-01 (240)
- test2 on test with 2 vnics on rts-qa-hac-01 (240)

Node

rts-qa-hac-02 (241) ▼

Preferred Node

None ▼

Create
Cancel

The admin can configure an advanced resource group with multiple vnics, vlan tags and use interfaces other than the default cluster data interface by clicking on the Advanced button in the top right of configuration window.

Create HA Resource Group

advanced resource group 1 X

VNIC	Over	VID	MTU	Description	
192.168.24.44/24	aggr0 (default) ▼	23	auto	priv3	🗑
192.168.5.22/24	ixgbe2 ▼	12	auto	legacy	🗑
<i>CIDR address</i>	aggr0 (default) ▼	<i>vid</i>	<i>auto</i>	<i>description</i>	🗑

VNIC address required (example 1.2.3.4/24).

Add VNIC

Pools Show All

- p01 on test on rts-qa-hac-01 (240)
- test on test on rts-qa-hac-01 (240)
- test2 on test with 2 vnics on rts-qa-hac-01 (240)

Node

rts-qa-hac-01 (240) ▼

Preferred Node

rts-qa-hac-01 (240) ▼

Create
Cancel

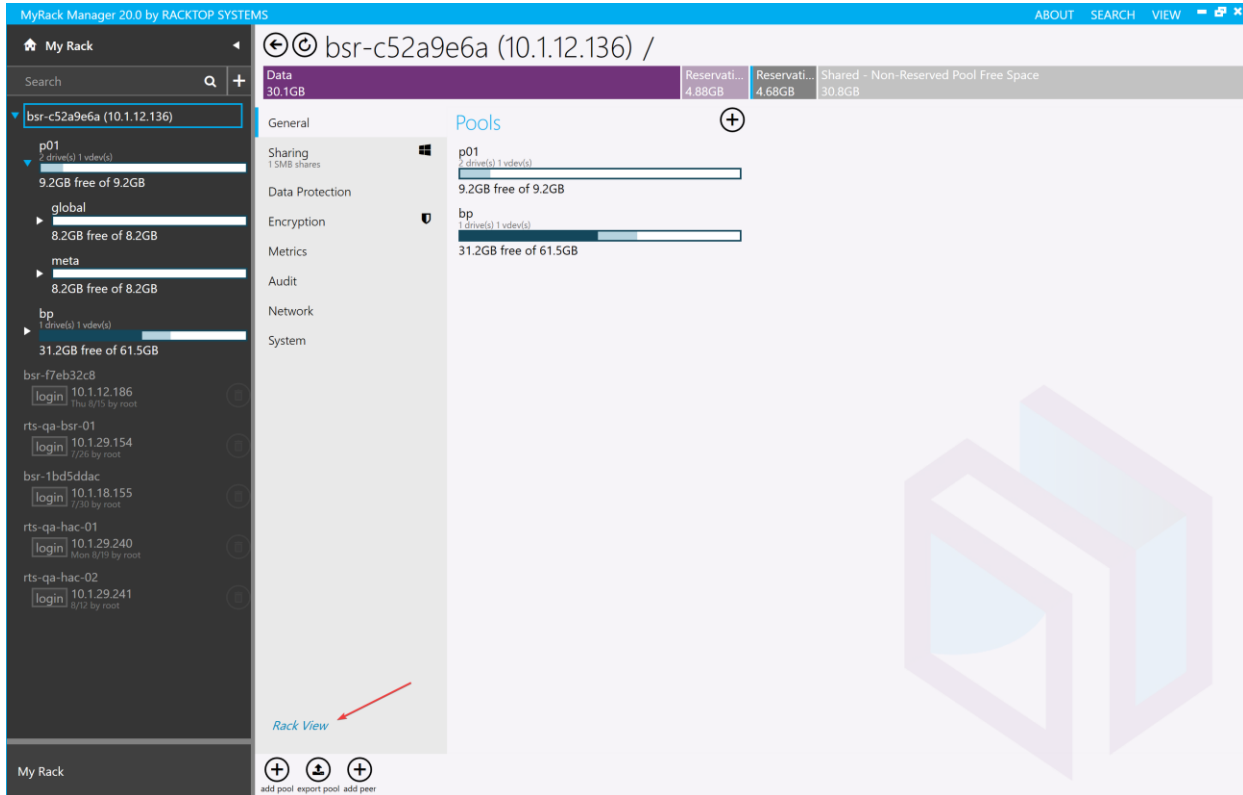
In addition to providing the IP and subnet for each vnic the admin will be able to configure

- Interface for vnic (Over)
- VLAN ID (VID)
- MTU Size for the vnic
- Network/VLAN Description

Appliance Level Links

Rack View

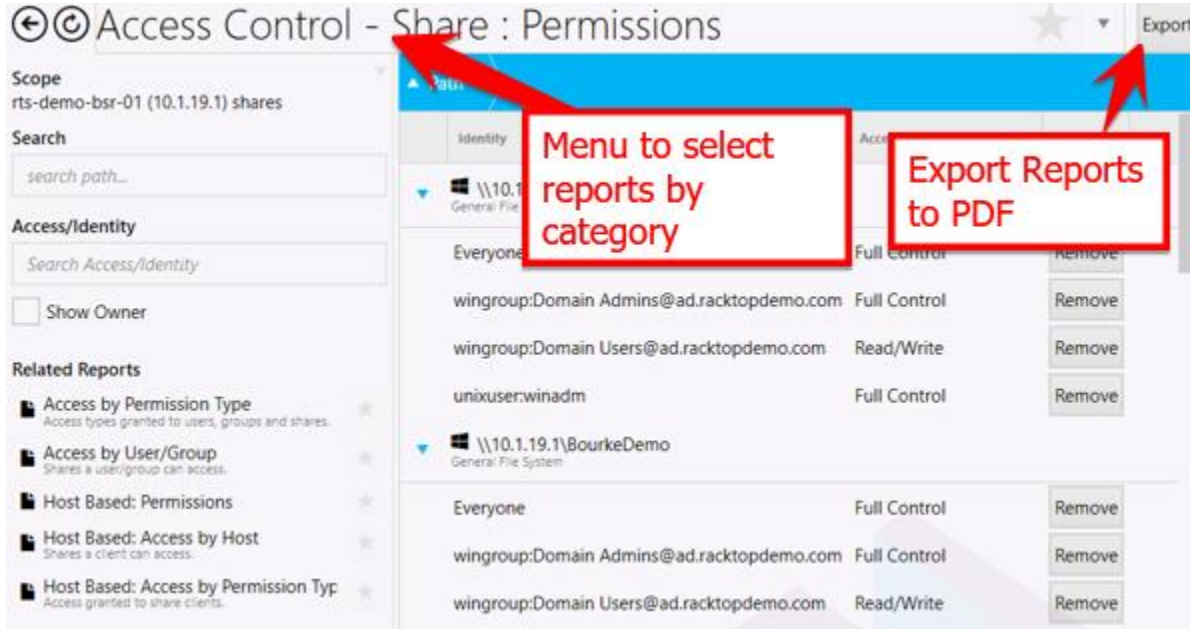
From any Appliance level tab, click on the 'Rack View' link to go to Rack View described later in this document.



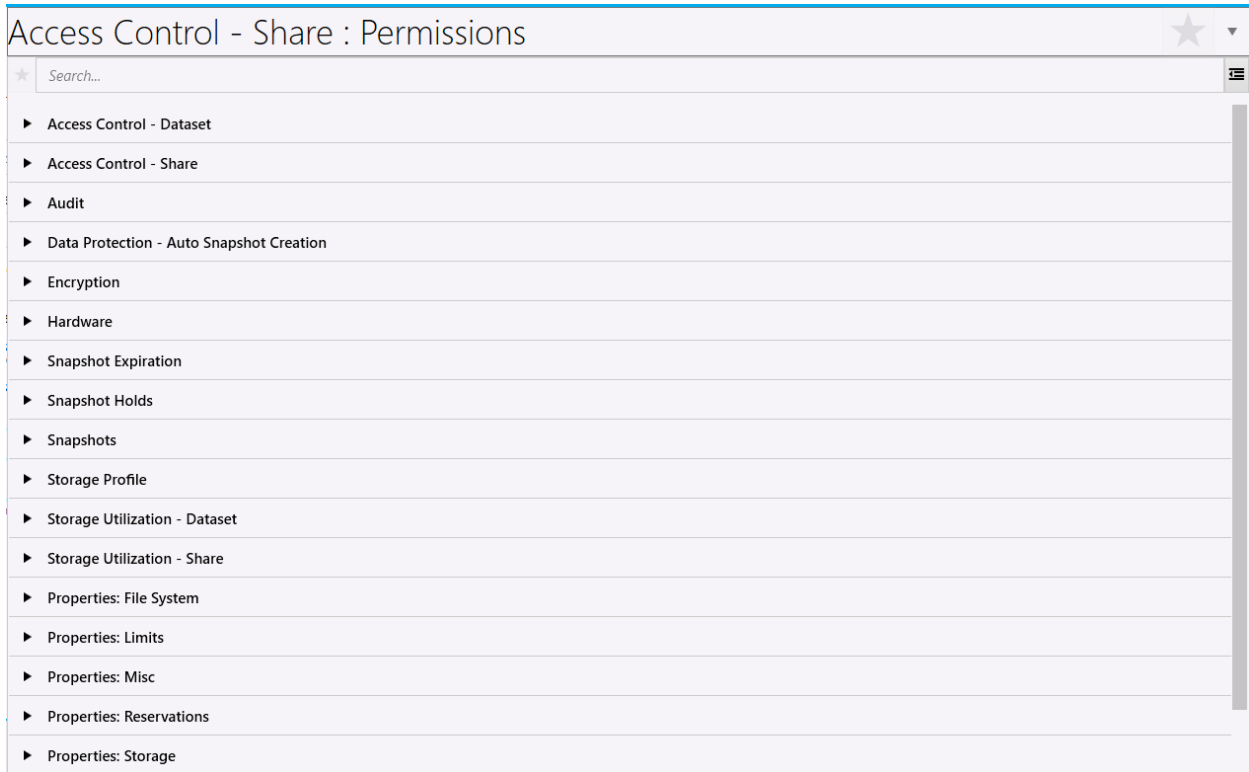
Compliance Reports

myRack Manager provides various exportable reports and can be accessed from the System Menu tab on the appliance level.

Compliance reports cover permissions management, data protection, data disposition reporting and other reports that are valuable for security and compliance with internal policies and government regulations. The compliance reports are designed to provide evidence of continuous compliance with standard data related controls.



You can also favorite your frequent reports by clicking the star outline.



Pool Level Only Menu Tabs

Selecting a pool from the navigation pane brings up another set of tabs.

Pool

This tab has information about the Pool's structure and performance.

The screenshot displays the MyRack Manager 20.0 interface. The top navigation bar shows the current rack and pool: `bsr-c52a9e6a (10.1.12.136) / p01 /`. The left sidebar lists various storage pools and their free space: `p01` (9.2GB free of 9.2GB), `global` (8.2GB free of 8.2GB), `meta` (8.2GB free of 8.2GB), `bp` (31.1GB free of 61.5GB), `bsr-f7eb32c8` (10.1.12.186), and `rts-qa-bsr-01` (10.1.29.154). The main content area shows the `p01` pool details, including its configuration (2 drive(s), 1 vdev(s)), sharing information (1 descendant SMB shares), and storage utilization (9.2GB free of 9.2GB). The right-hand panel provides options to expand the pool, add read/write cache mirrors, and spare drives, as well as a 'Start Scan' button for error checking.

From this tab you can expand the pool as well as add read and write cache devices. In the case of a problem or degraded pool it will give the admin the opportunity to replace a failed disk under the Notable vDevs section.

From pool tab admins can start a “Scan” to perform a scrub operation which reads every block of data and checks for errors against the block’s 256-bit checksum. This system process can impact system performance because it is read intensive. However, performing this scan will ensure stale data that hasn’t been accessed in long time is check for bit errors to avoid bit rot. This process happens automatically on every read operation from disk.

Progress of the scan can be monitored in real time.

The screenshot shows the 'Scan & Fix Errors' section. It indicates that a scrub operation is in progress, starting on Sun Sep 8 15:27:51 2019. The progress shows 914M scanned at 26.1M/s, 283M issued at 8.08M/s, with a total of 3.22T scanned, 0 repaired, and 0.01% done. No estimated completion time is provided. A 'Stop Scan' button is visible at the bottom.

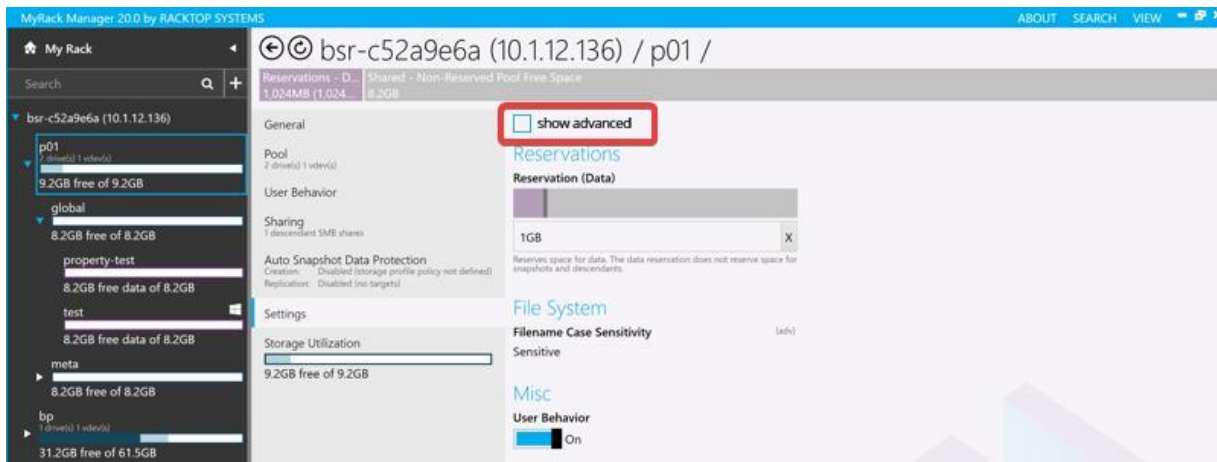
Sharing

The sharing tab shows the same information as the Sharing menu at the appliance level but scoped only to those shares on the selected pool.

Settings

This tab contains settings that apply to the pool including a pool level reservation. The pool reservation by default is set to 10% of the pool capacity up to 100GB. This is in place as a safety measure to prevent the pool from becoming completely full and making it difficult to do the necessary operations to remove data. When the pool becomes full the admin can release some or all of the Pool reservation.

There is a hidden checkbox at the top of the page, 'show advanced,' that will provide more options.



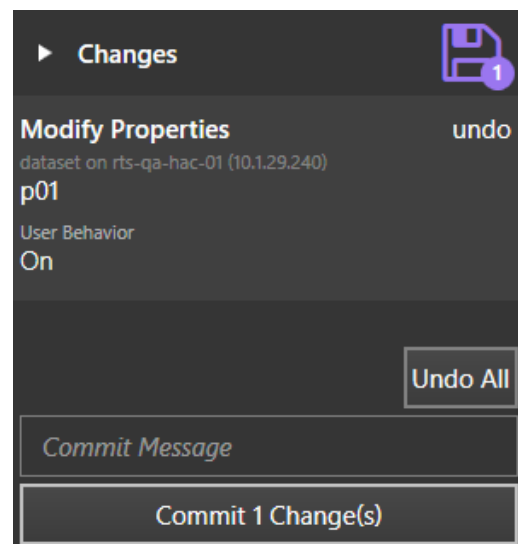
Enabling User Behavior

User Behavior can be enabled at the pool level or the dataset level. To enable it, navigate to the 'Settings' tab for a pool or a dataset. Click on the toggle switch and commit the changes.



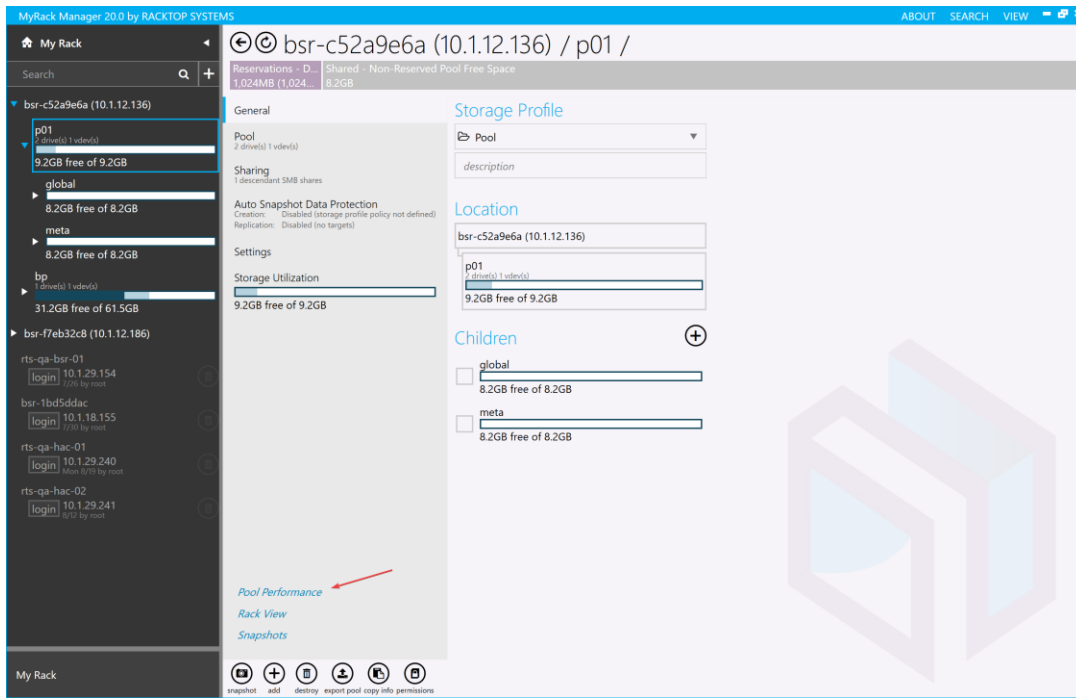
This action starts logging the behavior of the users at the level it was turned on for, and its descendants.

By default, data is stored in the meta dataset of the pool.

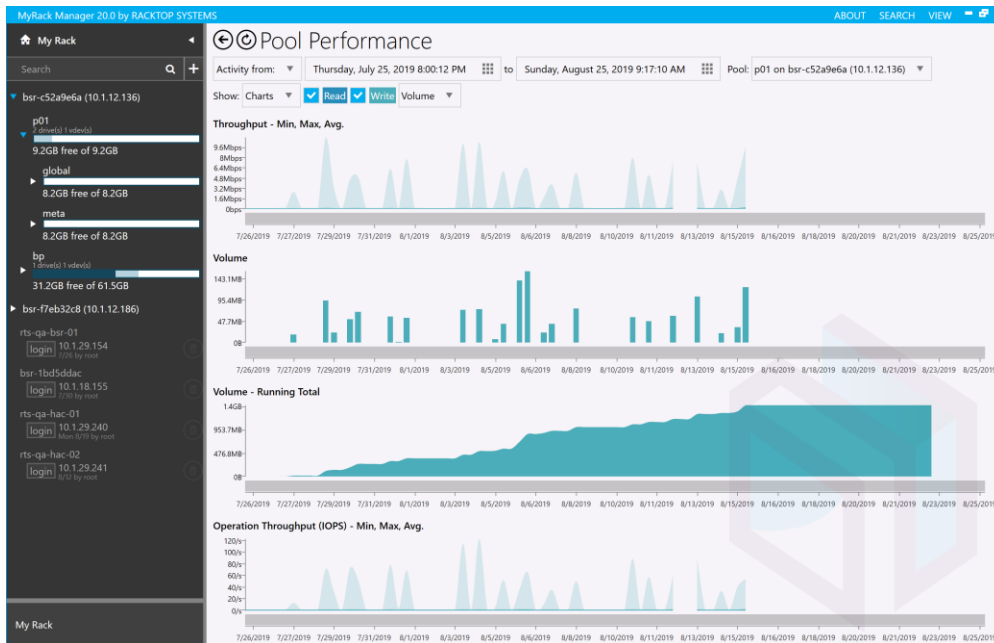


Pool Level Only Links

Pool Performance



Clicking on the 'Pool Performance' link leads to a page with charts and graphs about this pool's performance history. Admins can zoom in on the graph to look at specific time periods.

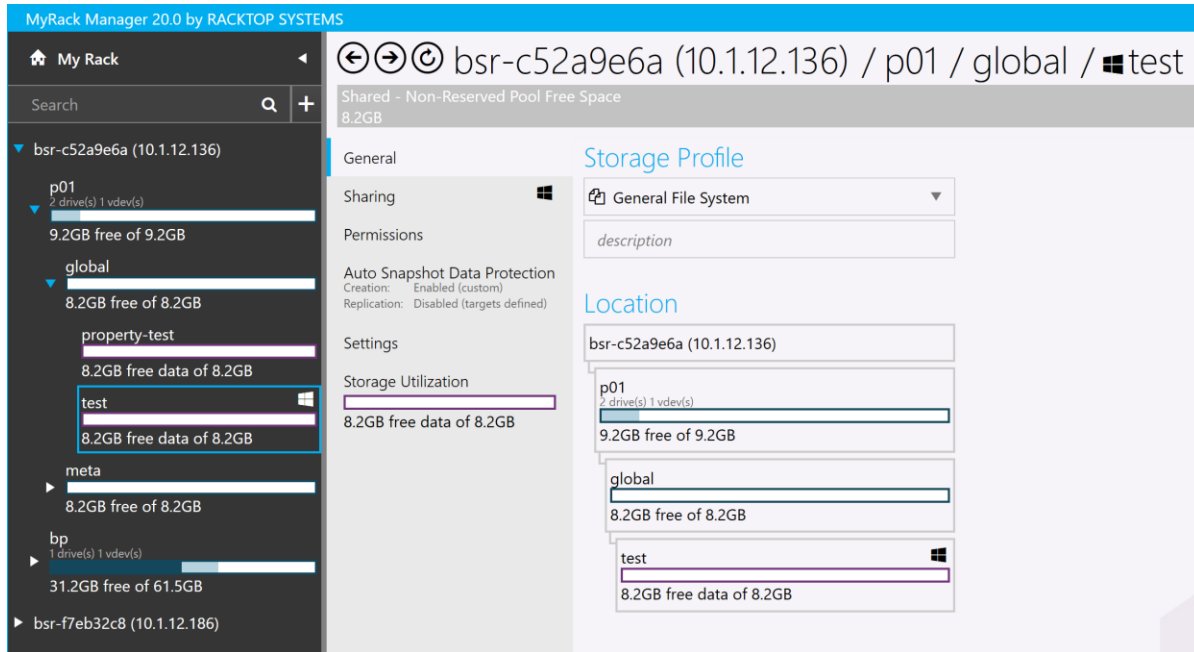


Pool and Dataset Level Menu Tabs

The following tabs are present when either a pool or a dataset is selected from the navigation pane.

General

This tab contains general info about the pool or dataset.



User Behavior

If enabled, this tab contains information about the behavior of users on the system.

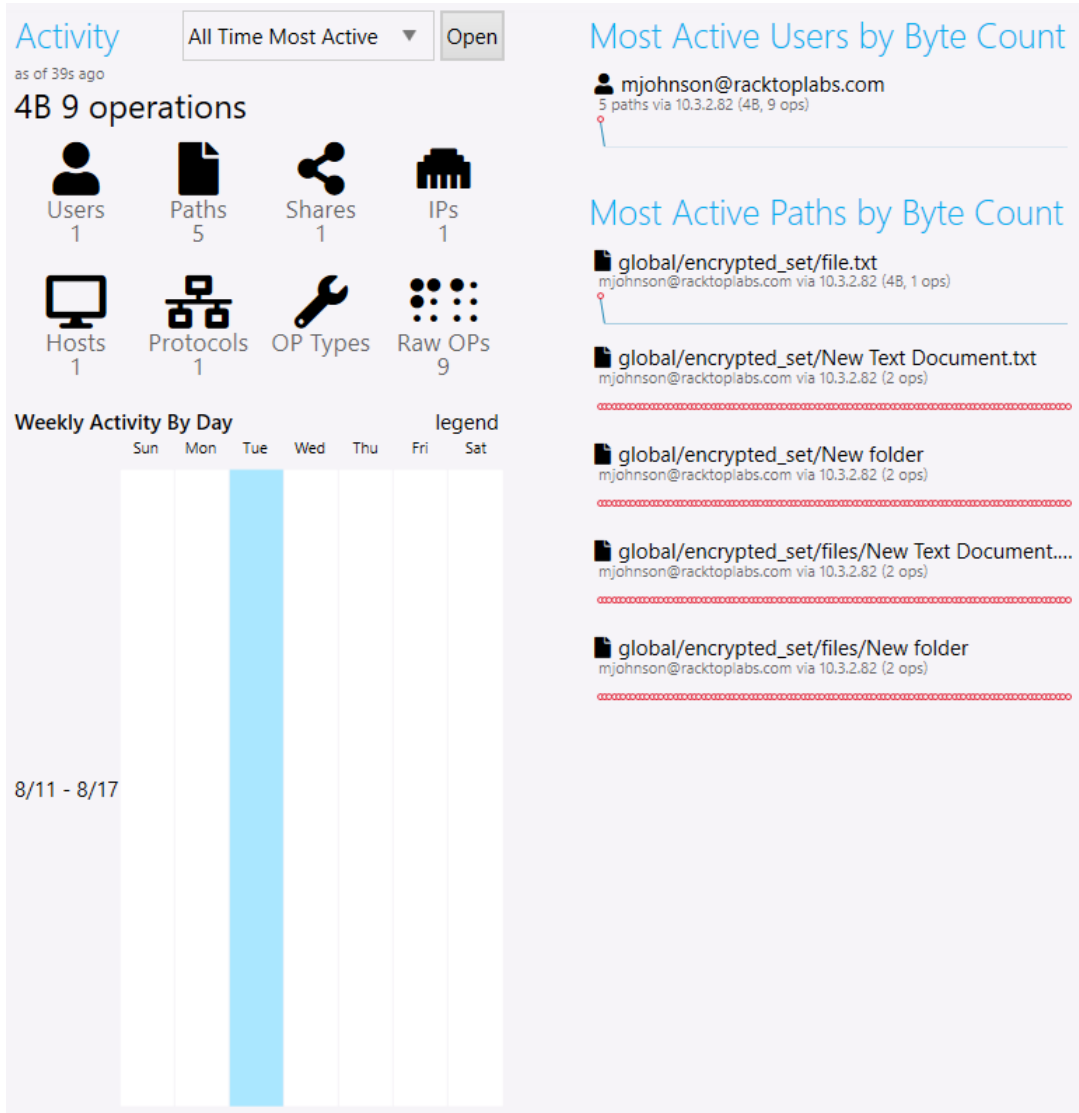
User Behavior Auditing and Analysis (UBA)

Overview

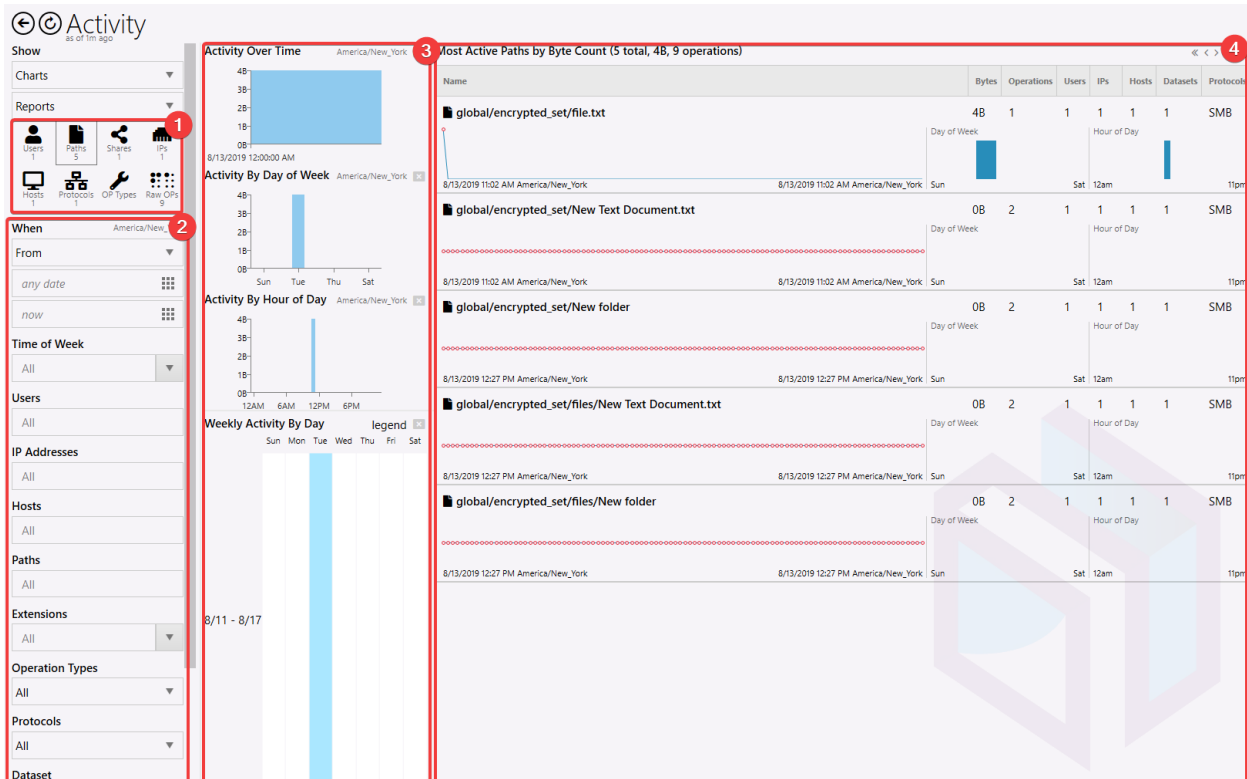
The User Behavior enables administrators, security and compliance officers to log the operations to each file made by applications and users such as file creation, movement, deletion, etc. It must be enabled in the settings tab of the pool or dataset before this will be viewable on the pool.

Viewing the User Behavior Audit

Once enabled, a new 'User Behavior' tab will be accessible from the pool/dataset screen. This tab is an overview of all the recent actions taken.

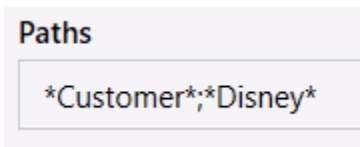


Most of the content here can be clicked on and will lead to the 'Activity' page.

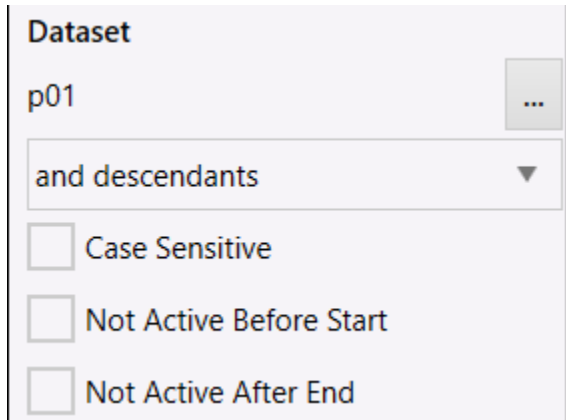


1. Activity categories can be selected.
2. List of filters that can be applied to specify which audit entries are shown.
3. Charts and graphs displaying activity information.
4. Here is where all the relevant entries will show up, depending upon the selected categories and filters.

Filter terms can be joined by using a semicolon (;) and an asterisk can be used as wildcard. For example, you can search for paths or file names with “Customer” OR “Disney” in the name with this entry:



At the bottom of the Search Bar you can change the scope by pool or dataset and time. You can also enable case sensitivity.



Dataset

p01

and descendants

Case Sensitive

Not Active Before Start

Not Active After End

Forwarding User Behavior

The user behavior activity can be forwarded to a SIEM or log centralization for off system processing and analysis. To configure UBA to forward to another host edit the configuration file in `/etc/racktop/ubcollectd/ubcollectd.conf`

[Syslog]

Protocol = "udp"

Server = "10.1.29.X:514"

CertFile = ""

Facility = "local0"

Enabled = true

Sharing

Sharing from the dataset level is where the admins configure the share protocol and in the case of SMB and AFP the share name for the dataset.

If the dataset is already enabled to share data via one protocol it will hide the other share types under the Show More Share Types hyperlink. To enable another file protocol click that first and you will be presented with more options.

SMB

For SMB shares you have the option to enable the dataset to be shared out as a top-level SMB Share. If you enable Access Based Enumeration (ABE) the system hides the share from anyone browsing via SMB who doesn't have read access to that share. Host Base Access control further restricts access by source IP.

[Show More Share Types](#)

SMB Share

Connect Using

Windows icon \\192.255.1.240\manual

On manual

Hide from users that don't have permission (ABE)

Host based access control
Example: @1.2.3.*; @1.2.3.4/24; *.foo.com

Read-only

Read/Write

Deny

AFP

Apple File Protocol uses Posix permissions and like SMB can use a different share name than the dataset name. To enable it as a time machine target you must enable it through myRack Manager first.

AFP supports host base access controls like SMB.

AFP is not supported on HA Resource Groups.

NFS

BrickStor supports NFSv3 and NFSv4.0/4.1/4.2. NFS 4 and above supports ACLs while the NFS v3 standard only supports host based access control and POSIX permissions. NFS shares must be the same name as the dataset and share the path of the dataset starting with /storage and then the pool name.

NFS Share

Connect Using
🔗 192.255.1.240:/storage/p01/global/manual

On
Control access by specifying IP and hostname criteria below.
Example: @1.2.3.*; @1.2.3.4/24; *.foo.com

Read-only

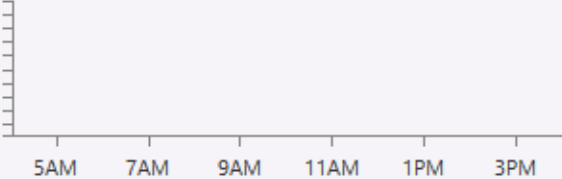
Read/Write

Full Control (Root)

Deny

Security Mode

Hide descendant datasets
 Data security labels

NFS Read/Write Volume


With NFS v4.2 clients BrickStor will support context security labels when the Data Security labels box is selected.

Clicking on the NFS Read/Write Volume will take you to performance metrics related to NFS and the dataset.

Permissions

The permissions menu allows admins to review and modify permissions for the selected dataset.

The screenshot shows the 'File System Permissions' interface. It features a title bar with a plus icon. Below the title, there are five rows of permission entries, each with a dropdown menu for the permission type and a dropdown for the user/group. The entries are: 1. Full Control (Owner), 2. List Folder Contents (Everyone), 3. Read/Write (wingroup:Administrators@BUILTIN), 4. Read/Write (winuser:bvillain@racktoplabs.com), and 5. Read/Write (unixuser:winadm). At the bottom, there are three buttons: 'Add Permission', 'Copy From', and 'Reset'. Below the buttons are two checkboxes: 'Recursively: Apply Reset Ownership'.

When joined to Active Directory or LDAP you can use AD user names and groups.

Group AD Permission

wingroup:<group@domain>

Single AD User Permission

winuser:<userid@domain>

Local BrickStor User Account

unixuser:<username>

You can recursively apply permissions to a dataset and its descendants and reset ownership by selecting the appropriate check boxes.

This is a close-up of the permission type dropdown menu. The dropdown is currently set to 'Read/Write' with the user 'winuser:bvillain@racktoplabs.com'. A red box highlights the text 'Click to Modify Permission Type' with a red arrow pointing to the dropdown arrow. The expanded menu shows the following options: Read (radio button), Read/Write (radio button, selected), Full Control (radio button), List Folder Contents (radio button), Traverse Folder (radio button), Deny (radio button), Deny Modify (radio button), Custom (radio button), Remove (trash icon), and Move Up (up arrow icon).

Copy Permissions from Another Dataset

Admins can copy the permissions of another dataset to the selected data set with the Copy From button. This feature will allow you to copy the permissions of any dataset on any appliance you are currently logged into.

Auto Snapshot Data Protection

This tab allows the user to change the data protection policy from the default storage protection profile to a custom data protection policy specifically for that dataset. It is also the menu in which the admin can chose the replication target(s) for the dataset based on available peers. Replication targets can be inherited so if this is a dataset where global has already been defined with a replication target or in the case where it is a subdirectory with a defined replication target this will show up in the Auto Snapshot Replication portion of the tab.

The screenshot shows the myRack Manager interface for a dataset named 'bsr-c52a9e6a (10.1.12.136) / p01 / global / property-test'. The interface is divided into several sections:

- General:** Shows 'Shared - Non-Reserved Pool Free Space' with '8.2GB' available.
- Auto Snapshot Creation:**
 - next run: 1:22p, last run: 9:22a, last status: skipped.
 - Use profile protection policy (dropdown menu).
 - Automatic snapshot frequency and retention has been optimized for the selected storage profile. To edit these settings, choose custom.
 - Frequency: Every 4 hour(s)
 - Retention: 5 day(s)
 - Daily consolidation: no day(s)
 - Weekly: 4 week(s)
 - Monthly: 12 month(s)
 - Yearly: no year(s)
- Auto Snapshot Replication:**
 - Off (Automatic replication has been disabled).
 - Replication Targets: No automatic replication targets.
 - Buttons: Add Target, Remove All.
- Recent Data Restoration:** Pending (0), Failed (0), Succeeded (0).

When replication is enabled the screen will look like the below screen capture. By clicking on the [SNAPS] button under replication targets the GUI will take you to the snapshots page on the target node as long as you have network access from where the myRack manager GUI is running.

You can remove targets and replace targets from this screen.

Clicking on the target will take you to a window that will show you all datasets on the appliance that are replicating to that target.

Auto Snapshot Creation log

next run 1:30p
last run 9:30a
last status skipped

Use profile protection policy log

Automatic snapshot frequency and retention has been optimized for the selected storage profile. To edit these settings, choose custom.

On

Frequency	Retention
Every 4 hour(s) ▼	5 day(s) +
Daily consolidation −	no day(s) +
Weekly −	5 week(s) +
Monthly −	12 month(s) +
Yearly −	no year(s) +

These settings only apply to new snapshots. Existing snapshots will expire based on the settings at the time of snapshot creation. Sub-daily snapshots will be skipped when no change occurs.

Auto Replicated Snapshots

Have same retention ▼

Auto Snapshot Replication log

On log

Automatic replication has been enabled for self and descendants.

Normal priority on all targets ▼

Replication Targets log

Automatic replication targets have been configured for self and descendants.

poolDR on 10.1.19.2 [SNAPS] [LOG] 🗑

common snapshot: @8/25/2019 1:30 AM

Add Target Remove All

Recent Data Restoration log

Pending (0) Failed (0) Succeeded (0)

MS

← bsr-c52a9e6a (10.1.12.136) / p01 / glo

Shared - Non-Reserved Pool Free Space
8.2GB

General show advanced

User Behavior

Sharing

Permissions

Auto Snapshot Data Protection
Creation: Enabled (storage profile)
Replication: Disabled (no targets)

Settings

Storage Utilization
8.2GB free data of 8.2GB

Limits

Quota X

Limit space consumed by data, snapshots, reservations and descendants.

Quota (Data) X

Limit space consumed by data. The data quota does not limit space used by snapshots and descendants.

Reservations

Reservation

Shared - Non-Reserved Pool Free Space
8.2GB

0B

Reserve space for data, snapshots and descendants. Space already reserved for data or by descendants is counted against the reservation.

Reservation (Data)

0B

Reserves space for data. The data reservation does not reserve space for snapshots and descendants.

Settings

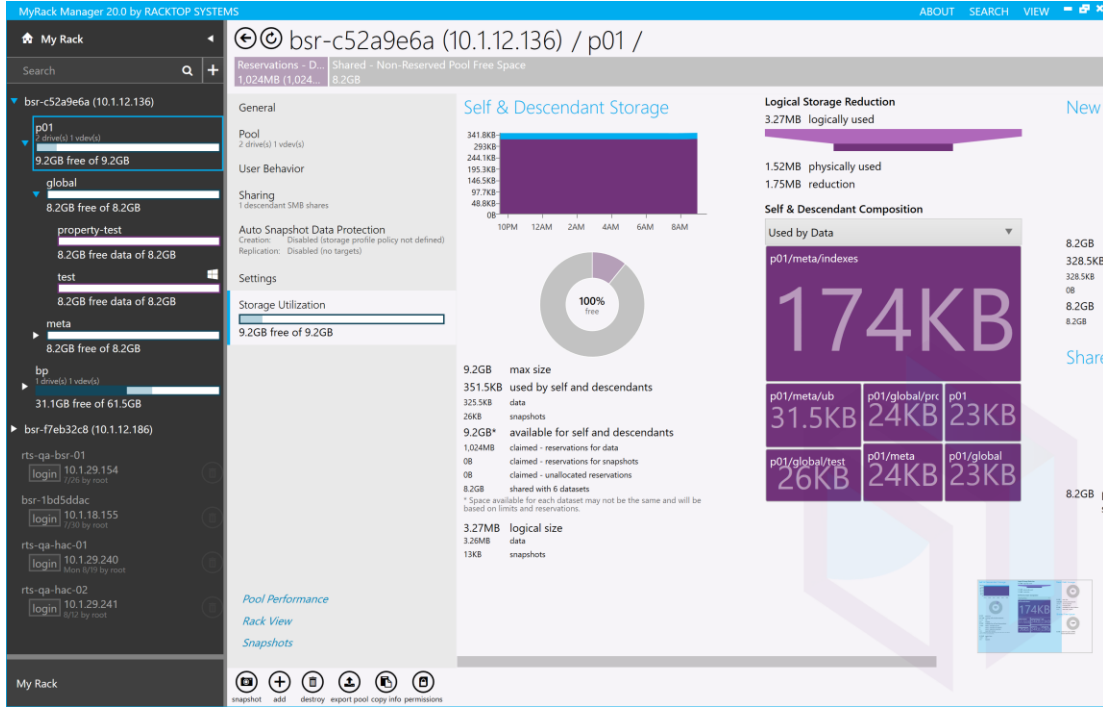
The settings tab when a dataset is selected allows you to configure quotas and reservations. You can quota only the data or you can quota the data with snapshots and descendants. You can also set reservations on the dataset here for both instead of thinly provisioning the dataset.

You can type a number and scale such as MB, GB, TB or you can use the slider above the text box to set the quota or reservation.

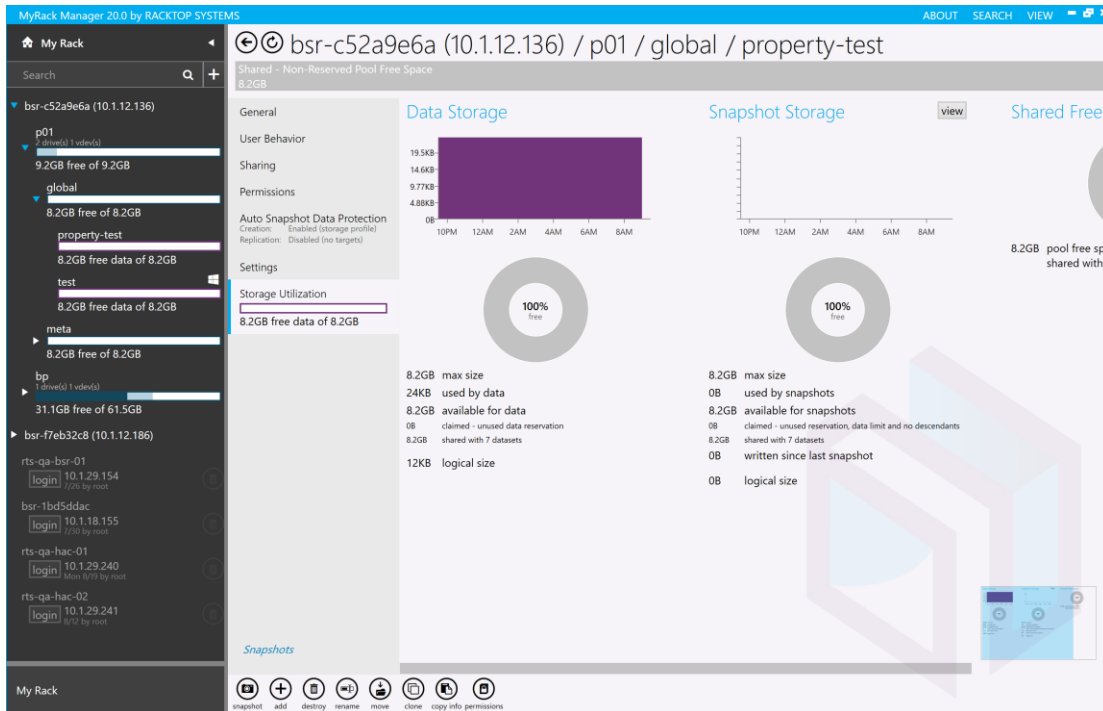
Storage Utilization

This tab shows information about the physical storage taken up by the pool or dataset.

Pool:



Dataset:



Pool and Dataset Links

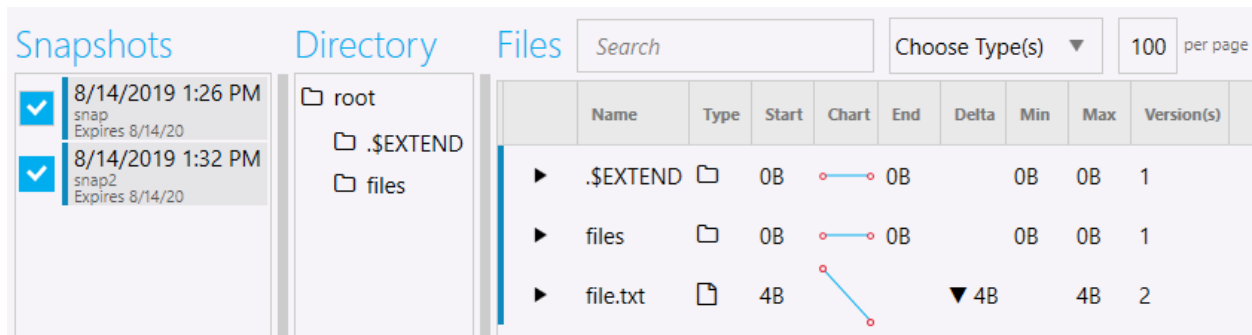
Snapshots

Snapshot Indexing

After snapshots have been created of a dataset, they can be accessed and viewed. After selecting a dataset from the My Rack panel, click on the 'Snapshots' button near the bottom.

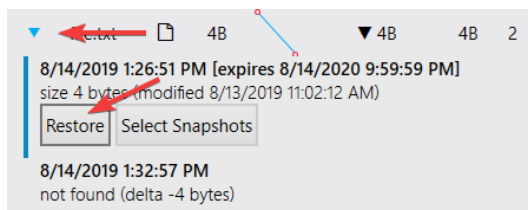


This brings up the snapshots screen for the selected dataset. At the top, filters can be set to only view snapshots from a certain time range. On the left, all the snapshots that match the filter parameters will be listed. Each one can be toggled on and off, and all the files present in the selected snapshots will be displayed in the panel on the right. Each file has a chart associated with it that shows its size over the course of the selected snapshots.

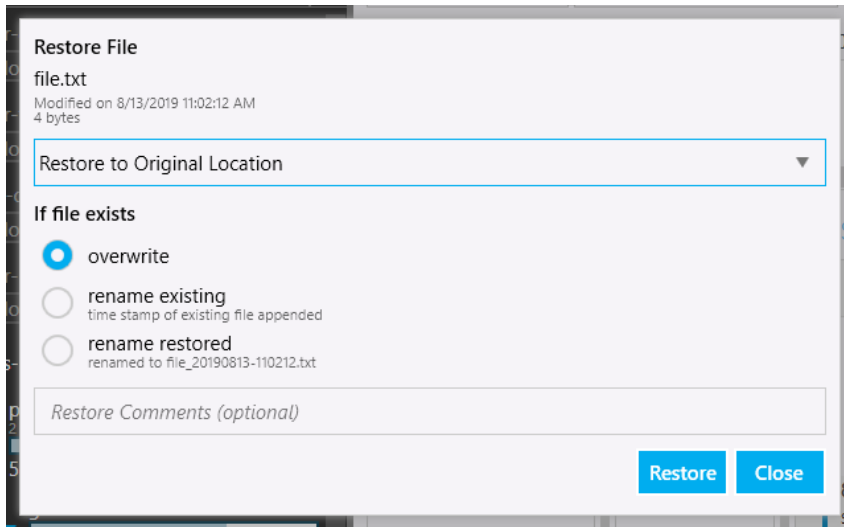


File Restoration

From the snapshots page, any item in any snapshot can be restored. To do this, click on the dropdown arrow on an item in the snapshot, and select 'Restore.'

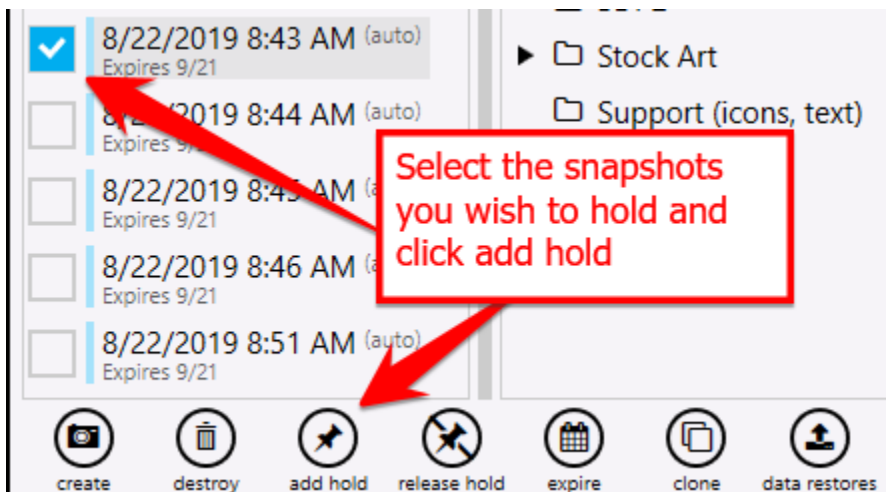


In the dialog box that shows up, choose whether the restored file should overwrite any existing file, rename the existing, or rename the restored file. Select 'Restore' to complete the action.



Snapshot Holds

It is sometimes necessary to hold snapshots past the normal expiration period. They can be assigned a tag that will be used to report on and enable an admin to remove all holds across all datasets on the appliance with that hold tag. You can also set an expiration on the hold tag itself. No snapshot will be removed from the dataset if there is a hold tag applied.

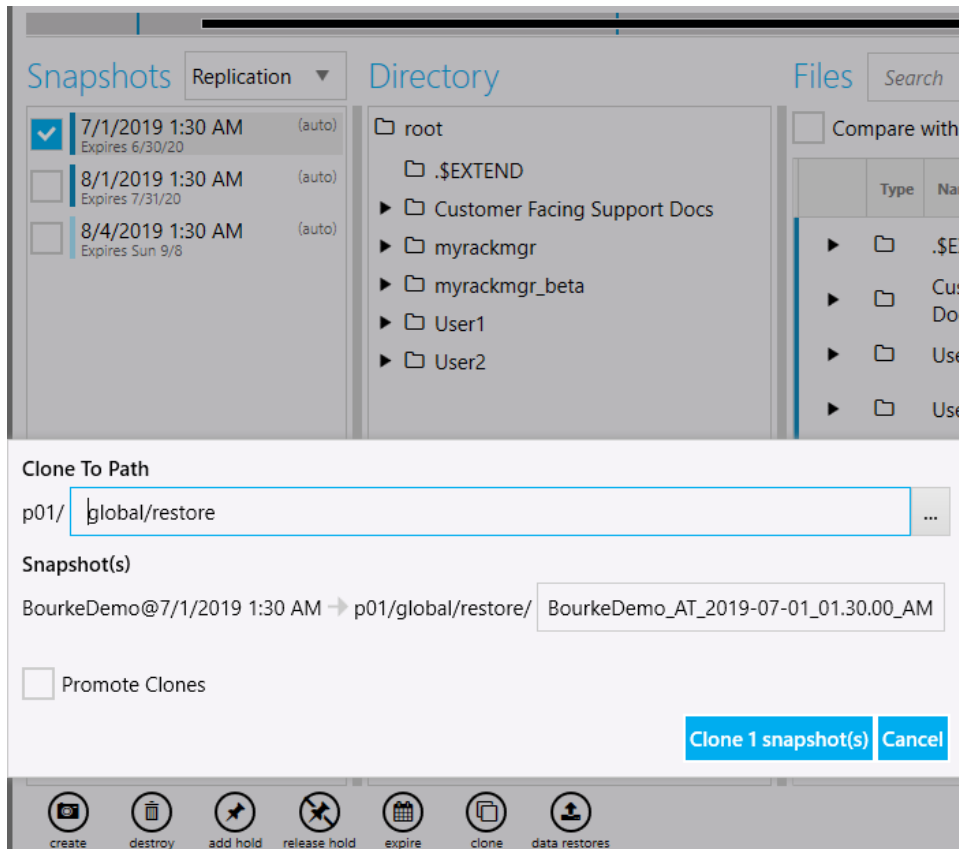


To release a hold tag you can just click release hold on the appropriate data sets.

If you delete a dataset you will delete the snapshots with it. If there are snapshots with a hold tag in the dataset pending destruction it will ask you to remove and release the holds before it can proceed destroying the dataset.

Clones

You can select a snapshot to clone which will create a writeable version of the snapshot without modifying the snapshot. Only changes to the clone will take additional capacity on disk. You can choose the path to create the clone. It must be on the same pool as the snapshot. Clones are the way to retrieve a file or files out of the snapshot on a replica because they are not mounted.



Be careful when promoting a clone. You should only promote a clone when you want all the snapshots prior to the snapshot to be linked to the clone and not the original active dataset. This operation is not reversible. It may also break replication if done improperly and you lose the common snapshot between the original and the replica.

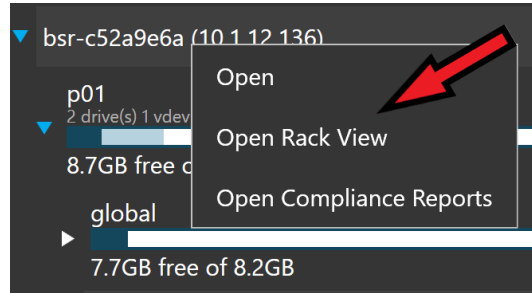
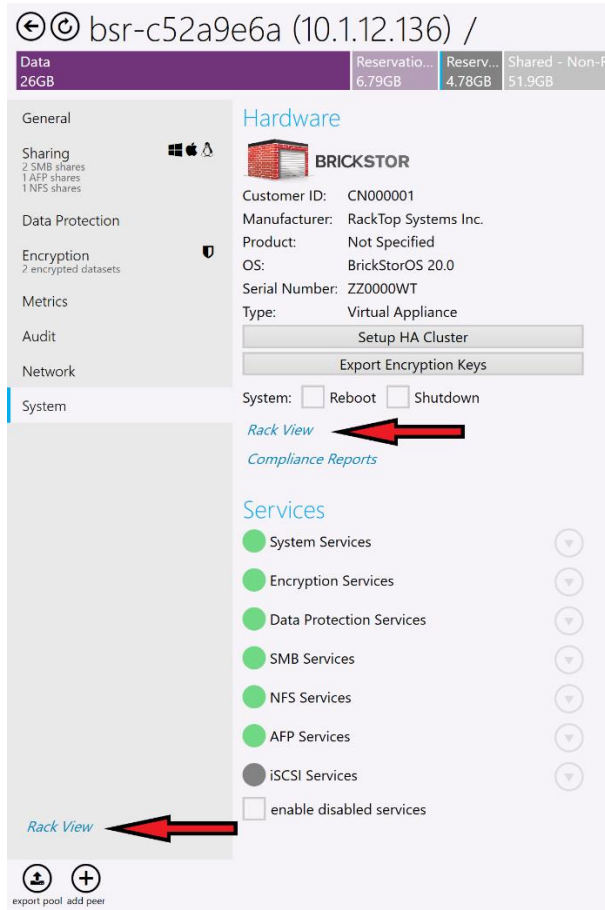
Clones are a rapid way to create an entire dataset based on a point in time. This is a common method used to recover from a ransomware attack. They can also be used to create a version of a dataset to test an upgrade or run destructive tests and analysis against data without affecting the golden copy of data.

Rack View

myRack Manager features the capability to easily view and modify your appliance hardware called Rack View. Rack View allows users to add or modify pools and vdevs and gives visuals that allow users to see what changes will occur to the system's hardware prior to committing them.

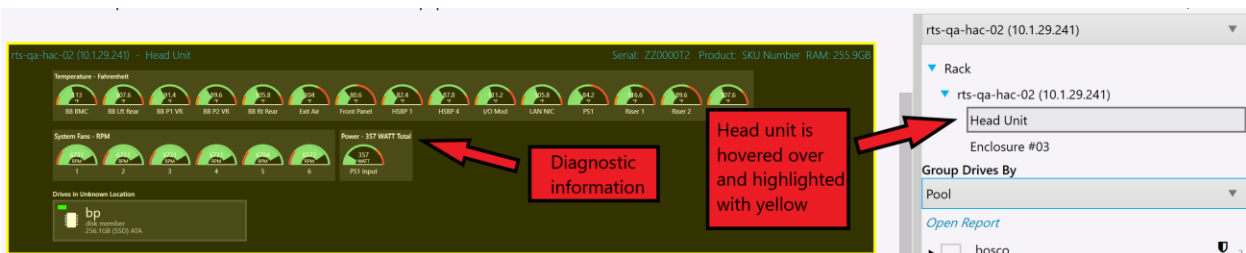
Accessing Rack View

To access Rack View, simply click the Rack View button under the Hardware section of a system or right click the system from the My Rack tab and select Open Hardware.

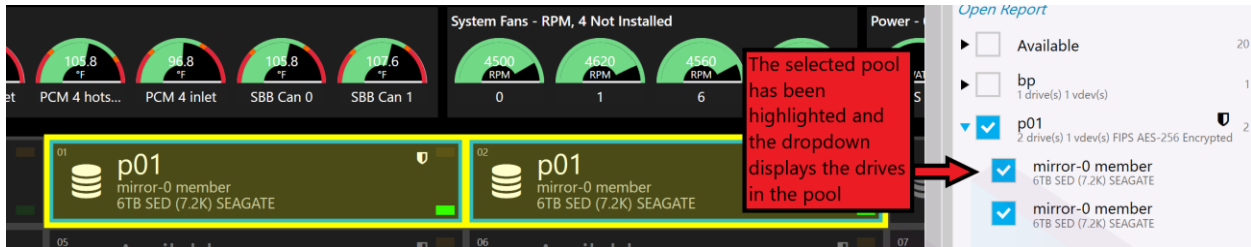


The Rack View Interface

Rack View will display the current hardware that is on your system including the head unit, JBODs, and any drives that are within these appliances. It will also display various diagnostic information such as the values of temperature sensors in the system and the fan speeds. On the upper right-hand side, you can select which appliance you want to zoom to. The appliance will be highlighted in yellow when the mouse is hovered over it and left clicking will zoom to the appliance.



The right-hand side of Rack View also allows you to group the drives in the appliances based on certain properties such as pool, make, and vdev type. To change the grouping type, select the dropdown under Group Drives By and then select how you want to group them. When hovering over one of these groups, affiliated drives will be highlighted and left clicking will zoom to the drives. You can also expand these groups with the arrow and select individual drives that are a part of the group.



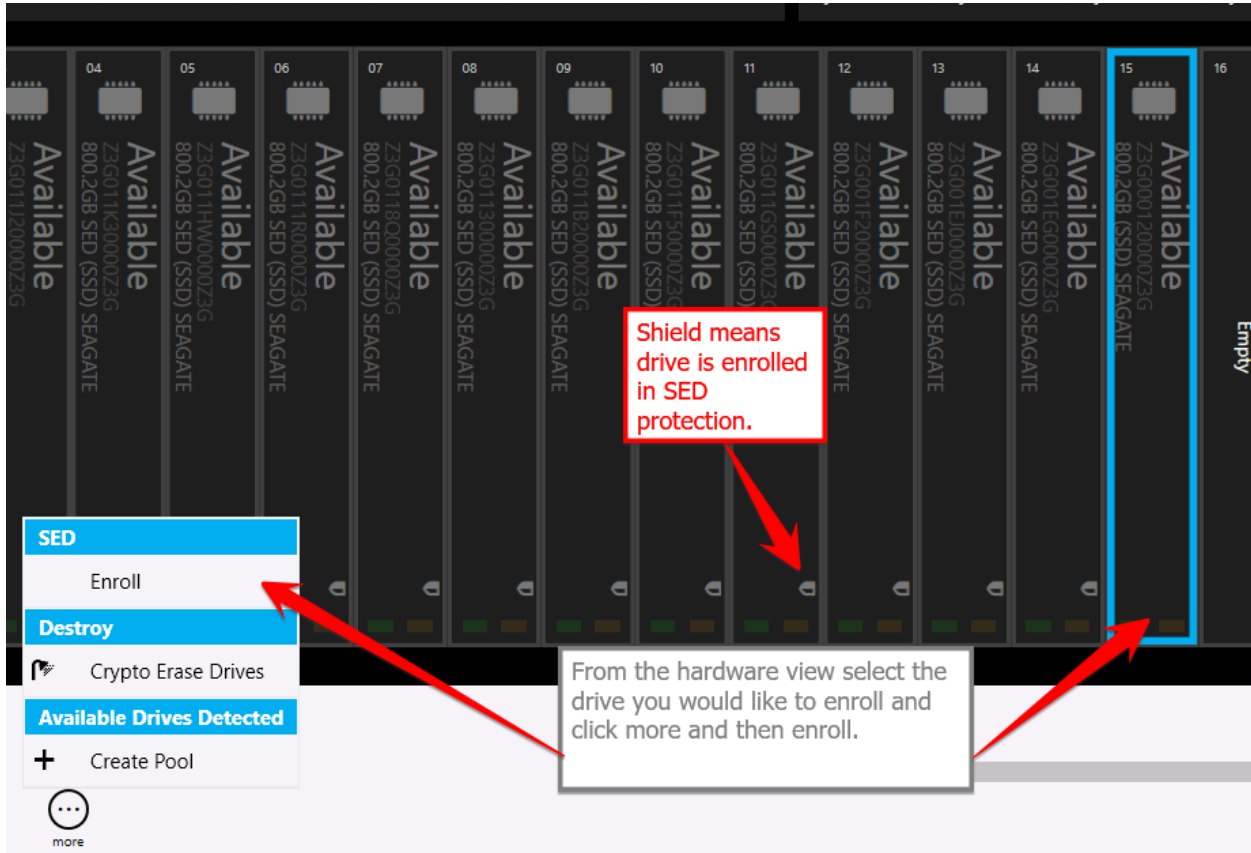
Self-Encrypting Drive Management

BrickStor can leverage TCG FIPS 140-2 certified self-encrypting drives for increased security. To manage the keys and disks within BrickStorOS does require a special license from RackTop and appropriate FIPS drives. TCG licensed systems come with drives encrypted using a factory generated key. Self-Encrypting Drives placed in a system that are not licensed will not lock when power is removed.

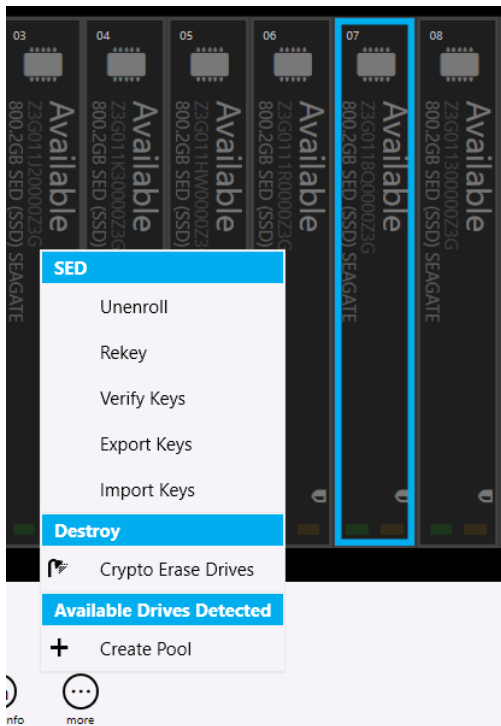
TCG Must be licensed and the Key Manager must be properly configured before you can

Drive Enrollment

Once the key manager is configured drives can be enrolled in the system. Each drive will receive a unique key used to unlock the self-encrypting drive known as the key encryption key (KEK) from the key manager and configure the drive to auto lock when power is removed from the drive. To enroll drives or a pool in the system go to the hardware view page of the UI. If you select a drive that is not in a pool you can select multiple drives and enroll the ones you choose to enroll. If you select a drive that is already a member of a pool it will enroll all drives that are a member of that pool.



Other Self Encrypting Drive Operations



Unenroll – Removes drive from SED management and sets the drive to default PIN and sets the drive to stay unlocked.

Rekey – Requests a new key from the key manager and changes the KEK PIN on the drive.

Verify Key – Verify the KEK unlocks the drive and is available from the key management service

Export Keys – Will provide a password protected file with the KEK PINS that can be imported later for backup purposes or to another node so that the other node can unlock the drives. This is required in HA using the internal key management service.

Import Keys – Allows you to import keys that were exported from the same node or another node into the internal key management database. This is performed for HA nodes to share keys between the heads. This can also be used

import keys to a replacement head node.

See the BrickStor SED Usage guide for more details related to Self-Encrypting Systems.

Exporting and Backing Up Keys

When using the BrickStor internal key manager it is important to back up the keys and store them in an alternate location.

The `/etc/racktop/keymgrd.conf` file allows users to set the location of the internal key file.

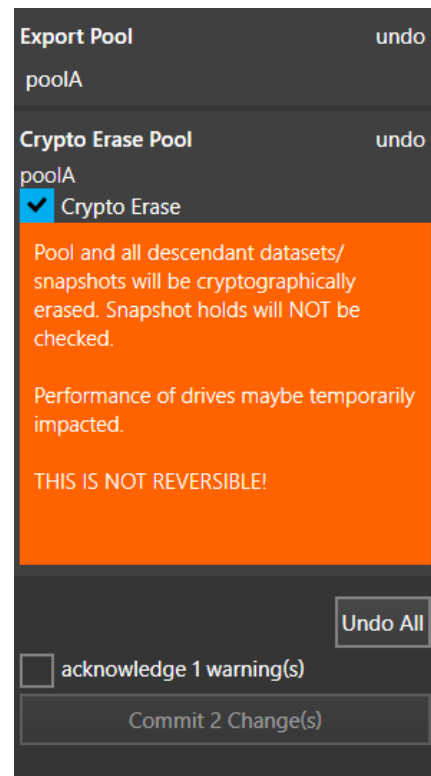
The configuration file also allows users to configure the BrickStor to rotate KEKs on a scheduled interval. This is only recommended when using an external key manager in order to ensure you have backup copies of the keys.

Cryptographically Erasing SEDs

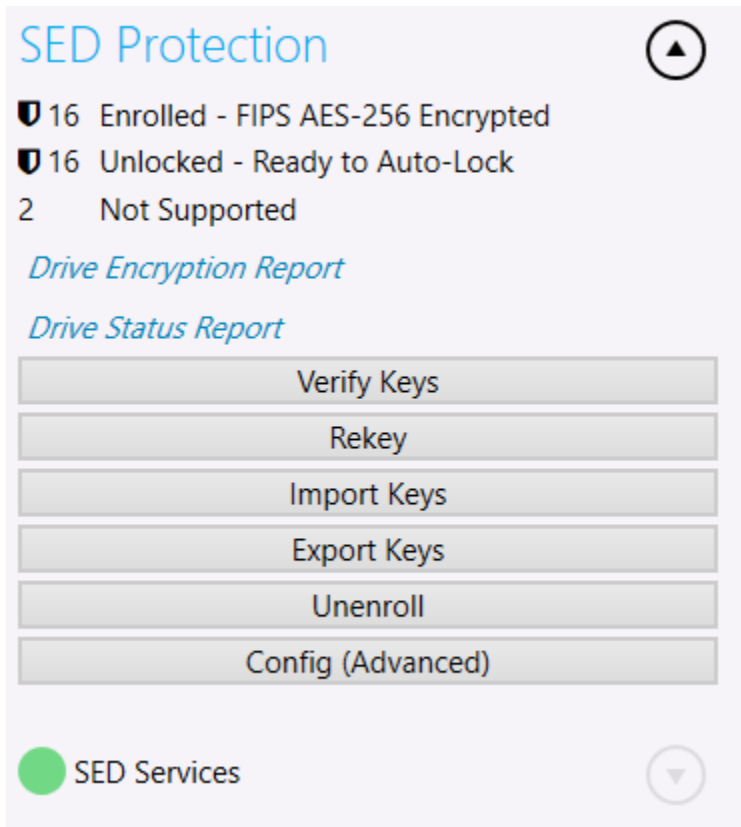
Users can Crypto Erase SEDs which will reset the pins and put them in an unenrolled state. To manage the drive again just enroll the drive.

As part of a pool destroy users can select the crypto erase option. This Option is irreversible. Data is permanently destroyed and unrecoverable. However, if you don't select the crypto erase option the data is potentially recoverable in the future off each drive.

If the KEK PIN has been lost for a drive a crypto erase is the only option to put the drive back into a usable state because the drive will become erased and unlocked.



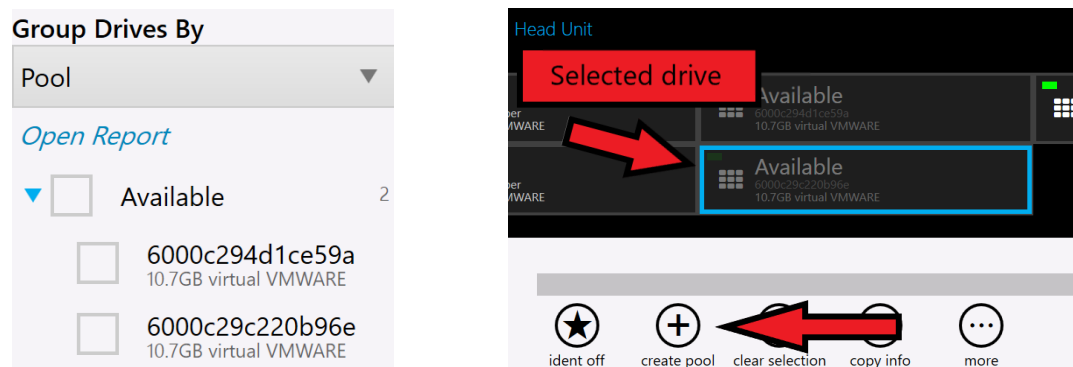
SED Protection on the Main Pane



Under the general tab of myRack Manager users can perform various SED configuration options as well review reports about which drives are enrolled in SED management and the current status of each drive.

Creating a Pool within the Rack View

To create a pool, first select an available drive by left clicking a drive labeled as Available or by selecting a drive from the right-hand dropdown of 'Available' when sorted by Pool. The selected drive will have a blue border and the icon create pool will appear at the bottom of the screen.



Clicking create pool will open the Create Pool dialog box where you must set a pool name and can change the type of vdev, the number of vdevs, how many drives are in each vdev, as well as how many

spares you want the pool to have. By default, it will choose drives from alternating enclosures, but you can uncheck this box to select specific drives for the pool. When everything is configured, click create to queue the changes.

Create Pool

p01 X

Type

mirror

Auto choose drives from alternating enclosures

Drive Type

6TB SED (7.2K) SEAGATE

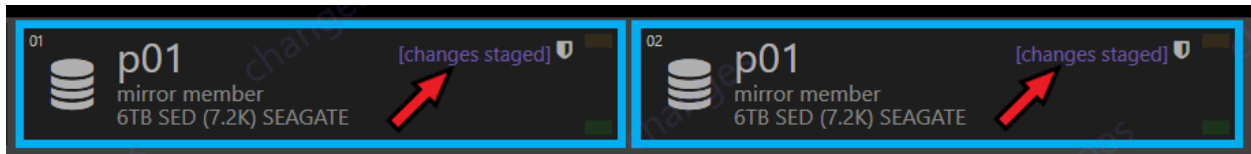
- 1 + vdevs

- 2 + drives per vdev

- 0 + spare drives

Create Cancel

Rack View will display the queued changes and any pool that will be affected by changes will have the [changes staged] indicator on it.

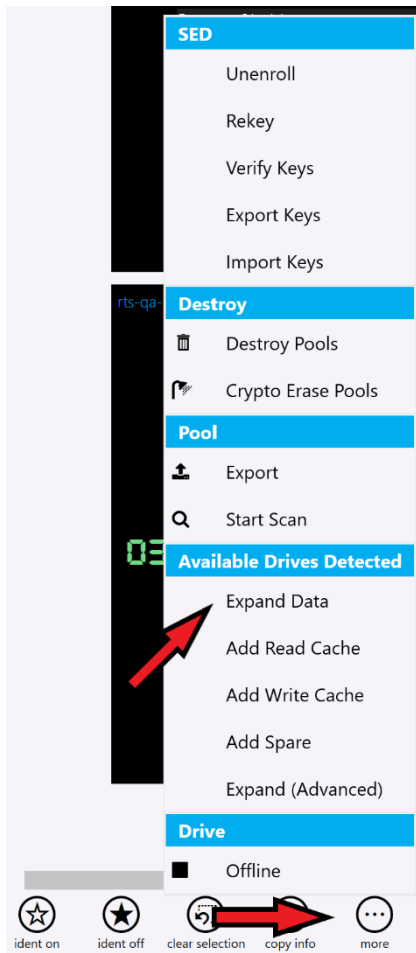


To finalize the creation of the pool, go to the Changes tab on the left-hand side of myRack Manager and click Commit Change(s).

Modifying an Existing Pool

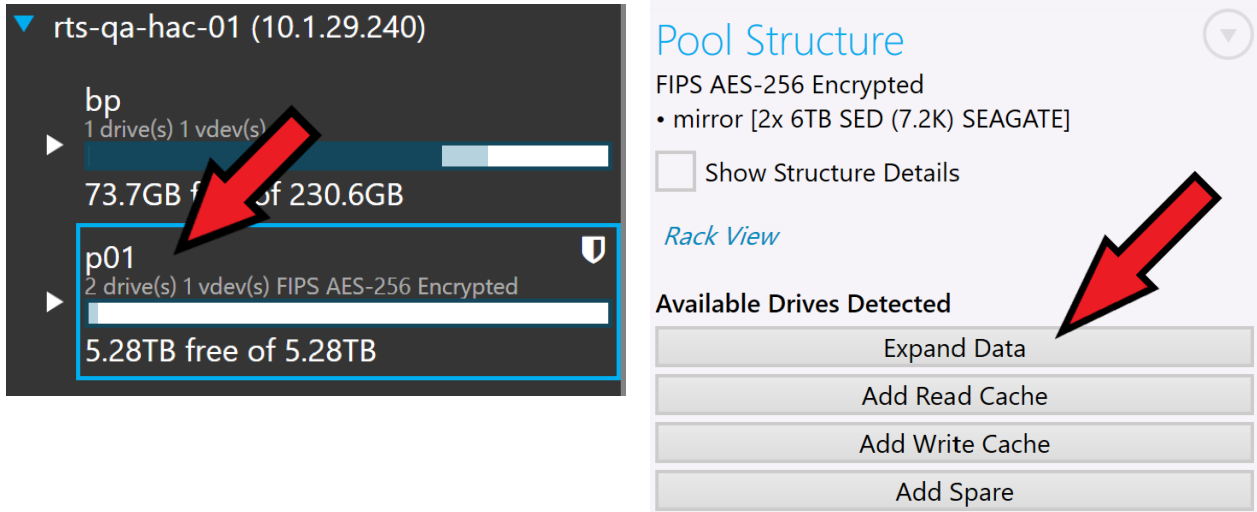
myRack Manager features several ways to modify pools that are currently on the system.

Expanding a Pool

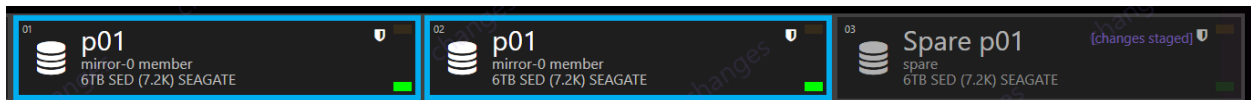
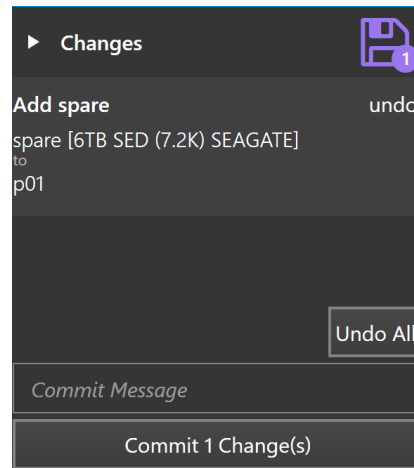
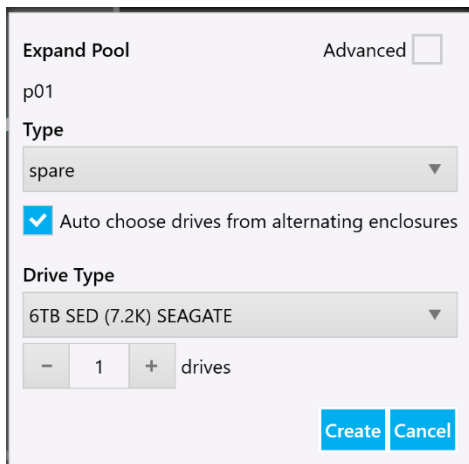


There are multiple ways to expand a pool. The first is to select the pool in Rack View, select 'more' from the bottom bar, and then click any of the available expansion options.

The second option is to select the pool from the My Rack tab on the left-hand side of myRack Manager and click either the Expand Data, Add Read Cache, Add Write Cache, or Add Spare button under the Pool heading, depending on what you would like to add to expand the pool (will only appear if the correct types of drives are available).



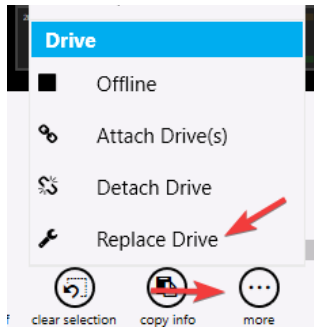
This will bring up the Expand Pool dialog box where you can choose to expand the pool by adding more vdevs, read and write caches, or spares. When the desired settings have been configured, click create to queue the change.



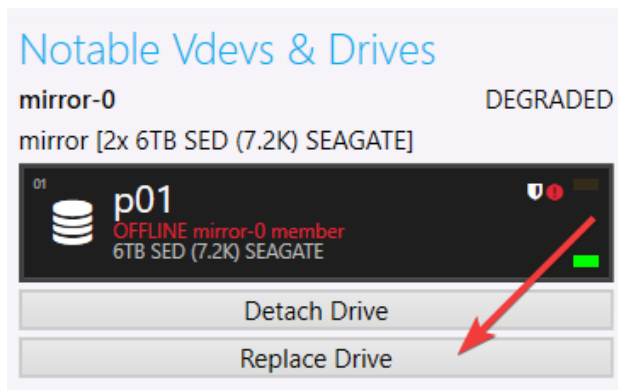
All changes in the queue will be indicated in Rack View and must be committed using the changes tab on the left-hand side of myRack Manager.

Replacing a Drive

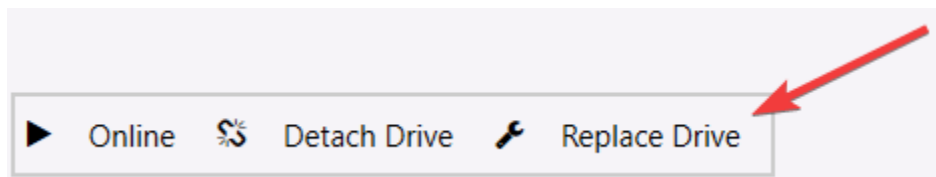
If a drive becomes disabled or faulted it may be necessary to replace the drive with another available drive in the system. Select the drive you wish to replace in Rack View, click 'more,' and click 'Replace Drive'.



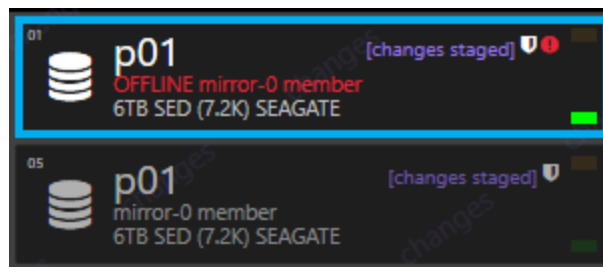
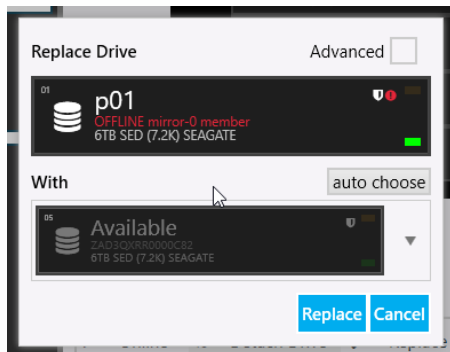
Or, if the drive is offline, you can navigate to the degraded pool in the My Rack tab on the left-hand side of the screen and click the Replace Drive button under the 'Notable Vdevs & Drives' heading.



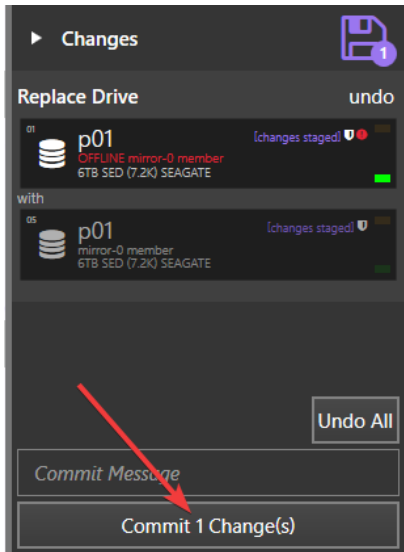
Selecting an offline drive from Rack View will also bring up actions that can be performed on it.



This will bring up the Replace Drive dialog box where you can select the drive to use as the replacement then click the Replace button to queue the change.

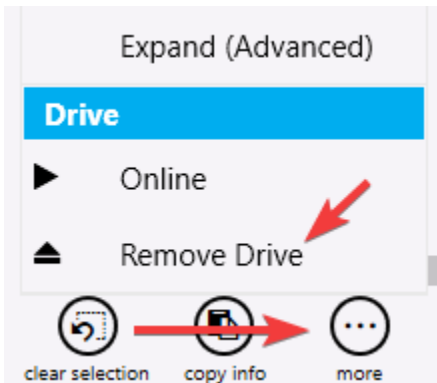


The change will be indicated in Rack View and will not be committed until the Commit Changes button is clicked on the Changes tab.

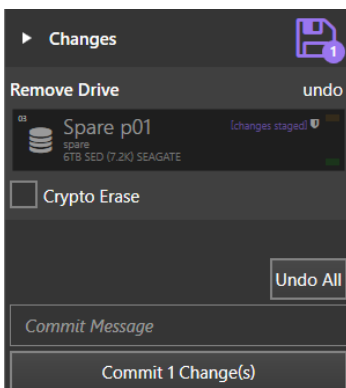


Removing a Spare

If a pool has a spare drive that no longer requires one, it can be removed to free up the drive by selecting the spare in the Rack View, selecting 'more,' and clicking the 'Remove Drive' button.

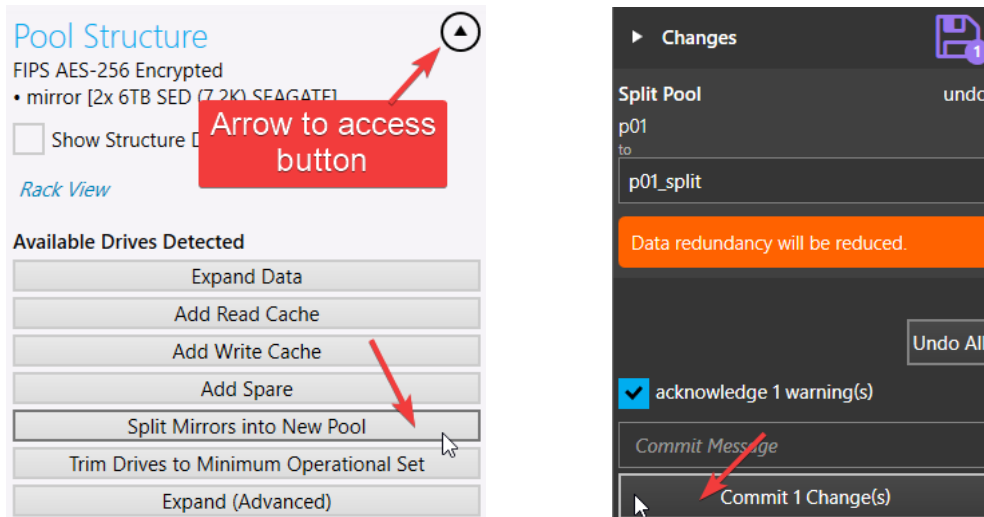


The change will be indicated in Rack View and will not be committed until you click the Commit Changes button in the Changes tab on the left-hand side.



Splitting a Mirrored Pool

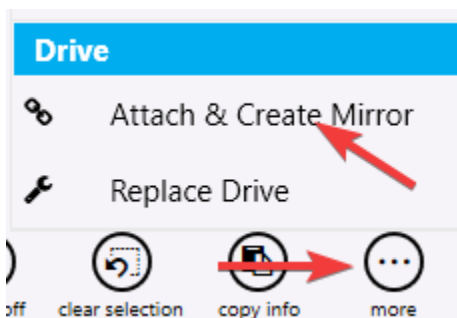
A pool consisting of mirror vdevs can be split into two pools with no redundancy that contain the same data. **Note that this is only recommended in certain scenarios as the lack of redundancy increases the risk of data loss.** To split a mirrored pool, navigate to the pool from the My Rack tab on the left-hand side and click the Split Mirrors into New Pool button under the Pool heading (you will need to click the arrow button to the right of the Pool heading to access this).



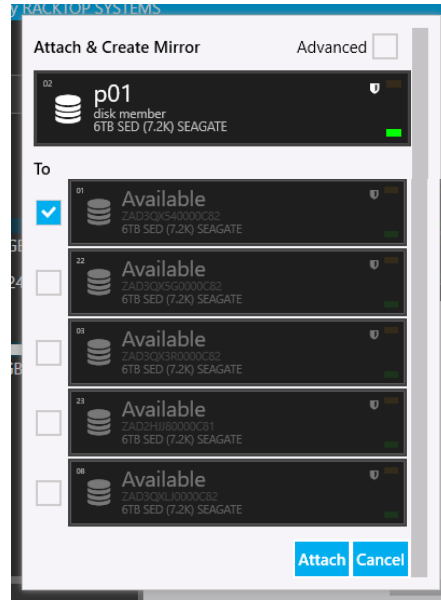
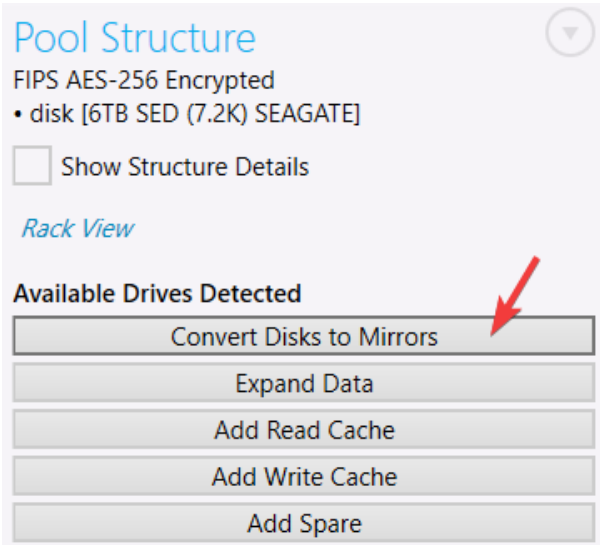
From the changes tab on the right-hand side you can change the name of the new pool that will result from the split and commit the changes with the Commit Changes button (by default the new pool created this way will be exported).

Attaching a Drive to a Pool

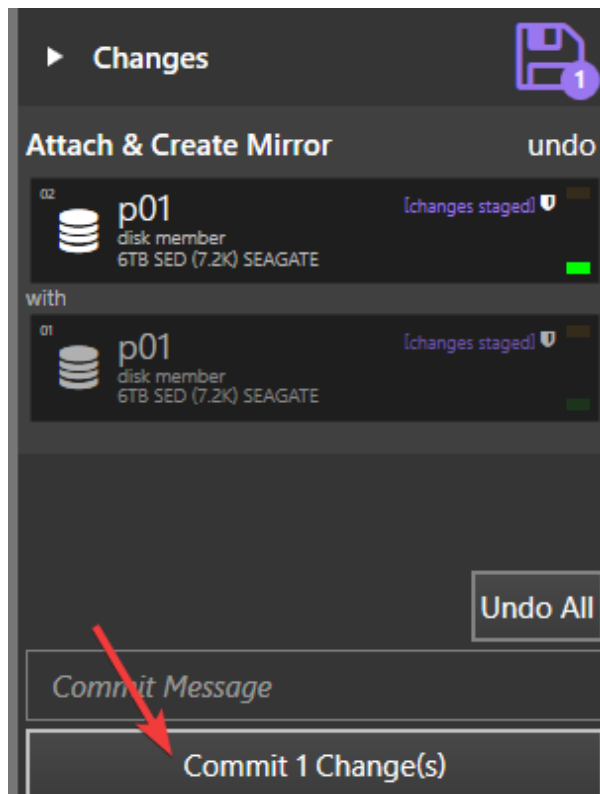
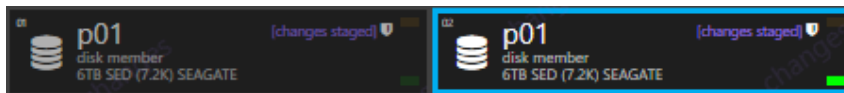
A pool with no redundancy can be converted to a mirrored pool, if there are enough available drives, in order to reduce the risk of data loss. To do this, select the pool in Rack View, select 'more', and click the 'Attach & Create Mirror' button,



Or navigate to the pool from the My Rack tab on the left-hand side and click the Convert Disks to Mirrors button under the Pool heading.



If done through Rack View, you will need to select the drive to attach yourself. When done through the pool's page it will select a drive for you automatically. The change will be indicated in Rack View and will not be committed until you click the Commit Changes button in the Changes tab on the right-hand side.

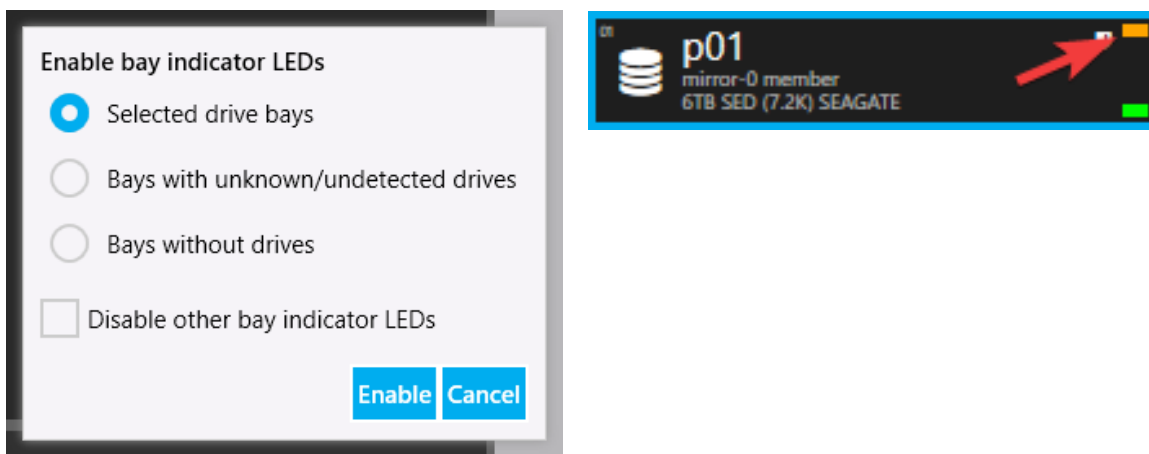


Toggle Identifying Lights

Rack View allows you to toggle a physical indicating light on each drive to assist with identifying the correct drives on the machine. You can either select one drive by clicking directly on it in Rack View, or multiple drives using the Group Drives By interface on the right-hand side. Once the appropriate drives have been selected click the identify button at the bottom of the screen.



This will bring up the Enable bay indicator LEDs dialog box, where you can turn on the lights for either the selected bays, bays with unknown drives, or bays without drives. You can also choose to disable all other indicator lights to ensure only the desired drives have their lights enabled.

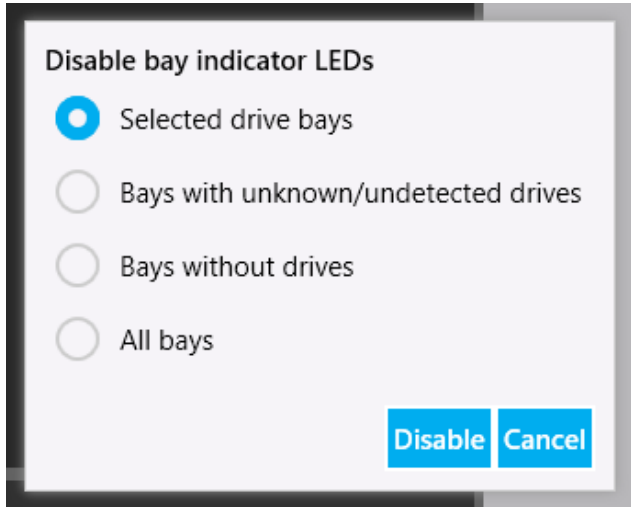


Drives with their indicating LEDs enabled will have a blinking orange indicator on Rack View as well as on the physical drive on the appliance.

To disable the identifying lights, select the desired drives like before and click the identify off button.

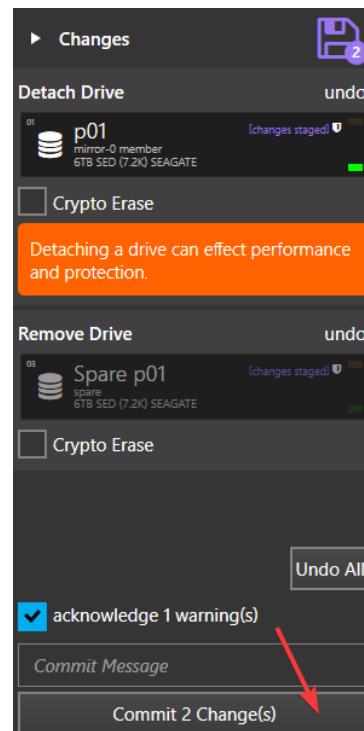
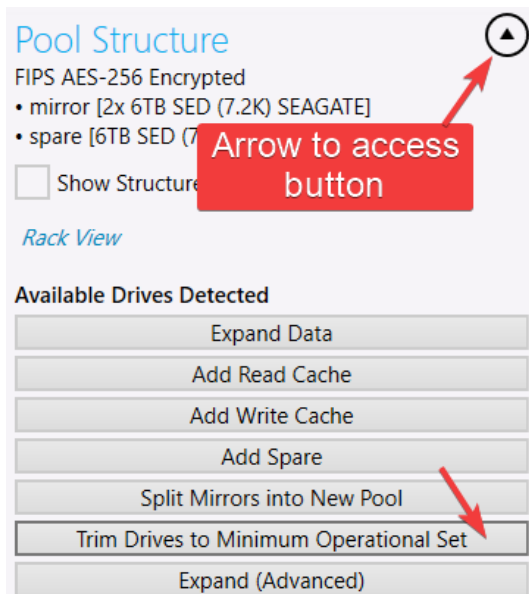


This will bring up the Disable bay indicator LEDs dialog box where you can turn off the lights on either the selected bays, bays with unknown drives, bays without drives, or all bays in general.



Trimming a Pool

If a pool is going to be retired or is no longer necessary and to be removed, it can be trimmed to the minimum operational set of drives. **This will remove all redundancy and additional data protection and should only be done in specific scenarios.** To trim a pool, navigate to the pool from the My Rack tab on the left hand side and click the Trim Drives to Minimum Operational Set button under the Pool heading (you will need to click the arrow button to the right of the Pool heading to access this).

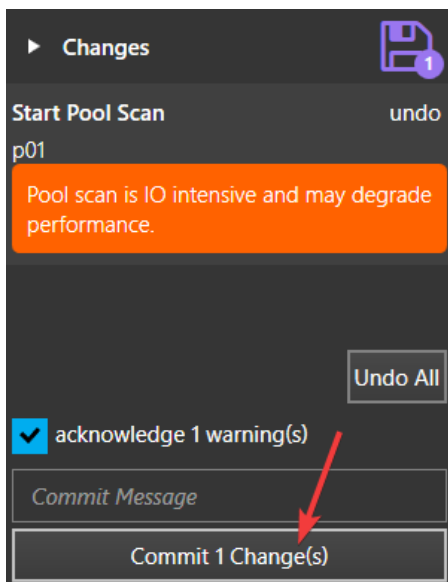
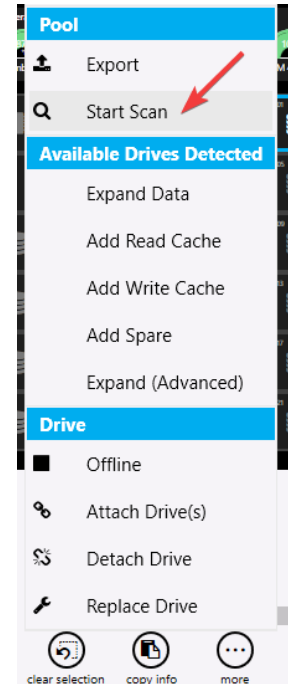
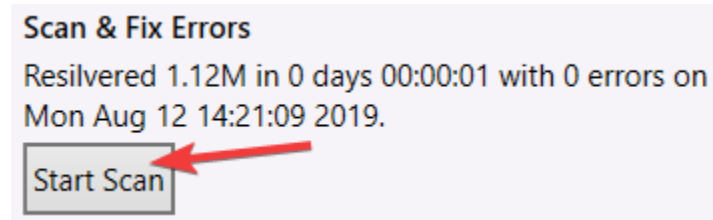


The steps it will take to trim the pool will be listed in the changes tab on the left-hand side and no changes will take effect until the Commit Changes button is clicked.

Scanning and Repairing a Pool

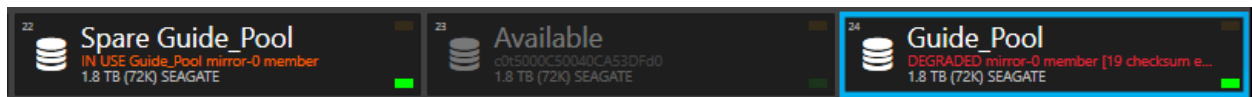
A pool can be checked for faults or problems and corrected using the scan pool feature. To scan a pool for potential faults, either select the pool in Rack View and click the more button at the bottom of the rack view and click Start Scan.

The button is also available on the Pool Tab.



The scan will not be started until you click the Commit Changes button in the Changes tab on the left-hand side.

If the scan detects a faulty drive in the pool, it will mark the drive as degraded and replace it with a spare drive if one is available.



From the pool's screen on the My Rack tab, the faulted drive will appear under Notable Vdevs & Drives. You can choose to promote the spare drive and detach the faulted drive from the pool, replace the faulted drive with another available drive on the system and return the spare to be a spare for the pool, or you can clear the errors on the drive if the problem has been corrected and return the spare. These options can also be found at the bottom of the screen in Rack View.

The screenshot displays the myRack Manager interface. At the top is a toolbar with icons for various actions: 'ident on', 'ident off', 'online', 'offline', 'clear errors', 'attach', 'detach', 'replace', 'expand pool', 'split pool', 'destroy pool', 'scan pool', 'export pool', and 'clear selection'. Red arrows point to the 'clear errors', 'detach', and 'replace' icons. Below the toolbar are two panels. The left panel, titled 'Notable Vdevs & Drives', shows a 'mirror-0' in a 'DEGRADED' state with three 1.8 TB (72K) SEAGATE drives. It lists two drives: 'Guide_Pool' (ID 24) which is 'DEGRADED mirror-0 member [19 checksum error(s)...]' and 'Spare Guide_Pool' (ID 22) which is 'IN USE Guide_Pool mirror-0 member'. Below each drive are buttons for 'Detach Original & Promote Spare', 'Replace Drive & Return Spare', and 'Clear Errors'. The right panel, titled 'Detach Original & Promote Spare', shows a confirmation dialog for the 'Guide_Pool' drive. It contains an orange warning box: 'Spare will become permanent replacement. You may want to add a new spare to maintain the same level of protection.' Below this is an 'Undo All' button, a checked checkbox for 'acknowledge 1 warning(s)', and a 'Commit 1 Change(s)' button. A red arrow points to the 'Commit 1 Change(s)' button.

Each of these changes will require you to click the Commit Changes button in the Changes tab on the left-hand side to complete the action.

High Availability (HA) Cluster Setup and Management

The BrickStor High Availability Cluster consists of four main components

BrickStor Head Nodes (2) – The Nodes act as the storage controllers. They are the components of the cluster that manage the storage and provide network connectivity to the clients. They have connections to both the admin network and the data network. They also have a direct ethernet connection between them called the heartbeat path.

The Witness – The witness is used to act as the third party in the quorum to break a tie for automatic failovers. A witness server is a service application that runs on Windows Server or Linux on a physical host or as a VM that is not using any part of the HA cluster for storage.

Shared Storage – The architecture of BrickStor relies on two controllers with shared storage visible by both Head Nodes. Shared storage is carved into Storage Pools. The Storage Pools are the discrete group of data disks and cache disks. A storage pool can only be managed by one Head Node at a time. Having a storage pool imported on both nodes at the same time can cause data corruption and is why there are many checks to prevent split brain scenarios or a case where a pool is inappropriately imported into two nodes at the same time.

Resource Group – A Resource Group is a logical grouping of storage pool(s) with an IP address bound to a vnic. A resource group may have more than one vnic. During the resource group configuration admins can specify the default interface or a custom interface for each vnic within the resource group. There is no hard limit on the number of vnics allowed per resource group but an unusually large number of vnics may affect failover times because each vnic must be reconstituted on failover. The Resource Group moves between nodes manually or during automatic failover. Clients can connect to that pool using regardless of the node that is managing it by connecting to that IP. Admins can put a resource group into a disabled state. When resource groups are in a disable state all pools are exported and the vnics are disabled meaning those IPs are not advertised on any interface.

Pool Status within an HA cluster

A storage pool can be in one of five states when managing an HA cluster.

1. **Member of a Resource Group** – Pool is part of a resource group and managed by the HA cluster. Pool is imported on one node and will failover to the other node with resource group.
2. **Disabled Member of Resource Group** – The pool is a member of a resource group but purposely exported from the node managing the resource group. Data on the pool is not available until the pool is enabled which will cause the cluster to import the pool into the node with the resource group.
3. **Unmapped Pool** – Pool is a member of the HA cluster, the pool is protected from being imported on more than one node at a time, the Pool is not currently a member of any resource group and the pool is not imported on either node. (A pool can be destroyed in this state and will display as missing)

4. **Removed from Cluster** – Pool is not a member of the HA cluster and is not protected by any HA services. From this state the Pool can be safely destroyed or mapped into the HA cluster.
5. **Missing** – The pool is not seen by either node in the cluster. This can be the result of the drives being physically removed from the cluster without removing the pool from the HA cluster first or if a pool is destroyed before removing it from the HA cluster.

Standard Network Interfaces

The system requires at least one admin vnic interface, a data vnic and a physical heartbeat interface.

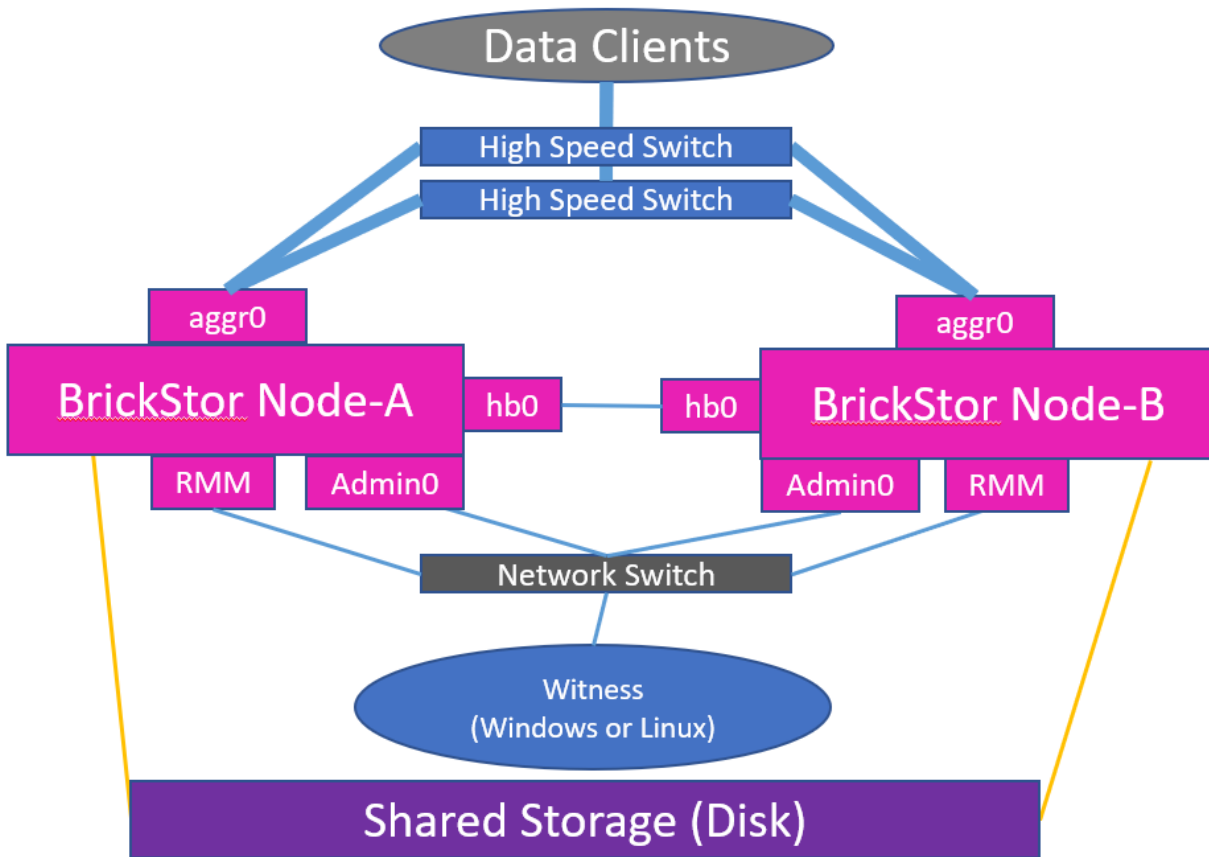
The standard practice is to have an “admin0” interface that is connected via a 1Gb or higher NIC for management functions. A second physical NIC with a cable connected directly to the other node in the cluster is the heartbeat connection and assigned vnic “hb0” with a local only routing IP space. And, an aggregate (“Data Aggregate”) is created for the data vnic’s over a high speed pair of network interfaces.

The IP of the Resource Group is placed on the data aggregate of the node managing the Resource Group. During an automatic or manual failover, the Resource Group vnic is removed from the losing node and added to the gaining node’s data aggregate. The gaining node sends out aggressive ARPs to alert the network of the new IP location.

It is possible to put a second admin interface that is statically located on the data aggregate providing an IP address that would be always available on the data network. This can be implemented when it is desired to have the witness contact the node via the data network.

HA Cluster Architecture

Best practice is to have the witness on the same subnet as admin0 to avoid any obstacles in between the cluster and the witness. The resource group vnics are placed over an aggregate with connections to two different high-speed switches for network redundancy and availability. The RMM is the out of band management port (ILO, iDRAC) that should also be on the same subnet as the admin network.



HA Scenarios

Loss of Network Connectivity

Loss of the admin0 interface will not cause any data availability issues. It will just prevent a user from managing the system through a UI by connecting to admin0. The user could connect to any other network interface that the UI can connect to via the network. Alternatively, the admin could address the problem with admin0 and restore the cluster to a fully healthy state.

Normal Status Checking:

Once the cluster is formed the three participants (Node-1, Node-2, and witness) will check on the health and status of the other nodes. They each use their network connections to connect to the other nodes via any available means. The node will connect over the admin interface to verify it is pingable and active. The two nodes use their direct connect (heartbeat) connection to check health as well. They also use IPMI to check the health and status as a third check. Each node checks to see that it can communicate to the witness. Meanwhile the witness is checking the status of each node through the admin interface.

Loss of Network Connectivity

Loss of the admin0 interface will not cause any data availability issues. It will just prevent a user from managing the system through a UI by connecting to admin0. The user could connect to any other network interface that the UI can connect to via the network. Alternatively, the admin could address the problem with admin0 and restore the cluster to a fully healthy state. Also, the admin can initiate a manual failover by connecting to either node even if the other node is only connected to the healthy node via the heartbeat connection.

Initiating a Manual Failover

The admin can move resources through the GUI by connecting to either node and moving Resource Groups to the desired node. If both nodes are up and operational then the admin can move Resource Groups through the GUI without having the witness online.

Automatic Failovers

If one node goes down the surviving node will automatically take over the Resource Groups as long as the cluster is set to move resources on failure and the witness and healthy node can communicate with each other to confirm with each other the failing node is unavailable. It is recommended to put the cluster into move resources and disable node mode, so that the node must be manually checked by an admin before it is put back into service.

Performing Maintenance

When performing maintenance on a node you can disable the node making it unavailable to receive Resource Groups. You can have it move Resource Groups automatically to the other node when disabling it.

Witness Configuration

Copies of the Witness binary for Windows and Linux are available by going to the web page of the BrickStor Appliance on port 8443. Point your browser to <https://<BrickStor Admin0 IP>:8443>

Windows Witness

Setup Steps

Witness Service

1. Retrieve a copy of the windows hiavd executable from the BrickStor in directory
2. Copy hiavd.exe to C:\hiavd
3. At the cmd prompt Type: hiavd.exe –install (to remove hiavd Type: hiavd –remove)
4. Type: sc failure "hiavd" actions= restart/60000/restart/60000/restart/60000/60000 reset= 0
5. Type: sc start hiavd (to start the service, then do it again to make sure service is started, on the Recovery tab, make sure there is a 1 for Restart service after)

Witness Networking Setup

1. Configure NIC
2. Open Firewall port to allow ICMP

- a. Firewall & Network Protection
 - b. Advance Setting
 - c. Inbound Rules
 - d. File and Printer Sharing (Echo Request ICMPv4-In)
 - e. Click the Scope tab and under Remote IP Address, Click Any IP Address
 - f. Click Advance tab and under Edge Traversal, select Allow Edge Traversal
 - g. Click OK
3. Configure port 4746 for inbound/outbound traffic
 - a. Create a firewall rule for port 4746 (inbound/outbound)
 - b. Type: netstat -a (look for listening port 4746)
 4. Enable Remote Desktop
 - a. In the search bar Type: System Properties
 - b. Click "Remote Desktop" and switch on to turn it on

Linux Witness

Setup Steps

If you wish to install the witness service on Linux you can either use the OVA provided by RackTop or install the binary and create a service that automatically starts with the witness. You must have a witness for each cluster.

To Configure the OVA deployed on Cent OS 7.6 perform the following steps:

1. Type name of virtual machine and drag and drop the below files:
2. "RackTop HA Witness _Inx_Signed.ovf"
3. "**RackTop_HA_Witness__Inx_Signed-disk1.vmdk**"
4. Click "**Next**"
5. Click "Select Datastore" and click "**Next**"
6. Click "**I agree**" and click "**Next**"
7. Click "Select Network" and click "**Next**"
8. Click "**Finish**"
9. Click VM and open **Web Console**
10. Logon using the following credentials:
11. UserID: **racktop**

12. Password: **racktop**
13. Type: **sudo nmtui** (to set hostname and update the IP configuration)
14. Type password for racktop user at the prompt.
15. Select "Set system hostname" and press <ENTER>
16. Type, host name and press <ENTER>
17. Press <ENTER> again
18. Select "Edit a connection" and press <ENTER>
19. Select "Wired connection 1", tab to "Edit" and press <ENTER>
20. Tab to "IPv4 Configuration", highlight "<Automatic>" and press <ENTER>
21. Select "Manual" and tab to "Show" and press <ENTER>
22. Tab to "Addresses", and click <ENTER> to configure.
23. Tab to "Gateway", and click <ENTER> to configure.
24. Tab to "DNS Servers", and click <ENTER> to configure.
25. Tab to "OK", and click <ENTER> to configure.
26. Tab to "<Back>", and click <ENTER>
27. Select "Quit", tab to "<OK>" and press <ENTER>.
28. Type: **sudo reboot**

Finishing the HA Cluster Setup

Once the following steps are complete you are ready to create the HA cluster from the System Tab of the GUI on one of the BrickStor Nodes in the cluster.

- Admin0 online and working on each node
- The heartbeat (hb0) interface has been created on each node and there is a network cable connecting both nodes to each other directly
- Witness service has been installed and is running
- A test pool has been created on one node
- The default data interface has been created on both nodes with the same name

Go to the system menu tab and click Setup HA Cluster and fill in the appropriate information and passwords into the following screen.

Setup HA Cluster

General Requirements:

- All members powered on and able to ping each other via non-crossover address.

Node Requirements:

- Connected to shared enclosure with one or more disks.
- Common pool visible but not imported by both nodes.
- Connected via crossover Ethernet cable.
- Staged hb0 vnic created on physical crossover interface.

Witness Requirements:

- HA service running and listening on HA comms port.
- Not member of another cluster.

Local Node		Remote Node		Witness	
10 . 1 . 29 . 240	X	address	10 . 1 . 29 . 241	X	address
192 . 255 . 0 . 1	X	crossover	192 . 255 . 0 . 2	X	10 . 1 . 29 . 133
<i>password</i>		root pwd	<i>password</i>		X

Common Resource Group Physical Interface
Interface name on nodes for HA resource group creation.

aggr0

Common HA Comms Port (advanced)
All members will use this port to communicate with each other (default 4746).

4,746

You can also reconfigure an existing cluster using this screen. Reasons to reconfigure the cluster include changing the IP or Port of the witness, after the admin0 address of a node is changed, choose a different default interface for resource groups.

Data Protection Best Practices

1. When setting up replication, especially for larger data sets where data is being written, snapshots should be set to run more frequently than you may run them during normal operation. Each snapshot becomes a replication job, and since more frequent snapshots will be smaller, there is less likely to be a failure to replicate due to network errors or latency. Any replication retransmits are also more likely to be successful.

2. In cases where an encrypted data set is being replicated, keys should be exported from the local BrickStor and imported on the remote BrickStor so that the data can be recovered there.
3. Use the advanced configuration parameters to optimize your replication:
 - Priorities can be set to determine which data sets will replicate first
 - Bandwidth throttling can be configured to optimize how much bandwidth is used and at what times of day, so that you can take advantage of low traffic periods and avoid high traffic periods.
 - Optimize snapshot retention periods on both ends
 - On the local system, make sure that snapshots are not aging out before they are replicated.
 - On the remote system, you may want longer retention periods, but this will also consume storage, so consider this balance.
4. Replication peers should be on an appropriate data network that will be available and not interfere with other network traffic.
5. Setup bsradm notify for snapshot reporting so that you can be sure your replications are successful.

Encryption Best Practices

For Users with the Local Key Manager

1. Regularly export the keys from the local key manager and save them in a safe controlled location off the BrickStor. In an HA cluster export and import the keys from both nodes to the other node and then export the keys from one node for backup. This should be done any time new encrypted datasets are created.
2. Import dataset keys to remote systems that are replication targets for fast recovery
3. Do not enable automatic key rotation
4. Enable key import and key export
5. Do not enable crypto-erase unless this is something you will need to do as part of regular operations
6. Do not enable unenroll drives so that nobody except an admin who modifies the config first can allow that operation
7. Periodically review the drive status report and the dataset encryption report
8. Manually perform a rekey based on organizational policies for encryption key rotation
9. Test recovery of files on the replication target to verify access to data during a non-critical time

For Users with the Local Key Manager

1. Verify your external key manager has appropriate backups and COOP plans.
2. Enable automatic key rotation
3. Determine if you want to enable key export based on your security posture and if you need them for COOP planning

4. Do not enable crypto-erase unless this is something you will need to do as part of regular operations
5. Verify replication targets can access appropriate dataset encryption keys on the key manager or export them and import them to the replication targets key manager.
6. Do not enable unenroll drives so that nobody except an admin who modifies the config first can allow that operation
7. Periodically review the drive status report and the dataset encryption report
8. Test recovery of files on the replication target to verify access to data during a non-critical time

High Availability (HA) Best Practices

1. Configure a dedicated witness for each HA cluster that is not using any of the cluster's storage. Eliminate routing between the witness and cluster
2. Use at least two resource groups to make the cluster active/active and balance them across both nodes.
3. Place the witness on the same local subnet as the out of band management (ILO,RMM) and admin0
4. Use an aggregate for the data network across two switches to improve network availability
5. Configure the cluster to disable the node after an automatic resource move
6. Don't try to fail the pool over manually in cases where the pool itself is degraded. Degraded pools can take longer to import. Figure out the root cause of why the pool is not healthy first
7. Fail a pool over if the system services on the node are failing but the pool is healthy. You can shut the node to a powered off state quickly by using the out of band management and perform a non-graceful shutdown. Failed or hung services may prevent the node from shutting down.

Command Line Operations

This section covers operations that can be performed via the command line. These operations are often performed via the GUI or the install tool as well.

Configuring Ethernet Address on Physical Interfaces

Network configuration model of any BrickStor appliance is essentially identical and as follows. All systems have an even number, two or four onboard 1GbE or 10GbE RJ45 interfaces clustered together to the right of the power supplies. Onboard 10GbE, 25GbE, etc. fiber or copper interfaces are typically included with every system and either reside on one of the two riser cards or installed into a special interface on the system board, on the far-right edge of the chassis. Under normal circumstances the leftmost Ethernet interface is automatically configured for the purposes of management access, which includes 'ssh' for some initial setup and for diagnostic as well as some feature configuration not currently accessible via the graphical management interface, as well as for access via the graphical management interface.

All physical interfaces are abstracted with one or multiple virtual interfaces, which do not typically share the MAC address of their underlying physical interface. Relationship of virtual interfaces to physical is

many to one, meaning a single physical interface may possess one or multiple virtual interfaces. Multiple virtual interface over a single physical interface are common in VLAN tagging scenarios, which we discuss later in the document.

It is also possible to bond links into what we commonly refer to as an aggregate, more commonly known as a LAG (Link Aggregation Group), with or without LACP capability. Due to various complexities of configuration and the wide range of possible configurations we will not discuss the details of this configuration in this guide. If this is a requirement, please contact RackTop support for details.

Multiple virtual interfaces do not share a MAC address, instead each is assigned a randomly generated address, unless one is explicitly provided. We will not discuss the details of this customization here. If this is a requirement, please contact RackTop support for details.

The appliance ships with at least one already existing virtual interface named 'admin0', configured to automatically obtain an IP address via DHCP to ease initial configuration whenever possible. This is the *primary* management interface, meaning this is the interface used to manage the machine as an administrator, not meant for data traffic normally.

At least one *data* interface is required to expose files via one or more supported protocols: AFP, NFS and SMB. For all interfaces other than management, which typically will already exist and will not need to be created or re-created, use the naming convention 'dataXX', where 'XX' is a non-negative numeric value starting with 0. All physical interfaces are suitable candidates, sans the first interface used for management as discussed previously. Needs vary, but a typical configuration will use 10GbE, 25GbE, and similar high bandwidth interfaces for all data access.

Virtual interfaces on BrickStorOS are configured via 'dladm', which must be created before a physical link can be used. After a system has been connected to network equipment, information about state of physical interfaces can be seen with a command in the following example. A typical output follows this general appearance:

```
# dladm show-phys
```

LINK	MEDIA	STATE	SPEED	DUPLEX	DEVICE
ixgbe0	Ethernet	down	0	unknown	ixgbe0
ixgbe1	Ethernet	down	0	unknown	ixgbe1
igb0	Ethernet	up	1000	full	igb0
igb1	Ethernet	unknown	0	half	igb1
igb2	Ethernet	unknown	0	half	igb2
igb3	Ethernet	unknown	0	half	igb3

In the above example it can be seen that the link named 'igb0' is up and configured at 1GbE. This is the

physical interface on which virtual interface 'admin0' is provisioned. Typically, data interfaces will follow the naming convention prescribed earlier and use high speed interfaces, commonly identified as 'ixgbeXX', where 'XX' is a non-negative numeric value starting with 0. Following is an example of establishing such a data interface over physical interface called 'ixgbe0'.

```
# dladm create-vnic -l ixgbe0 data0
```

Once a virtual interface has been created, an IP address must be assigned to this interface. IP interfaces on BrickStorOS are configured via 'ipadm'. The default 'admin0' IP interface cannot be modified since it is a temporary interface from an ipadm standpoint, instead it needs to be created persistently if a static IP address assignment is required. If you need to create a static IP, perform the following either via ssh, while connected via an IP address assigned to another interface, or directly via console of virtual console:

```
# ipadm delete-if admin0
```

```
# ipadm create-if admin0
```

```
# ipadm create-addr -T static -a local=x.x.x.x/24 admin0/v4
```

Where in the last command 'x.x.x.x/24' is the IP address/CIDR and 'admin0/v4' is the interface name and IP version (4 or 6). Upon creation of an IP interface, two addresses are configured, IPv4 and IPv6. For all intents and purposes IPv6 interface should be ignored usually.

VLAN Tagging

If VLAN tagging is setup on the port for trunking, you can create an interface like shown:

```
# dladm show-link
```

This will give you a list of available links for the next step, which is:

```
# dladm create-vlan -l ixgbe0 -v 10 vlan10
```

Replace ixgbe0 with an appropriate physical interface from your system and vlan10 with the name for your vlan. Note: vlan name must lead with a letter and also contain at least one number.

Link Aggregation (Bonding)

If link aggregation is required, first create an aggregate and then create a vnic on top of it:

```
# dladm create-aggr -l ixgbe0 -l ixgbe1 0
```

Where '0' denotes the number that will be placed in the name 'aggr0'. After that, create a vnic on top of the aggregate:

```
# dladm create-vnic -l aggr0 data0
```

Configuring Default Gateway

If, in the previous steps, you have deleted an interface, you may not have a default gateway if the interface that was deleted was the only one on its subnet. You can find your default gateway by using:

```
# netstat -rn
```

Routing Table: IPv4

Destination	Gateway	Flags	Ref	Use	Interface
default	10.1.12.254	UG	3	5761	
10.1.12.0	10.1.12.196	U	7	2008782	admin0
127.0.0.1	127.0.0.1	UH	2	70	lo0

Routing Table: IPv6

Destination/Mask	Gateway	Flags	Ref	Use	If
::1	::1	UH	2	10	lo0

From there, under the flags column you are looking for a 'G', which stands for gateway and a 'default' designation under the 'Destination' column. You can add a new permanent default route using the following:

```
# route -p add default x.x.x.x
```

Where x.x.x.x is your default gateway. You can now see your default route in 'netstat -rn'

BSRAPID Configuration

By default BSRAPID is configured to listen on all interfaces on port 8443. However, the service can be configured to listen on a different port and on specific IP addresses. To configure this behavior, change the Listen Address in `/etc/racktop/bsrapid/bsrapid.conf`

Configure BSRAPID to listen on any interface with port 5443

```
ListenAddress = ":5443"
```

Configure BSRAPID to listen only on 10.1.12.120 port 5443

```
ListenAddress = "10.1.12.120:5443"
```

Time Zone Setup

Set the time zone of BrickStor through the command line by editing the following file.

```
# tzselect
```

Then follow the prompts. A reboot is required for the changes to take effect.

NTP Setup

Preparing to Setup and Sync Time

First disable the NTP service so that you can synchronize time for the system to the NTP server. By default, the NTP service is configured to get time from the pool.ntp.org service.

You can enable from the command line or the GUI. To enable by command line:

```
# svcadm disable ntp
```

Next run the `'ntpdate'` command to synchronize time. This should show a current offset.

Note: ntp service must be disabled for ntpdate to work

```
# ntpdate <IP of Time Server>
```

If the offset was very large you can run the `ntpdate` command again to verify that clock was adjusted accordingly and offset now should be very small.

Example:

```
# ntpdate pool.ntp.org
```

```
10 Sep 08:30:08 ntpdate[7063]: step time server 129.6.15.28 offset -  
17971.406299 sec
```

```
# ntpdate pool.ntp.org
```

```
10 Sep 08:30:31 ntpdate[7064]: adjust time server 129.6.15.29 offset 0.002656  
sec
```

Problems with SMB authentication or AD join may be related to BrickStor's time being 5 minutes or more out of sync with Active Directory time.

Hosts Entries

Setting up hosts entries

Most of the time this should not be necessary, but in the exceptional cases where host name resolution is required and cannot be accomplished via DNS, static entries may be added to allow for local

resolution. This activity is accomplished via *'bsradm'* as follows, where *'192.168.0.1'* is the address and *'othernode'* is name resolving to this address:

```
# bsradm hosts add --ip 192.168.0.1 --names othernode
```

Note: this may be a required step if DNS is not setup and you are connecting to an NFS datastore from ESXi.

RMM (Remote Terminal) IP Address

Your BrickStor storage appliance comes equipped with a Remote Management Module frequently abbreviated to RMM. RackTop recommends connecting this Ethernet interface as well as the *'admin0'* management Ethernet interface to a dedicated management network, if one is available. Separation of management and administration concerns from data access is a recommended best practice. This enables you to access the appliance as if you were standing in front of it with a crash cart or KVM, even when services such as SSH are down. You can use RMM to power cycle the machine or see and use the console. If RMM is already configured, you can find the IP address with this command from the terminal:

```
# bsradm hw rmm
```

IpSource: DHCP Address

IpAddress: 192.168.0.101

SubnetMask: 255.255.255.0

MacAddress: 00:1e:67:50:c7:c1

SnmpCommunityString: public

DefaultGateway: 192.168.0.1

Vlan: 0

Once you have the IP address, you can login to RMM via your browser. You will need to use Java to access the console and this will most likely require adding a security exception for the IP address in the Java control panel.

Creating Local Accounts

As root you can create local accounts that can be used for controlled access to shares as well as providing access to administrative functions such as the ability to manage BrickStor with the myRack Manager.

```
#useradd <username>
```

To set the user's password:

```
#passwd <username>
```


Add Local Accounts to Bsradmins Group

To allow a given local account administrative access of a BrickStor appliance via myRack Manager, this account must be in the *'bsradmins'* group of the appliance. To add a user to the group, run the following command, replacing username placeholder with actual local account name:

```
# usermod -G bsradmins <username>
```

Adding and removing e-mail addresses from Notification List

To add e-mail addresses to receive notifications from the BrickStor appliance, use the following command format at the terminal:

```
# bsradm notify add <email address> -all
```

Other options besides the "all" notifications options are:

```
--system      Add to system notification list
--reports     Add to reports notification list
--faults      Add to faults notification list
```

To list users and their notification types, use:

```
# bsradm notify show
```

And to remove users from their notification, use:

```
# bsradm notify remove <email address> --all
```

Joining Active Directory

The first step for making a CIFS share available for users is to join Active Directory, which requires several configuration steps before joining the domain will be possible. A machine account will be created for a BrickStor upon successful domain join operation. This machine account will enable users to passthrough authenticate and be either permitted or denied access to shares without requiring separate authentication against the BrickStor. In other words, once users are logged into Active Directory, their authentication information is stored on their system and in Active Directory, and no further authentication prompts are necessary in order to access shares on a domain-joined BrickStor.

Active Directory requires certain attributes of name resolution, which usually means the BrickStor must be configured to resolve names against domain in the given instance of Active Directory to which it will be bound. BrickStor's domain setting must also be set to name of domain being joined.

First, validate what is currently configured, because no change may be necessary. Check currently configured domain with the following command:

```
# bsradm dns domain get
```

If the value reported is correct, that is, it matches the Active Directory domain name, no change is necessary. If, however a modification is necessary, change should be made with the following command, replacing placeholder 'domain.tld' with actual fully qualified Active Directory domain name:

```
# bsradm dns domain set <domain.tld>
```

Next, confirm that correct DNS resolvers are configured, and if not, make necessary changes. In most environments at least two DNS servers will be configured and BrickStor must point to these resolvers, which in typical Active Directory configurations will be domain controllers also, or commonly member servers with a dedicated DNS function.

First, validate what is currently configured, because no change may be necessary. Check currently configured domain name resolution servers with the following command:

```
# bsradm dns ns show
```

If values reported are correct, no further resolver changes should be necessary. If, however DNS servers need changing, use the following commands to add/remove entries, replacing placeholder 'address' with IP address of system being added or removed.

```
# bsradm dns ns add <address>
```

```
# bsradm dns ns remove <address>
```

Note: NTP must be correctly configured with accurate synchronized timing with the Domain Controller before you can join the Domain Successfully

The command for joining the storage appliance to the domain is:

```
# smbadm join -y -u <Administrator Account> <domain.tld>
```

Where 'Administrator' is the name of the user you want to use to join the domain. This account is only used to create the computer object and does not need to be a service account. You will be prompted for a password. If the join fails, please double check your username and password and the settings in /etc/resolv.conf.

To view the domain type:

```
# smbadm list
```

Also verify that forward and reverse lookups are correct within active directory for the BrickStor.

iSCSI Share Configuration

It is common practice to separate block and file traffic on different physical interfaces however this is not required.

Creating a Default Target and Target Portal Group

Create a target portal group to restrict the target to your data0 (data, not management) IP address:

```
# itadm create-tpg global <x.x.x.x>
```

Where <x.x.x.x> is the IP address associated with data0. Next, we need to create the default target. To create the target, type the following:

```
# itadm create-target
```

Now check the status of your targets to make sure everything is okay:

```
# itadm list-target -v
```

TARGET NAME	STATE	SESSIONS
iqn.2010-03.com.racktopsystems:02:c434c8d7-5643-6364-af5d-cb0bae33d531	online	0
alias:	-	
auth:	none (defaults)	
targetchapuser:	-	
targetchapsecret:	unset	
tpg-tags:	default	

Next, modify your target to be part of the target portal group:

```
# itadm modify-target -t global <iqn>
```

Where <iqn> is the target listed in the previous step. From here, you should be able to manage the rest from the MyRack Manager GUI.

Example:

```
# itadm modify-target -t global iqn.2010-03.com.racktopsystems:02:c434c8d7-5643-6364-af5d-cb0bae33d531
```

```
# itadm list-target -v
```

TARGET NAME	STATE	SESSIONS
iqn.2010-03.com.racktopsystems:02:c434c8d7-5643-6364-af5d-cb0bae33d531	online	0
alias:	-	
auth:	none (defaults)	
targetchapuser:	-	
targetchapsecret:	unset	
tpg-tags:	global = 2	

Configuration & Performance Implications

RAID Performance

BrickStor uses mirrors and RAID-Z for disk level redundancy within vdevs.

RAIDZ

RAID-Z vdevs are a variant of RAID-5 and RAID-6:

- You can choose the number of data disks and the number of parity disks. Today, the number of parity disks is limited to 3 (RAID-Z3).
- Each data block that is handed over to ZFS is split up into its own stripe of multiple disk blocks at the disk level, across the RAID-Z vdev. This is important to keep in mind: Each individual I/O operation at the file system level will be mapped to multiple, parallel and smaller I/O operations across members of the RAID-Z vdev.
- When writing to a RAID-Z vdev, ZFS will use a best fit algorithm when the vdev is less than 90% full.
- Write transactions in ZFS are always atomic, even when using RAID-Z: Each write operation is only finished if the überblock has been successfully written to disk. This means there's no possibility to suffer from the traditional RAID-5 write hole, in which a power-failure can cause a partially (and therefore broken) written RAID-5 set of blocks.
- Due to the copy-on-write nature of ZFS, there's no read-modify-write cycle for changing blocks on disk: ZFS writes are always full stripe writes to free blocks. This allows ZFS to choose blocks that are in sequence on the disk, essentially turning random writes into sequential writes, maximizing disk write capabilities.

Just like traditional RAID-5 and RAID-6, you can lose up to 1 disk or 2 disks respectively without losing any data using RAID-Z1 and RAID-Z2. And just like ZFS mirroring, for each block at the file system level, ZFS can try to reconstruct data out of partially working disks, as long as it can find a critical number of blocks to reconstruct the original RAID-Z group.

Performance of RAIDZ

When the system writes to a pool it writes to the vdevs in a stripe. A Vdev in a RAID-Z configuration will have the IOPS and performance characteristics of the single slowest disk in that vdev (it will not be a summation of the disks). This is because a read from disk requires a piece of data from every disk in the vdev to complete the read. So, a pool with 3 vdevs in a RAID-Z1 with 5 disks per vDEV will have the raw IOPS performance of 3 disks. You may see better performance than this through caching, but this is the most amount of raw IOPS the pool can deliver from disk. The more vdev's in the pool the better the performance.

Performance of Mirrors

When the vdev's are configured as mirrors the configuration of the pool is equivalent to RAID-10. A pool with mirrored vdev's will always outperform other configurations. A read from disk only needs data

from one disk in the mirror. As with RAID-Z, the more vdevs the better performance will be. Resilver times with mirrored vdevs will be faster than with RAID-Z and will have less of a performance impact on the overall system during resilvering.

RackTop recommends the use of mirrored vdevs in environments with high random IO such as virtualization because it provides the highest performance.

Compression

Compression is performed inline and at the block level. It is transparent to all other layers of the storage system. Each block is compressed independently and all-zero blocks are converted into file holes. To prevent “inflation” of already-compressed or incompressible blocks, BrickStor maintains a 12.5% compression ratio threshold below which blocks are written in uncompressed format.

BrickStor supports compression via the LZJB, GZIP (levels 1-9), LZE, and LZ4. RackTop finds that LZ4 works very well, balancing speed and compression performance. It is common to realize a 1.3 to 1.6 compression ration with highly compressible data which not only optimizes storage density but also improves write performance due to the reduction in disk IO.

RackTop recommends always using compression because any CPU penalty is typically outweighed by the savings in storage and bandwidth to the disk.

Deduplication

Deduplication is performed inline and at the block level, also like compression, deduplication is transparent to all other layers of the storage system. For deduplication to work as expected the blocks written to the system must be aligned. Deduplication even when turned off will not reverse the deduplication of blocks already written to the system. This can only be accomplished through copying or moving the data. Deduplication negatively impacts the system performance if data is not significantly duplicative because an extra operation must be done to look if it is a duplicate block for writes and if it is the last block for deletes. Additionally, the deduplication table must be stored in RAM. This takes up space that could otherwise be used for metadata and caching. Should the deduplication not all fit in RAM then system performance will degrade sharply because every read and write operation will require the system to reread the dedup table from disk.

Deduplication is only supported on All Flash Pools.

Clones

ZFS clones create an active version of a snapshot. By creating a snapshot of a base VM and using clones of that same snapshot you can have an unlimited number of copies of the same base virtual machine without taking up more storage capacity. The only increased storage footprint will come from the deltas or differences between clones. Additionally, since each VM will reference the same set of base data blocks the system and user will benefit from caching since all VM’s will be utilizing the same blocks of data.

Imbalance of vdev Capacity

If you wish to grow the capacity of a volume by adding another vdev you should do so by adding a vdev of equivalent size to the other vdevs in the pool. If the other vdevs are already past 90% capacity they will still be slow because data will not automatically balance or spread across all vdevs after the additional capacity is added. To force a rebalance in a VMware environment you can perform a vmotion or storage migration. With the Copy On Write Characteristics of ZFS, the pool will automatically rebalance across all vdevs.

Performance Monitoring

There are several scripts included with BrickStor for monitoring the performance of the storage portion of the system. Dtrace and kstat are powerful tools for analyzing the storage performance and behavior.

IOStat

IOStat is one of the most common tools to assess disk performance. Running `!iostat -xn 5` from the command line will result in an output similar to the following.

```

extended device statistics
r/s  w/s  kr/s  kw/s  wait  actv  wsvc_t  asvc_t  %w  %b  device
2.7 142.4  1.0  879.8  0.6  0.0  4.0  0.1  14  2  c2t0d0
2.0  1.3  11.3  0.0  0.0  0.0  0.0  0.0  0  0  c0t5000C5002E4A47DAd0
2.0  1.3  11.3  0.0  0.0  0.0  0.0  0.0  0  0  c0t5000C5002E4B60D8d0
2.0  1.3  11.3  0.0  0.0  0.0  0.0  0.0  0  0  c0t5000C5002E46D2C5d0
2.0  1.3  11.3  0.0  0.0  0.0  0.0  0.0  0  0  c0t5000C5002E4AC53Cd0
2.0  1.3  11.3  0.0  0.0  0.0  0.0  0.0  0  0  c2t1d0
2.0  1.3  11.3  0.0  0.0  0.0  0.0  0.0  0  0  c0t5000C5002E4B0940d0
2.0  1.3  11.3  0.0  0.0  0.0  0.0  0.0  0  0  c0t5000C5002E4665F8d0
2.0 24.3  11.3 137.9  0.0  0.2  0.0  6.7  0  8  c10t50000393C8C93AF6d0
2.0 24.7  11.3 137.9  0.0  0.2  0.0  6.5  0  8  c10t50000393C8C91A5Ad0
0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0  0  c10t50000393C8C918E6d0
2.0  6.0  11.3 109.4  0.0  0.0  0.0  0.1  0  0  c0t5001517BB2863DF8d0
2.0  1.3  11.3  0.0  0.0  0.0  0.0  0.0  0  0  c0t5001517BB27C697Cd0
2.0 36.3  11.3 269.7  0.0  0.5  0.0 13.6  0 18  c10t50000393C8C9184Ed0
2.0 37.3  11.3 269.7  0.0  0.4  0.0  9.0  0 11  c10t50000393C8C93BAEd0
    
```

Key values to look for are %w, %b, asvc_t. It is normal for these values to increment beyond zero in a heavily loaded system, but they should usually be in relation to the overall system load. If the system is not heavily utilized, and these values are consistently high on a single device, it may indicate that the device is experiencing a hardware issue.

Zpool iostat

Zpool iostat shows extended details about a ZFS pool. Running `!zpool iostat -v 3` from the command line will result in an output similar to the following:

```

                                capacity  operations  bandwidth  latency
pool          alloc free read write read write read write
-----
poolA          1.18T 6.98T  0  0  0  0  0.00  0.00
mirror         403G 2.33T  0  0  0  0  0.00  0.00
c0t5000C5002E46D2C5d0  -  -  0  0  0  0  0.00  0.00
    
```

```

c0t5000C5002E4A47DAd0 - - 0 0 0 0 0.00 0.00
mirror          403G 2.33T 0 0 0 0 0.00 0.00
c0t5000C5002E4AC53Cd0 - - 0 0 0 0 0.00 0.00
c0t5000C5002E4B60D8d0 - - 0 0 0 0 0.00 0.00
mirror          403G 2.33T 0 0 0 0 0.00 0.00
c0t5000C5002E4B0940d0 - - 0 0 0 0 0.00 0.00
c0t5000C5002E4665F8d0 - - 0 0 0 0 0.00 0.00
c2t1d0          784K 29.7G 0 0 0 0 0.00 0.00
-----
syspool          16.1G 216G 0 2 0 11.9K 0.00 0.22
c2t0d0s0        16.1G 216G 0 2 0 11.9K 0.00 0.22
-----
vmpool01         160G 3.47T 0 131 0 633K 0.00 27.46
mirror          130G 1.69T 0 19 0 68.3K 0.00 24.90
c10t50000393C8C91A5Ad0 - - 0 13 0 69.6K 0.00 10.56
c10t50000393C8C93AF6d0 - - 0 13 0 69.6K 0.00 9.15
c0t5001517BB2863DF8d0 2.89M 22.2G 0 9 0 82.0K 0.00 0.11
mirror          30.3G 1.78T 0 36 0 159K 0.00 36.40
c10t50000393C8C93BAEd0 - - 0 20 0 160K 0.00 14.72
c10t50000393C8C9184Ed0 - - 0 19 0 160K 0.00 21.07
cache           - - - - - - -
c0t5001517BB27C697Cd0 99.1G 12.7G 0 0 0 0 0.00 0.00
-----

```

In this instance, you would be looking for latency values > 100 for extended periods of time as an indicator of an overloaded system.

Default System Service Ports and Protocols

Service	Description / Purpose	Direction	Port
DNS	Domain Name Service	both	UDP 53
NTP	Time synchronization	both	UDP 123
AFP	Apple client access	in	TCP 548
NFS/portmap	NFS client access	in	TCP/UDP 2049
NFS/rpc	NFS client access	In	TCP/UDP 111
NFS/lockmgr	NFS client access	In	TCP/UDP 4045
iSCSI	iSCSI client/initiator access	In	TCP 3260 and 3205
SMB	SMB/CIFS client access	in	TCP/UDP 139, 445
LDAP	Access to directory service servers	out	TCP/UDP, 389, 636
Kerberos	Authentication	out	UDP 88
SSH	Management and Replication data receive	in	TCP 22
TCP Replication	Replication send	out	TCP 22, 8444
mail	Notification emails	out	TCP 25, 587
syslog	Logging	out	TCP/UDP 514
bsrapid	Used for MyRack Manager (https)	in	TCP 8443
influxdb	Used for MyRack Manager (charts)	in	TCP 8086, 8088
hiavd	High Availability (between HA nodes)	both	TCP 4746
KMIP	Access to key management server	out	TCP 5696, 8445
SNMP	Monitoring with SNMP	both	UDP 161
SNMP traps	Sending alerts to SNMP stations	out	UDP 162
HTTPS	Call Home for Software Updates (https://api.myracktop.com)	out	TCP 443