



Red Hat Ceph Storage 3.0

Release Notes

Release notes for Red Hat Ceph Storage 3.0

Red Hat Ceph Storage 3.0 Release Notes

Release notes for Red Hat Ceph Storage 3.0

Legal Notice

Copyright © 2018 Red Hat, Inc.

The text of and illustrations in this document are licensed by Red Hat under a Creative Commons Attribution–Share Alike 3.0 Unported license ("CC-BY-SA"). An explanation of CC-BY-SA is available at

<http://creativecommons.org/licenses/by-sa/3.0/>

. In accordance with CC-BY-SA, if you distribute this document or an adaptation of it, you must provide the URL for the original version.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, JBoss, OpenShift, Fedora, the Infinity logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux ® is the registered trademark of Linus Torvalds in the United States and other countries.

Java ® is a registered trademark of Oracle and/or its affiliates.

XFS ® is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL ® is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Node.js ® is an official trademark of Joyent. Red Hat Software Collections is not formally related to or endorsed by the official Joyent Node.js open source or commercial project.

The OpenStack ® Word Mark and OpenStack logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation, or the OpenStack community.

All other trademarks are the property of their respective owners.

Abstract

The Release Notes document describes the major features and enhancements implemented in Red Hat Ceph Storage in a particular release. The document also includes known issues and bug fixes.

Table of Contents

CHAPTER 1. INTRODUCTION	3
CHAPTER 2. ACKNOWLEDGMENTS	4
CHAPTER 3. MAJOR UPDATES	5
CHAPTER 4. TECHNOLOGY PREVIEWS	11
CHAPTER 5. DEPRECATED FUNCTIONALITY	12
CHAPTER 6. KNOWN ISSUES	13
CHAPTER 7. NOTABLE BUG FIXES	22
CHAPTER 8. SOURCES	26

CHAPTER 1. INTRODUCTION

Red Hat Ceph Storage is a massively scalable, open, software-defined storage platform that combines the most stable version of the Ceph storage system with a Ceph management platform, deployment utilities, and support services.

The Red Hat Ceph Storage documentation is available at <https://access.redhat.com/documentation/en/red-hat-ceph-storage/>.

CHAPTER 2. ACKNOWLEDGMENTS

Red Hat Ceph Storage version 3.0 contains many contributions from the Red Hat Ceph Storage team. Additionally, the Ceph project is seeing amazing growth in the quality and quantity of contributions from individuals and organizations in the Ceph community. We would like to thank all members of the Red Hat Ceph Storage team, all of the individual contributors in the Ceph community, and additionally (but not limited to) the contributions from organizations such as:

- Intel
- Fujitsu
- UnitedStack
- Yahoo
- UbuntuKylin
- Mellanox
- CERN
- Deutsche Telekom
- Mirantis
- SanDisk

CHAPTER 3. MAJOR UPDATES

This section lists all major updates, enhancements, and new features introduced in this release of Red Hat Ceph Storage.

New ways to identify client versions

This update adds the following features that help with identifying client versions to determine which clients use an old version of Red Hat Ceph Storage.

- The **ceph osd set-require-min-compat-client** command adds the ability to set a minimum required release for clients to prevent new connections from older clients. By default it is set to **jewel**. To view its value, use the **ceph osd dump** command.
- The **ceph features** command that reports the total number of clients and daemons and their features and releases.
- If the debugging level for Monitors is set to **10** (**debug mon = 10**), addresses and features of connecting and disconnecting clients are logged to log file on a local file system.

A new --pg-num option for the osdmapprool utility

The **osdmapprool** utility now includes the **--pg-num** option that can be used with the **--test-map-pgs** option. This allows the user to test placement policies with a different number of placement groups (PGs) than are in the OSD map.

Option to add a limit on RBD snapshots

A new option to set a limit on the number of snapshots on a RADOS Block Device (RBD) image is now supported. Use the option **snap limit --limit** with the **rbd** command to set the limit.

Ansible now supports removing Monitors and OSDs

You can use the **ceph-ansible** utility to remove Monitors and OSDs from a Ceph cluster. For details, see the [Removing Monitors with Ansible](#) and [Removing OSDs with Ansible](#) sections in the Red Hat Ceph Storage 3 Administration Guide. The same procedures apply also for removing Monitors and OSDs from a containerized Ceph cluster.

The iSCSI gateway is now fully supported

Red Hat Ceph Storage 3.0 adds full support for the iSCSI gateway. These iSCSI initiators are supported:

- Red Hat Enterprise Linux 7.4
- VMware ESX 6.5
- Microsoft Windows Server 2016
- Red Hat Virtualization 4.x

For details, see the [Using and iSCSI Gateway](#) chapter in the Block Device Guide for Red Hat Ceph Storage 3.

The rbd export-diff and rbd import-diff commands now support parallelism

The **rbd export-diff** and **rbd import-diff** commands have been improved to being capable of fully parallel operations. As a result, the commands now benefit from concurrency across the cluster. The commands are executed in parallel by default. To configure the amount of parallelism, use the **--rbd-**

`concurrent-management-ops <number>` option when using the commands.

Support for deploying logical volumes as OSDs

A new utility, `ceph-volume`, is now supported. The utility enables deployment of logical volumes as OSDs on Red Hat Enterprise Linux. For details, see the [Using the ceph-volume Utility to Deploy OSDs](#) chapter in the Block Device Guide for Red Hat Ceph Storage. Note that `ceph-volume` does not support deploying logical volumes as OSDs in containers. In addition, `ceph-volume` is not tested on Ubuntu 16.04.03.

Bucket owners can grant permissions to other users

With this update, bucket owners can provide read access to their buckets to another user. For details, see the [Ceph - How to grant access for multiple S3 users to access a single bucket](#) solution on the Red Hat Enterprise Linux.

On a CephFS with only one data pool, the `ceph df` command shows characteristics of that pool

On Ceph File Systems that contain only one data pool, the `ceph df` command shows results that reflect the file storage spaces used and available in that data pool. This new functionality is available for FUSE clients only for now and will be available for kernel clients in a future release of Red Hat Enterprise Linux.

Promoting and demoting all images in a pool at once

You can now promote or demote all images in a pool at the same time by using the following commands:

```
rbd mirror pool promote <pool>
rbd mirror pool demote <pool>
```

This is especially useful in an event of a failover, when all non-primary images must be promoted to primary ones.

Ansible now automatically sets online repositories for Ubuntu

This update automates the process of setting up online repositories for Red Hat Ceph Storage on Ubuntu nodes. To set up the repositories, set the following parameters in the `all.yml` file located in the `/usr/share/ceph-ansible/group_vars/` directory:

```
ceph_origin: repository
ceph_repository: rhcs
ceph_repository_type: cdn
ceph_rhcs_cdn_debian_repo:
https://customername:customerpasswd@rhcs.download.redhat.com
```

Specify your customer name and password.

For details, see the [Installation Guide for Ubuntu](#).

A Red Hat Ceph Storage cluster can be deployed from an Ubuntu node by using Ansible

Previously, Red Hat did not provide the `ceph-ansible` package for Ubuntu. With this update, you can use the Ansible automation application to deploy a Ceph cluster from an Ubuntu node.

For details, see the [Installing a Red Hat Ceph Storage Cluster](#) section in the Installation Guide for Ubuntu.

A new compact command

With this update, the OSD administration socket supports the **compact** command. A large number of **omap** create and delete operations can cause the normal compaction of the **levelDB** database during those operations to be too slow to keep up with the workload. As a result, **levelDB** can grow very large and inhibit performance. The **compact** command compacts the **omap** database (**levelDB** or **RocksDB**) to a smaller size to provide more consistent performance.

Installing NFS Ganesha by using Ansible is supported

You can now install the NFS Ganesha interface by using the **ceph-ansible** playbook. For additional details, see the **all.yml** and **nfss.yml** file in the **/usr/share/ceph-ansible/** directory on the Ansible administration node.

RocksDB now replaces levelDB

This update changes the default back end for the **omap** database from the **levelDB** to the **RocksDB** database. **RocksDB** uses the multi-threading mechanism in compaction so that it better handles the situation when the **omap** directories become very large (more than 40 G). **LevelDB** compaction takes a lot of time in such a situation and causes OSDs to time out.

Simplified creation of CephFS client keyring

A new command, **ceph fs authorize**, is now supported. The command simplifies creation of **cephx** capabilities for a Ceph File System (CephFS) client user. For example, to grant the **client.1** user read and write access to MDS nodes and read access to Monitor and OSD nodes on a Ceph File System named **cephfs**:

```
# ceph fs authorize cephfs client.1 rw r
```

Use this command only when creating new users. It is not possible to modify existing users with **ceph fs authorize**.

Granting access to Ceph Block Device images has been simplified

The **ceph auth get-or-create** command now supports two profiles, **rbd** and **rbd-read-only**. When using these profiles, **cephx** capabilities are created automatically without the need to specify them directly. For example, to create a **client.1** user with required capabilities for Monitors and OSDs:

```
ceph auth get-or-create client.1 mon 'profile rbd' osd 'profile rbd [pool=<pool>]'
```

OSDs support the **rbd** and **rbd-read-only** profiles. Monitors support only the **rbd** profile.

MDS cache limits can be configured in bytes

New configuration options are now supported that enable configuring Metadata Server (MDS) cache limits to be configured in bytes, not only in inodes count. For details, see the [Understanding MDS Cache Size Limits](#) section in the Ceph File System Guide for Red Hat Ceph Storage 3. Note that limiting the MDS cache by the inodes count is now deprecated.

Improvements in the cluster log

The cluster log has been improved. Certain unnecessary messages, such as audit log, PGMap 5 second, or print on every **osdmap** epoch, have been removed. Other messages were improved to use a more human-readable format. Also, a message is not logged when health checks fail. In addition, a new command, **log last**, is now supported. The command shows the recent log messages.

Ceph health checks are more easily integrated with external alerting systems

Ceph's built-in health checks have been refactored to enable more robust integration with external alerting systems. For each condition that is checked, there is now a unique status code, for example **PG_AVAILABILITY**.



NOTE

Any external script that was relying on the JSON syntax of the **ceph status** or **ceph health** command output must be updated for the new format. To ease migration, set the **mon_health_pre luminous_compat** parameter to **True** on Monitors to instruct **ceph status** and **ceph health** to generate old-style health output in addition to the new output.

Deleting images and snapshots from full clusters is now easier

When a cluster reaches its **full_ratio**, the following commands can be used to remove Ceph Block Device images and snapshots:

- **ceph rbd remove**
- **ceph rbd snap rm**
- **ceph rbd snap unprotect**
- **ceph rbd snap purge**

The Ceph Object Gateway now supports NFSv3 protocol

The Ceph Object Gateway now provides the ability to export Simple Storage Service (S3) object namespaces by using NFS version 3 alongside the existing NFS version 4. For details, see the [Exporting the Namespace to NFS-Ganesha](#) section of the Red Hat Ceph Storage 3 Object Gateway Guide for Red Hat Enterprise Linux.

Support for data compression

The Ceph Object Gateway now supports data compression at rest. For details, see the *Compression* section in the [Object Gateway Guide for Red Hat Enterprise Linux](#) or [Ubuntu](#).

Support for S3 Bucket Policy

Support for Simple Storage Service (S3) Bucket Policy has been added. Note that the support has the following limitations:

- Identity and Access Management (IAM) for users and groups is not supported
- String interpolation is not supported
- Only a subset of condition keys is supported

For details see the [Bucket Policies](#) section in the Developer Guide for Red Hat Ceph Storage 3.

nfs-ganesha rebased to 2.5

The **nfs-ganesha** package has been upgraded to upstream version 2.5, which provides a number of bug fixes and enhancements over the previous version.

NFSv4 recovery state data can be stored in Ceph RADOS

NFS version 4 (NFSv4) recovery state data such as, `clientids`, can now be stored in Ceph RADOS objects. This change increases the resilience of clustered NFS servers exposing Ceph storage resources.

New "radosgw-admin user list" command

Previously, the command that listed users and subusers required the user's uid as an input. This approach required extra commands. This release introduces the `radosgw-admin user list` command, which lists all users and subusers without requiring any uids.

S3 object expiration is now supported

The Ceph Object Gateway now supports the Amazon Simple Storage Service (S3) object expiration. For details see the [Object Gateway S3 Application Programming Interface \(API\)](#) chapter and the [Bucket Lifecycle](#) section in the Developer Guide for Red Hat Ceph Storage 3.

Support for S3 server-side encryption

The Ceph Object Gateway now supports the Amazon Simple Storage Service (S3) server-side encryption. For details, see the [S3 API Server-side Encryption](#) section in the Developer Guide for Red Hat Ceph Storage 3.

Support for the Red Hat Ceph Storage Dashboard

The Red Hat Ceph Storage Dashboard provides a monitoring dashboard for Ceph clusters to visualize the cluster state. The dashboard is accessible from a web browser and provides a number of metrics and graphs about the state of the cluster, Monitors, OSDs, Pools, or network.

For details, see the [Monitoring Ceph Clusters with Red Hat Ceph Storage Dashboard](#) section in the Administration Guide for Red Hat Ceph Storage 3.

The async messenger

The `async` messenger is used by default instead of the `simple` one. For details see the [Messaging](#) and [Async Messenger Settings](#) section in the Configuration Guide for Red Hat Ceph Storage 3.

Support for dynamic bucket resharding

The Ceph Object Gateway now supports the `rgw_dynamic_resharding` parameter. The process for dynamic bucket resharding periodically checks all the Ceph Object Gateway buckets and detects buckets that require resharding. If a bucket has grown larger than specified by the `rgw_max_objs_per_shard` parameter, the Ceph Object Gateway reshards the bucket dynamically in the background. For details, see the [Dynamic Bucket Index Resharding in RHCS 3](#) section in the Object Gateway Guide for Red Hat Enterprise Linux.

Note that dynamic bucket resharding is disabled in multi-site configuration.

The Ceph File System is now fully supported

The Ceph File System (CephFS) is a file system compatible with POSIX standards that provides a file access to a Ceph Storage Cluster. With this new version, CephFS is now fully supported. For details about CephFS, see the [Ceph File System Guide](#) for Red Hat Ceph Storage 3.

Scrubbing is blocked for any PG if the primary or any replica OSDs are recovering

The `osd_scrub_during_recovery` parameter now defaults to `false`, so that when an OSD is recovering, the scrubbing process is not initialized on it. Previously, `osd_scrub_during_recovery`

was set to **true** by default allowing scrubbing and recovery to run simultaneously. In addition, in previous releases if the user set **osd_scrub_during_recovery** to **false**, only the primary OSD was checked for recovery activity.

A new ceph-medic utility

A new utility, **ceph-medic**, is now available and fully supported. The utility detects common issues with a Ceph Storage Cluster that prevents the cluster from functioning properly. For details, see the [Installing and Using ceph-medic to Diagnose a Ceph Storage Cluster](#) chapter in the Troubleshooting Guide for Red Hat Ceph Storage 3.

Colocation of containerized Ceph daemons

With this release, you can colocate specific containerized Ceph daemons with OSD daemons on the same node. This approach significantly improves total cost of ownership (TCO) at small scale, reduces the minimum configuration from six nodes to three, makes upgrading more convenient, and provides better resource isolation. Also, each daemon has system resources reserved to avoid the "noisy neighbor" effect.

For details, see the [Colocation of Containerized Ceph Daemons](#) chapter in the Container Guide for Red Hat Ceph Storage 3.

Support for Ceph Manager

Ceph Manager (**ceph-mgr**) is a new daemon that takes over some of the Monitor's workload and introduces an interface for optional Python modules. Administrators must deploy at least two **ceph-mgr** daemons, or more typically, one **ceph-mgr** daemon on each node where they run a **ceph-mon** daemon. For details, see the [Installation Guide for Red Hat Enterprise Linux](#) or [Ubuntu](#).

Support for the RESTful plug-in

RESTful is a plug-in for the **ceph-mgr** daemon that provides an API for interacting with Ceph clusters.

For details, see the [Ceph Management API: Reference and Integration Guide](#).

CHAPTER 4. TECHNOLOGY PREVIEWS

This section provides an overview of Technology Preview features introduced or updated in this release of Red Hat Ceph Storage.



IMPORTANT

Technology Preview features are not supported with Red Hat production service level agreements (SLAs), might not be functionally complete, and Red Hat does not recommend to use them for production. These features provide early access to upcoming product features, enabling customers to test functionality and provide feedback during the development process.

For more information on Red Hat Technology Preview features support scope, see <https://access.redhat.com/support/offerings/techpreview/>.

OSD BlueStore

BlueStore is a new back end for the OSD daemons that allows for storing objects directly on the block devices. Because BlueStore does not need any file system interface, it improves performance of Ceph Storage Clusters.

To learn more about the BlueStore OSD back end, see the [OSD BlueStore \(Technology Preview\)](#) chapter in the Administration Guide.

Support for RBD mirroring to multiple secondary clusters

Mirroring RADOS Block Devices (RBD) from one primary cluster to multiple secondary clusters is now supported as a technology preview.

Erasure Coding for Ceph Block Devices

Erasure coding for Ceph Block Devices is now supported as a Technology Preview. For details, see the [Erasure Coding with Overwrites \(Technology Preview\)](#) section in the Storage Strategies Guide for Red Hat Ceph Storage 3.

CHAPTER 5. DEPRECATED FUNCTIONALITY

This section provides an overview of functionality that has been deprecated in all minor releases up to this release of Red Hat Ceph Storage.

The Red Hat Storage Console

The Red Hat Storage Console does not support Red Hat Ceph Storage 3. Use the Ansible automation application with the **ceph-ansible** playbooks to install a Red Hat Storage Ceph cluster. For details, see the [Installation Guide for Red Hat Enterprise Linux](#) or [Ubuntu](#).

For cluster monitoring, you can use the Red Hat Ceph Storage Dashboard that provides a monitoring dashboard to visualize the state of a cluster. For details, see the [Monitoring Ceph Clusters with Red Hat Ceph Storage Dashboard](#) section in the Administration Guide.

The **ceph-installer** utility

The **ceph-installer** utility has been deprecated. **ceph-installer** is a command line utility to install and configure Ceph using an HTTP REST API.

CHAPTER 6. KNOWN ISSUES

This section documents known issues found in this release of Red Hat Ceph Storage.

Ansible does not properly handle unresponsive tasks

Certain tasks, for example adding monitors with the same host name, cause the **ceph-ansible** utility to become unresponsive. Currently, there is no timeout set after which the unresponsive tasks is marked as failed. ([BZ#1313935](#))

Certain image features are not supported with the RBD kernel module

The following image features are not supported with the current version of the RADOS Block Device (RBD) kernel module (**krbd**) that is included in Red Hat Enterprise Linux 7.4:

- **object-map**
- **deep-flatten**
- **journaling**
- **fast-diff**

RBDs may be created with these features enabled. As a consequence, an attempt to map the kernel RBDs by running the **rbd map** command fails.

To work around this issue, disable the unsupported features by setting the **rbd_default_features = 1** option in the Ceph configuration file for kernel RBDs or dynamically disable them by running the following command:

```
rbd feature disable <image> <feature>
```

This issue is a limitation only in kernel RBDs, and the features work as expected with user-space RBDs.

NFS Ganesha does not show bucket size or number of blocks

NFS Ganesha, the NFS interface of the Ceph Object Gateway, lists buckets as directories. However, the interface always shows that the directory size and the number of blocks is **0**, even if some data is written to the buckets. ([BZ#1359408](#))

An LDAP user can access buckets created by a local RGW user with the same name

The RADOS Object Gateway (RGW) does not differentiate between a local RGW user and an LDAP user with the same name. As a consequence, the LDAP user can access the buckets created by the local RGW user.

To work around this issue, use different names for RGW and LDAP users. ([BZ#1361754](#))

The GNU tar utility currently cannot extract archives directly into the Ceph Object Gateway NFS mounted file systems

The current version of the GNU tar utility makes overlapping write operations when extracting files. This behavior breaks the strict sequential write restriction in the current version of the Ceph Object Gateway NFS. In addition, GNU tar reports these errors in the usual way, but it also by default continues extracting the files after reporting the errors. As a result, the extracted files can contain incorrect data.

To work around this problem, use alternate programs to copy file hierarchies into the Ceph Object Gateway NFS. Recursive copying by using the `cp -r` command works correctly. Non-GNU archive utilities might be able to correctly extract the tar archives, but none have been verified. ([BZ#1418606](#))

Old zone group name is sometimes displayed alongside with the new one

In a multi-site configuration when a zone group is renamed, other zones can in some cases continue to display the old zone group name in the output of the `radosgw-admin zonegroup list` command.

To work around this issue:

1. Verify that the new zone group name is present on each cluster.
2. Remove the old zone group name:

```
$ rados -p .rgw.root rm zonegroups_names.<old-name>
```

([BZ#1423402](#))

Failover and failback cause data sync issues in multi-site environments

In environments using the Ceph Object Gateway multi-site feature, failover and failback cause data sync to stall. This is because the `radosgw-admin sync status` command reports that **data sync is behind** for an extended period of time.

To work around this issue, use the `radosgw-admin data sync init` command and restart the Gateways. ([BZ#1459967](#))

It is not possible to remove directories stored on S3 versioned buckets by using `rm`

The mechanism that is used to check for non-empty directories prior to unlinking them works incorrectly in combination with the Ceph Object Gateway Simple Storage Service (S3) versioned buckets. As a consequence, directory trees on versioned buckets cannot be recursively removed with a command such as `rm -rf`. To work around this problem, remove any objects in versioned buckets by using the S3 interface. ([BZ#1489301](#))

Deleting directories that contain symbolic links is slow

An attempt to delete directories and subdirectories on a Ceph File System that include a number of hard links by using the `rm -rf` command is significantly slower than deleting directories that do not contain any hard links. ([BZ#1491246](#))

Resized LUNs are not immediately visible to initiators when using the iSCSI gateway

When using the iSCSI gateway, resized logical unit numbers (LUNs) are not immediately visible to initiators. This means the initiators are not able to see the additional space allocated to a LUN. To work around this issue, restart the iSCSI gateway after resizing a LUN to expose it to the initiators, or always add new LUNs when increasing storage capacity. All targets must be updated before utilizing the new space by the initiators. ([BZ#1492342](#))

The Ceph Object Gateway requires applications to write sequentially

The Ceph Object Gateway requires applications to write sequentially from offset 0 to the end of a file. Attempting to write out of order causes the upload operation to fail. To work around this issue, use utilities like `cp`, `cat`, or `rsync` when copying files into NFS space. Always mount with the `sync` option. ([BZ#1492589](#))

The Expiration, Days S3 Lifecycle parameter cannot be set too

The Ceph Object Gateway does not accept the value of `0` for the **Expiration, Days** Lifecycle configuration parameter. Consequently, setting the expiration to `0` cannot be used to trigger background delete operation of objects.

To work around this problem, delete objects directly. ([BZ#1493476](#))

Load on MDS daemons is not always balanced fairly or evenly in multiple active MDS configurations

In certain cases, the MDS balancers offload too much metadata to another active daemon or none at all. ([BZ#1494256](#))

User space issues make `df` calculations less accurate for kernel client users

User space improvements in `df` calculations have been accepted in the upstream kernel, but have not yet been packaged downstream. The `df` command reports more accurate free space data when a Ceph File System is mounted with the `ceph-fuse` utility. When mounted with the kernel client, `df` reports the same, less accurate data as in previous versions. To work around this problem, kernel client users can use the `ceph df` command and examine the relevant data pools to determine free space more accurately. ([BZ#1494987](#))

An iSCSI initiator can send more than `max_data_area_mb` worth of data when a Ceph cluster is under heavy load causing a temporary performance drop

When a Ceph cluster is under heavy load, an iSCSI initiator might send more data than specified by the `max_data_area_mb` parameter. Once the `max_data_area_mb` limit has been reached, the `target_core_user` module returns queue full statuses for commands. The initiators might not fairly retry these commands and they can hit initiator side time outs and be failed in the multipath layer. The multipath layer will retry the commands on another path while other commands are still being executed on the original path. This causes a temporary performance drop, and in some extreme cases in Linux environment the `multipathd` daemon can terminate unexpectedly.

If the `multipathd` daemon crashes, restart it manually:

```
# systemctl restart multipathd
```

([BZ#1500757](#))

The Ceph iSCSI gateway only supports clusters named "ceph"

The Ceph iSCSI gateway expects the default cluster name, that is "ceph". If a cluster uses a different name, the Ceph iSCSI gateway does not properly connect to the cluster. To work around this problem, use the default cluster name, or manually copy the content of the `/etc/ceph/<cluster-name>.conf` file to the `/etc/ceph/ceph.conf` file in addition to the associated keyrings. ([BZ#1502021](#))

The `stat` command returns `ID: 0` for CephFS FUSE clients

When a Ceph File System (CephFS) is mounted as a File System in User Space (FUSE) client, the `stat` command outputs `ID: 0` instead of a proper ID. ([BZ#1502384](#))

Having more than one path from an initiator to an iSCSI gateway is not supported

In the iSCSI gateway, `tcmu-runner` might return the same inquiry and Asymmetric logical unit access (ALUA) info for all iSCSI sessions to a target port group. This can cause the initiator or multipath layer to use the incorrect port info to reference the internal structures for paths and devices, which can result in failures, failover and failback failing, or incorrect multipath and SCSI log or tool output. Therefore, having more than one iSCSI session from an initiator to an iSCSI gateway is not supported. ([BZ#1502740](#))

Incorrect number of tcmu-runner daemons reported after iSCSI target LUNs fail and recover

After iSCSI target Logical Unit Numbers (LUNs) recover from a failure, the `ceph -s` command in certain cases outputs an incorrect number of `tcmu-runner` daemons. ([BZ#1503411](#))

The tcmu-runner daemon does not clean up its blacklisted entries upon recovery

When the path fails over from the Active/Optimized to Active/Non-Optimized path or vice-versa on a failback, the old target is blacklisted to prevent stale writes from occurring. These blacklist entries are not cleaned up after the `tcmu-runner` daemon recovers from being blacklisted, resulting in extraneous blacklisted clients until the entries expire after one hour. ([BZ#1503692](#))

delete_website_configuration cannot be enabled by setting the bucket policy DeleteBucketWebsite

In the Ceph Object Gateway, a user cannot enable `delete_website_configuration` on a bucket even when a bucket policy has been written granting them `S3:DeleteBucketWebsite` permission.

To work around this issue, you can use other methods of permitting, for example, by using admin operations, by bucket owner, or by ACL. ([BZ#1505400](#))

During a data rebalance of a Ceph cluster, the system might report degraded objects

Under certain circumstances, such as when an OSD is marked out, the number of degraded objects reported during a data rebalance of a Ceph cluster can be too high, in some cases implying a problem where none exists. ([BZ#1505457](#))

The iSCSI gateway can fail to scan or setup LUNs

When using the iSCSI gateway, the Linux initiators can return the `kzalloc` failures due to buffers being too large. In addition, the VMWare ESX initiators can return the `READ_CAP` failures due to not being able to copy the data. As a consequence, the iSCSI gateway fails to scan or setup Logical Unit Numbers (LUNs), find or rediscover devices, and add the devices back after path failures. ([BZ#1505942](#))

The RESTful API commands do not work as expected

The RESTful plug-in provides API to interact with a Ceph cluster. Currently, the API fails to change the `pgp_num` parameter. In addition, it indicates a failure when changing the `pg_num` parameter, despite `pg_num` being changed as expected. ([BZ#1506102](#))

Adding LVM-based OSDs fail on clusters with other names than "ceph"

An attempt to install a new Ceph cluster or add OSDs by using the `osd_scenario: lvm` parameter fails on clusters that use other names than the default "ceph". To work around this problem on new clusters, use the default cluster name ("ceph"). ([BZ#1507943](#))

The iSCSI gwc1i utility does not support hyphens in pool or image names

It is not possible to create a disk using a pool or image name that includes hyphens ("-") by using the iSCSI `gwc1i` utility. ([BZ#1508451](#))

Ansible creates unused systemd unit files

When installing the Ceph Object Gateway by using the `ceph-ansible` utility, `ceph-ansible` creates `systemd` unit files for the Ceph Object Gateway host corresponding to all Object Gateway instances located on other hosts. However, only the unit file that corresponds to the hostname of the Ceph Object

Gateway host is active. The rest of the unit files appear inactive, but this does not have any impact on the Ceph Object Gateways. ([BZ#1508460](#))

The `nfs-server` must be disabled on the NFS Ganesha node

When the `nfs-server` service is running on the NFS Ganesha node, an attempt to start the NFS Ganesha instance after its installation fails. To work around this issue, ensure that `nfs-server` is stopped and disabled on the NFS Ganesha node before installing NFS Ganesha. To do so:

```
# systemctl disable nfs-server
# systemctl stop nfs-server
```

([BZ#1508506](#))

Assigning LUNs and hosts to a hostgroup using the `iSCSIgwcli` utility prevents access to the LUNs upon reboot of the iSCSI gateway host

After assigning Logical Unit Numbers (LUNs) and hosts to a hostgroup by using the iSCSI `gwcli` utility, if the iSCSI gateway host is rebooted, the LUN mappings are not properly restored for the hosts. This issue prevents access to the LUNs. ([BZ#1508695](#))

`nfs-ganesha.service` fails to start after a crash or a process kill of NFS Ganesha

When the NFS Ganesha process terminates unexpectedly or it is killed, the `nfs-ganesha.service` daemon fails to start as expected. ([BZ#1508876](#))

The `ms_async_affinity_cores` option does not work

The `ms_async_affinity_cores` option is not implemented. Specifying it in the Ceph configuration file does not have any effect. ([BZ#1509130](#))

Ansible fails to install clusters that use custom group names in the Ansible inventory file

When the default values of the `mon_group_name` and `osd_group_name` parameters are changed in the `all.yml` file, Ansible fails to install a Ceph cluster. To avoid this issue, do not use custom group names in the Ansible inventory file by changing `mon_group_name` and `osd_group_name`. ([BZ#1509201](#))

`lvm` installation scenario does not work when deploying Ceph in containers

It is not possible to use the `osd_scenario: lvm` installation method to install a Ceph cluster in containers. ([BZ#1509230](#))

Compression ratio might not be the same on the destination site as on the source site

When data synced from the source to destination site is compressed, the compression ratio on the destination site might not be the same as on the source site. ([BZ#1509266](#))

`ceph log last` does not display the exact number of specified lines

The `ceph log last <number>` command shows the specified number of lines from the cluster log and cluster audit log, by default located at `/var/log/ceph/<cluster-name>/log` and `/var/log/ceph/<cluster-name>.audit.log`. Currently, the command does not display the exact number of specified lines. To work around this problem, use the `tail -<number> <log-file>` command. ([BZ#1509374](#))

`ceph-ansible` does not properly check for running containers

In an environment where the Docker application is not preinstalled, the **ceph-ansible** utility fails to deploy a Ceph Storage Cluster because it tries to restart **ceph-mgr** containers when deploying the **ceph-mon** role. This attempt fails because the **ceph-mgr** container is not deployed yet. In addition, the **docker ps** command returns the following error:

```
either you don't have docker-client or docker-client-common installed
```

Because **ceph-ansible** only checks if the output of **docker ps** exists, and not its content, **ceph-ansible** misinterprets this result for a running container. When the **ceph-ansible** handler is run later during Monitor deployment, the script it executes fails because no **ceph-mgr** container is found.

To work around this problem, make sure that Docker is installed before using **ceph-ansible**. For details, see the [Getting Docker in RHEL 7](#) section in the Getting Started with Containers guide for Red Hat Enterprise Linux Atomic Host 7. ([BZ#1510555](#))

Object leaking can occur after using `radosgw-admin bucket rm --purge-objects`

In the Ceph Object Gateway, the `radosgw-admin bucket rm --purge-objects` command is supposed to remove all object from a bucket. However, in some cases, some of the objects are left in the bucket. This is caused by the `RGWRados::gc_aio_operate()` operation abandoning on shutdown. To work around this problem, remove the objects by using the `rados rm` command. ([BZ#1514007](#))

The Red Hat Ceph Storage Dashboard cannot monitor iSCSI gateway nodes

The `cephmetrics-ansible` playbook does not install required Red Hat Ceph Storage Dashboard packages on iSCSI gateway nodes. As a consequence, the Red Hat Ceph Storage Dashboard cannot monitor the iSCSI gateways, and the "iSCSI Overview" dashboard is empty. ([BZ#1515153](#))

Ansible fails to upgrade NFS Ganesha nodes

Ansible fails to upgrade NFS Ganesha nodes because the `rolling-update.yml` playbook searches for the `/var/log/ganesha/` directory that does not exist. Consequently, the upgrading process terminates with the following error message:

```
"msg": "file (/var/log/ganesha) is absent, cannot continue"
```

To work around this problem, create `/var/log/ganesha/` manually. ([BZ#1518666](#))

The `--limit mdss` option does not create CephFS pools

When deploying the Metadata Server nodes by using the Ansible and the `--limit mdss` option, Ansible does not create the Ceph File System (CephFS) pools. To work around this problem, do not use `--limit mdss`. ([BZ#1518696](#))

Manual and dynamic resharding sometimes hangs

In the Ceph Object Gateway (RGW), manual and dynamic resharding hangs on a bucket that has versioning enabled. ([BZ#1535474](#))

Resharding a bucket that has ACLs set alters the bucket ACL

In the Ceph Object Gateway (RGW), resharding a bucket with access control list (ACL) set alters the bucket ACL. ([BZ#1536795](#))

Rebooting all Ceph nodes simultaneously will cause an authentication error

When performing a simultaneous reboot of all the Ceph nodes in the storage cluster, a resulting

`client.admin` authentication error will occur when issuing any Ceph-related commands from the command-line interface. To work around this issue, avoid rebooting all Ceph nodes simultaneously. ([BZ#1544808](#))

Purging a containerized Ceph installation using NVMe disks fails

When attempting to purge a containerized Ceph installation using NVME disks, the purge fails because there are a few places where NVMe disk naming is not taken into account. ([BZ#1547999](#))

When using the `rolling_update.yml` playbook to upgrade to Red Hat Ceph Storage 3.0 and from version 3.0 to other zStream releases of 3.0, users who use CephFS must manually upgrade the MDS cluster

Currently the Metadata Server (MDS) cluster does not have built-in versioning or file system flags to support seamless upgrades of the MDS nodes without potentially causing assertions or other faults due to incompatible messages or other functional differences. For this reason, it's necessary during any cluster upgrade to reduce the number of active MDS nodes for a file system to one, first so that two active MDS nodes do not communicate with different versions. Further, it's also necessary to take standbys offline as any new **CompatSet** flags will propagate via the MDSMap to all MDS nodes and cause older MDS nodes to suicide.

To upgrade the MDS cluster:

1. Reduce the number of ranks to 1:

```
ceph fs set <fs_name> max_mds 1
```

2. Deactivate all non-zero ranks, from the highest rank to the lowest, while waiting for each MDS to finish stopping:

```
ceph mds deactivate <fs_name>:<n>
ceph status # wait for MDS to finish stopping
```

3. Take all standbys offline using **systemctl**:

```
systemctl stop ceph-mds.target
ceph status # confirm only one MDS is online and is active
```

4. Upgrade the single active MDS and restart daemon using **systemctl**:

```
systemctl restart ceph-mds.target
```

5. Upgrade and start the standby daemons.

6. Restore the previous `max_mds` for your cluster:

```
ceph fs set <fs_name> max_mds <old_max_mds>
```

For steps on how to upgrade the MDS cluster in a container, refer to the [Updating Red Hat Ceph Storage deployed as a Container Image](#) Knowledgebase article. ([BZ#1550026](#))

Adding a new Ceph Manager node will fail when using the `AnsibleLimit` option

Adding a new Ceph Manager to an existing storage cluster using the Ansible `limit` option, tries to copy the Ceph Manager's keyring without generating it first. This causes the Ansible playbook to fail and the

new Ceph Manager node will not be configured properly. To workaround this issue, do not use the `limit` option while running the Ansible playbook. This will result in a newly generated keyring to be copied successfully. ([BZ#1552210](#))

For Red Hat Ceph Storage deployments running within containers, adding a new OSD will cause the new OSD daemon to continuously restart

Adding a new OSD to an existing Ceph Storage Cluster running within a container, will restart the new OSD daemon every 5 minutes. As a result, the storage cluster will not achieve a `HEALTH_OK` state. Currently, there is no workaround for this issue. This does not affect already running OSD daemons. ([BZ#1552699](#))

Reducing the number of active MDS daemons on CephFS can cause kernel clients I/O to hang

Reducing the number of active Metadata Server (MDS) daemons on a Ceph File System (CephFS) may cause kernel clients I/O to hang. If this happens, kernel clients are unable to connect MDS ranks greater than or equal to `max_mds`. To workaround this issue, raise `max_mds` to be greater than the highest rank. ([BZ#1559749](#))

Adding iSCSI gateways using the `gwcli` tool returns an error

Attempting to add an iSCSI gateway using the `gwcli` tool returns the error:

```
package validation checks - OS version is unsupported
```

To work around this issue, add iSCSI gateways with the parameter `skipchecks=true`. ([BZ#1561415](#))

Initiating the `ceph-ansible` playbook to expand the cluster sometimes fails on nodes with NVMe disks

When `osd_auto_discovery` is set to `true`, initiating the `ceph-ansible` playbook to expand the cluster causes the playbook to fail on nodes with NVMe disks because it is trying to reconfigure disks that are already being used by existing OSDs. This makes it impossible to add a new daemon collocating with an existing OSD that uses NVMe disks when `osd_auto_discovery` is set to `true`. To workaround this issue, configure a new daemon on a new node for which `osd_auto_discovery` is not set to `true`, and use the `--limit` parameter when initiating the playbook to expand the cluster. ([BZ#1561438](#))

`shrink-osd` playbook cannot shrink some OSDs

The `shrink-osd` Ansible playbook does not support shrinking OSDs backed by an NVMe drive. ([BZ#1561456](#))

`tcmu-runner` sometimes logs error messages

The `tcmu-runner` might sporadically log messages such as `Async lock drop` or `Could not break lock`. These logs can be ignored if they are not repeating more often than one time per hour. If the messages occur often, this can be indicative of a network path issue between one or more iSCSI initiators and the iSCSI targets and should be investigated. ([BZ#1564084](#))

Sometimes the `shrink-mon` Ansible playbook fails to remove a monitor from the monmap

The `shrink-mon` Ansible playbook will sometimes fail to remove a monitor from the monmap even though the playbook completes its run successfully. The cluster status shows the monitor intended to be deleted as down. To workaround this issue, launch the `shrink-mon` playbook again with the intention of removing the same monitor, or remove the monitor from the monmap manually. ([BZ#1564117](#))

It is not possible to expand a cluster when using `theosd_scenario: lvm` option

`ceph-ansible` is not idempotent when deploying OSDs using `ceph-volume` and the `lvm_volumes` config option. Therefore, if you deploy a cluster using the `lvm_osd_scenario` option, then you will not be able to expand the cluster. To workaround this issue, remove existing OSDs from the `lvm_volumes` config option so that they will not try to be recreated when deploying new OSDs. Cluster expansion will succeed as expected and create the new OSDs. ([BZ#1564214](#))

Upgrading a node in a Ceph cluster installed with `ceph-test` packages must have `ceph_test = true` in `/etc/ansible/hosts` file

When using the `ceph-ansible rolling_update.yml` playbook to upgrade a Ceph node in a RHEL cluster that was installed with `ceph-test` packages, set `ceph_test = true` in the `/etc/ansible/hosts` file for each node that has `ceph-test` package installed:

```
[mons]
mon_node1 ceph_test=true

[osds]
osd_node1 ceph_test=true
```

Not applicable for clients and MDS nodes. ([BZ#1564232](#))

The `shrink-osd.yml` playbook currently has no support for removing OSDs created by `ceph-volume`

The `shrink-osd.yml` playbook assumes all OSDs are created by `ceph-disk`. As a result, OSDs deployed using `ceph-volume` cannot be shrunk. ([BZ#1564444](#))

Increasing `max_mds` from 1 to 2 sometimes causes CephFS to be in degraded state

When increasing `max_mds` from 1 to 2, if the Metadata Server (MDS) daemon is in the starting/resolve state for a long period of time, then restarting the MDS daemon leads to assert. This causes the Ceph File System (CephFS) to be in degraded state. ([BZ#1566016](#))

Mounting of `nfs-ganesha` file server on a client sometimes fails

Mounting of `nfs-ganesha` file server on a client fails with `Connection Refused` when a containerized IPv6 Red Hat Ceph Storage cluster with an `nfs-ganesha-rgw` daemon is deployed using the `ceph-ansible` playbook. I/Os are then unable to run. ([BZ#1566082](#))

Client I/O sometimes fails for CephFS FUSE clients

Client I/O sometimes fails for Ceph File System (CephFS) as a File System in User Space (FUSE) clients with the error `transport endpoint shutdown` due to assert in the FUSE service. To workaround this issues, unmount and then remount CephFS FUSE, and then start the client I/Os. ([BZ#1567030](#))

The DataDog monitoring utility returns "HEALTH_WARN" even though the cluster is healthy

The DataDog monitoring utility uses the `overall_status` field to determine the health of a cluster. However, `overall_status` is deprecated in Red Hat Ceph Storage 3.0 in favor of the `status` field and therefore always returns the `HEALTH_WARN` error message. Consequently, DataDog reports `HEALTH_WARN` even in cases when the cluster is healthy.

CHAPTER 7. NOTABLE BUG FIXES

This section describes bugs fixed in this release of Red Hat Ceph Storage that have significant impact on users. In addition, it includes descriptions fixed known issues from previous versions.

Improvements in handling of full OSDs

When an OSD disk became so full that the OSD could not function, the OSD terminated unexpectedly with a confusing assert message. With this update:

- The error message has been improved.
- By default, no more than 25% of OSDs are automatically marked as **out**.
- The **statfs** calculation in FileStore or BlueStore back ends have been improved to better reflect the disk usage.

As a result, OSDs are less likely to become full and if they do, a more informative error message is added to the log. ([BZ#1332083](#))

Split threshold is now randomized

Previously, the split threshold was not randomized, so that many OSDs reached it at the same time. As a consequence, such OSDs incurred high latency because they all split directories at once. With this update, the split threshold is randomized which ensures that OSDs split directories over a large period of time. ([BZ#1337018](#))

Mirroring image metadata is supported

Image metadata are now replicated to a peer cluster as expected. ([BZ#1344212](#))

Dynamic feature updates are now replicated

When a feature was disabled or enabled on an already existing image and the image was mirrored to a peer cluster, the feature was not disabled or enabled on the replicated image. With this update, dynamic features updates are replicated as expected. ([BZ#1344262](#))

Disabling image features is no longer incorrectly allowed on non-primary images

With RADOS Block Device (RBD) mirroring enabled, non-primary images are expected to be read-only. Previously, an attempt to disable image features on non-primary images could cause an indefinite wait. This operation is now properly disallowed on non-primary images. As a result, an attempt to disable image features on such images fails with an appropriate error message. ([BZ#1353877](#))

The `rbd bench write` command no longer fails when `--io-size` is equal to the image size

Previously, the `rbd bench-write --io-size <size> <image>` command failed with a segmentation fault if the size specified by the `--io-size` option was greater than 4 GB. With this update, the option is restricted from being too large. ([BZ#1362014](#))

Creating a new pool after manually modifying the CRUSH map and removing a CRUSH ruleset no longer causes issues

Previously, creating a new pool after manually modifying the CRUSH map and removing a CRUSH ruleset caused the newly created pool to use `rule_id` rather than the specified `ruleset`. This led to other issues in the cluster, such as the inability to unprotect snapshots because the newly created pool

was in an incorrect state. The underlying issue has been fixed, and the newly created pools have the correct specified CRUSH ruleset and behave as expected. ([BZ#1369586](#))

AWS SDK for Golang applications work as expected with the Ceph Object Gateway

A bug in the URL processing in the Civetweb HTTP server caused certain kinds of Simple Storage Service (S3) requests to fail. The affected requests included for example a number of requests generated by clients of the Amazon Web Services (AWS) Software Development Kit (SDK) for Golang. Consequently, S3 applications written for AWS SDK for Golang did not interact correctly with the Ceph Object Gateway. This update fixes the handling of absolute URIs in Civetweb, and the AWS SDK for Golang applications work as expected with the Ceph Object Gateway. ([BZ#1387437](#))

The `--rbd-concurrent-management-ops` option works with the `rbd export` command

The `--rbd-concurrent-management-ops` option ensures that image export or import work in parallel. Previously, when `--rbd-concurrent-management-ops` was used with the `rbd export` command, it had no effect on the command performance. The underlying source code has been modified, and `--rbd-concurrent-management-ops` works as expected when exporting images by using `rbd export`. ([BZ#1410923](#))

`rolling_update` no longer sets and unsets flags in between each OSD upgrade

The `rolling_update` playbook of the `ceph-ansible` utility set and unset the `noout`, `noscrub`, and `nodeep-scrub` flags in between each OSD upgrade. If a scrubbing process was scheduled to start shortly or was in progress, setting these flags did not stop scrubbing immediately, and `rolling_update` waited until scrubbing was finished. This process was repeated on each OSD with scheduled scrubbing or scrubbing in progress. This behavior caused the upgrade process to take considerable time to finish. This update ensures that the flags are set before upgrading all OSDs, and are unset after all OSDs are upgraded. ([BZ#1450754](#))

Using IPv6 addressing is now supported with containerized Ceph clusters

Previously, an attempt to deploy a Ceph cluster as a container image failed if IPv6 addressing was used. With this update, IPv6 addressing is supported. ([BZ#1451786](#))

Delete operations are handled during recovery, not peering

When a large number of delete operations were in a client workload, a disk could be easily saturated during peering, which caused very high latency, because the delete operations did not go through the operations queue or do any batching. With this update the delete operations are handled during recovery, instead of peering. ([BZ#1451936](#))

A heartbeat message for Jumbo frames has been added

Previously, if a network included jumbo frames and the maximum transmission unit (MTU) was not configured properly on all network parts, a lot of problems, such as slow requests, and stuck peering and backfilling processes occurred. In addition, the OSD logs did not include any heartbeat timeout messages because the heartbeat message packet size is below 1500 bytes. This update adds a heartbeat message for Jumbo frames. ([BZ#1455711](#))

Upgrading a containerized Ceph cluster by using `rolling_update.yml` is supported

Previously, after upgrading a containerized Ceph cluster by using the `rolling_update.yml` playbook, the `ceph-mon` daemons were not restarted. As a consequence, they were unable to join the quorum after the upgrade. With this update, upgrading containerized Ceph clusters with `rolling_update.yml` works as expected. For details, see the [Upgrading a Red Hat Ceph Storage Cluster That Runs in Containers](#) section in the Container Guide for Red Hat Ceph Storage 3. ([BZ#1458024](#))

OSD activation no longer fails when running the `osd_disk_activate.sh` script in the Ceph container when a cluster name contains numbers

Previously, in the Ceph container image the `osd_disk_activate.sh` script considered all numbers included in a cluster name as an OSD ID. As a consequence, OSD activation failed when running the script because the script was seeking a keyring on a path based on an OSD ID that did not exist. The underlying issue has been fixed, and OSD activation no longer fails when the name of a cluster in a container contains numbers. ([BZ#1458512](#))

Unsupported playbooks are no longer available

The `/usr/share/ceph-ansible/infrastructure-playbooks/` directory no longer includes unsupported playbooks. ([BZ#1461551](#))

New health checks with more structure

Previously, during the installation of a Red Hat Ceph Storage cluster, Ceph raised spurious health warnings. The health checks have been improved to be more structured and no longer trigger health warnings on healthy clusters. ([BZ#1464964](#))

Ceph no longer creates pools by default

Previously, `rbd` pools were created by default upon Ceph cluster creation. This caused several problems, including unnecessary health warnings. Pools are now created only by the user based on their needs rather than by default. ([BZ#1464966](#))

Deleting objects no longer leaves stale bucket index entries

Previously, when objects were removed from the Ceph Object Gateway, the `radosgw` daemon could fail to remove the entries of the deleted objects due to a time scaling error. This bug has been fixed, and `radosgw` removes the bucket index entries as expected. ([BZ#1472874](#))

Large objects are no longer truncated

When creating large objects on large clusters, some of the objects were truncated at 512 KB size. Consequently, an attempt to read such objects failed with **Error 404**. This bug has been fixed, and large objects are no longer truncated. As a result, reading such objects works as expected. ([BZ#1473405](#))

The `--inconsistent-index` option has been restricted

Using the `--inconsistent-index` option with the `radosgw-admin bucket rm` command could cause corruption of the bucket index if the command failed or was stopped. With this update, usage of `--inconsistent-index` requires a confirmation from users (the `--yes-i-really-mean-it` option), and a warning is printed when attempting to use this option. ([BZ#1477311](#))

Restarting `rbd-mirror` is no longer required after a non-orderly shutdown

In RBD mirroring configuration, the local non-primary images could not be force promoted after a non-orderly shutdown of the remote cluster. Consequently, if this happened, and the `rbd-mirror` daemon was not restarted on the local cluster, it was not possible to promote the image because the `rbd-mirror` did not release the exclusive lock. This bug has been fixed, and restarting `rbd-mirror` is no longer required in this case. ([BZ#1479673](#))

Using the `site.yml` playbook with the `--limit` option works as expected

When using the `site.yml` playbook with the `--limit` option set to `osd`, `clients`, or `rgws` to deploy a cluster, the playbook created an incorrect configuration file with missing values. The playbook now uses

the **delegate_facts** option that allows the playbook to instruct hosts to get information from other hosts that are not part of the current play, in this case Monitor hosts. As a result, the playbook creates a proper configuration file in the described scenario. ([BZ#1482067](#))

The number of PGs per OSD is now limited

Previously, it was possible to create pools that included a large number of placement groups (PGs) which could overload the cluster. This update introduces a new configuration option, **mon_max_pg_per_osd**, that limits the number of PGs per OSD to 200. Creating pools or adjusting the **pg_num** parameter now fails if the change would make the number of PGs per OSD exceed the configured limit. You can adjust this option in the Ceph configuration file. In addition, the **mon_pg_warn_max_per_osd** option has been removed. ([BZ#1489064](#))

Slow OSD startup after upgrading to Red Hat Ceph Storage 3.0

Ceph Storage Clusters that have large **omap** databases experience slow OSD startup due to scanning and repairing during the upgrade from Red Hat Ceph Storage 2.x to 3.0. The rolling update may take longer than the specified time out of 5 minutes. Before running the Ansible **rolling_update.yml** playbook, set the **handler_health_osd_check_delay** option to 180 in the **group_vars/all.yml** file. ([BZ#1549293](#))

CHAPTER 8. SOURCES

The updated Red Hat Ceph Storage packages are available at the following locations:

- For Red Hat Enterprise Linux:
<http://ftp.redhat.com/redhat/linux/enterprise/7Server/en/RHCEPH/SRPMS/>
- For Ubuntu: <https://rhcs.download.redhat.com/ubuntu/>