Red Hat Ceph Storage 4.2

Release Notes

Release notes for Red Hat Ceph Storage 4.2

Red Hat Ceph Storage 4.2 Release Notes

Release notes for Red Hat Ceph Storage 4.2

Legal Notice

Abstract

The release notes describes the major features, enhancements, known issues, and bug fixes implemented for the Red Hat Ceph Storage 4.2 product release. These release notes include the previous release notes of the previous Red Hat Ceph Storage 4.2 releases up to the current release. Red Hat is committed to replacing problematic language in our code, documentation, and web properties. We are beginning with these four terms: master, slave, blacklist, and whitelist. Because of the enormity of this endeavor, these changes will be implemented gradually over several upcoming releases. For more details, see our CTO Chris Wright's message .

Table of Contents

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 4.2 contains many contributions from the Red Hat Ceph Storage team. In addition, 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

- Ubuntu Kylin

- Mellanox

- CERN

- Deutsche Telekom

- Mirantis

- SanDisk

- SUSE

CHAPTER 3. NEW FEATURES

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

3.1. THE CEPH ANSIBLE UTILITY

ceph-ansible playbook gathers logs from multiple nodes

With this release, the playbook gathers logs from multiple nodes in a large cluster automatically.

ceph-ansible performs additional connectivity check between two sites

With this update, **ceph-ansible** performs additional connectivity checks between two sites prior to a realm pull.

The purge playbook removes the unused Ceph files

With this release, the purge cluster playbook removes all the Ceph related unused files on the grafana-server node after purging the Red Hat Ceph Storage cluster.

Use the --skip-tags wait_all_osds_up option to skip the check that waits for all the OSDs to be up

With this release, during the upgrade of the storage cluster, by using the **--skip-tags wait_all_osds_up** option at Ansible runtime users can skip this check thereby preventing the failure of the **rolling_update.yml** playbook when you have a disk failure.

crush_rule for existing pools can be updated

Previously, the **crush_rule** value for a specific pool was set during the creation of the pool and it was not possible to update later. With this release, **crush_rule** value can be updated for an existing pool.

Custom crush_rule can be set for RADOS Gateway pools

With this release, RADOS gateway pools can have custom **crush_rule** values in addition to other pools like OpenStack, MDS and Client.

Set ceph_docker_http_proxy and ceph_docker_https_proxy to resolve proxy issues with container registry behind a HTTP(s) proxy

Previously, the environment variables defined in the **/etc/profile.d** directory were not loaded resulting in failure of login and pull registry operations. With this update, by setting the environment variables **ceph_docker_http_proxy** and/or **ceph_docker_https_proxy**, the container registry behind a HTTP(s) proxy works as expected.

Ceph Ansible works with Ansible 2.9 only

Previously, **ceph-ansible** supported 2.8 and 2.9 versions of Ansible as a migration solution. With this release, **ceph-ansible** supports Ansible 2.9 only.

Dashboard is set to HTTPS by default

Previously, the dashboard was set to **http**. With this release, the dashboard is set to **https** by default.

The ceph-mon service is unmasked before exiting the playbook

Previously, during a failure, the **ceph-mon** systemd service would be masked resulting in the playbook failure thereby preventing the service from restarting manually. With this release, the **ceph-mon** service

is unmasked before exiting the playbook during a failure and users can now manually restart the **ceph-mon** service before restarting the rolling update playbook.

3.2. CEPH MANAGEMENT DASHBOARD

View the user's bucket quota usage in the Red Hat Ceph Storage Dashboard

With this release, the Red Hat Ceph Storage Dashboard displays the user's bucket quota usage, including the current size, percentage used, and number of objects.

3.3. CEPH FILE SYSTEM

The mgr/volumes CLI can now be used to list cephx auth IDs

Earlier, **ceph_volume_client** interface was used to list the cephx auth IDs. This interface is now deprecated.

With this release, consumers like Manila can use the mgr/volume interface to list the cephx auth IDs that are granted access to the subvolumes.

Syntax

```
ceph fs subvolume authorized_list _VOLUME_NAME_ _SUB_VOLUME_NAME_ [--group_name=_GROUP_NAME_]
```

3.4. CEPH MANAGER PLUGINS

Internal python to C++ interface is modified to improve Ceph manager performance

Previously, **pg_dump** provided all the information thereby affecting the performance of Ceph manager. With this release, the internal python to C++ interface is modified and the modules provide information on **pg_ready**, **pg_stats**, **pool_stats**, and 'osd_ping_times`.

Progress module can be turned off

Previously, the progress module could not be turned off since it was an **always-on** manager module. With this release, the progress module can be turned off by using **ceph progress off** and turned on by using **ceph progress on**.

3.5. CEPH OBJECT GATEWAY

Ceph Object Gateway's default shard requests on bucket index, `rgw_bucket_index_max_aio`, increased to 128

Previously, outstanding shard requests on a bucket index was limited to 8 causing slow performance with listing buckets. With this release, the default number of shard requests on a bucket index, **rgw_bucket_index_max_aio**, has been increased from 8 to 128, thereby improving the bucket listing performance.

Cluster log information now includes latency information for buckets

Previously, cluster information in logs provided the latency for bucket requests, but did not specify latency information for each bucket. With this release, each line in the log includes the bucket name, object name, request ID, operation start time, and operation name.

This enhancement makes it easier for customers to gather this information when parsing logs. To calculate the latency of the operation, use an **awk** script to subtract the time of the log message from the time the operation started.

The Ceph Object Gateway log includes the access log for Beast

With this release, Beast, the front-end web server, now includes an Apache-style access log line in the Ceph Object Gateway log. This update to the log helps diagnose connection and client network issues.

Explicit request timeout for the Beast front end

Previously, slow client connections, such as clients connected over high-latency networks, might be dropped if they remained idle.

With this release, the new **request_timeout_ms** option in **/etc/ceph.conf** adds the ability to set an explicit timeout for the Beast front end. The default value for **request_timeout_ms** is 65 seconds.

Setting a larger request timeout can make the Ceph Object Gateway more tolerant of slow clients, and can result in fewer dropped connections.

List RGW objects with missing data

Previously, RGW objects that had data erroneously deleted would were unknown to administrators, so they could not determine how best to address this issue. With this release, cluster administrators can use **rgw-gap-list** to list candidate RGW objects that may have missing data.

3.6. MULTI-SITE CEPH OBJECT GATEWAY

▌[RFE] RGW Multi-site data sync optimizations |

TODO https://bugzilla.redhat.com/show_bug.cgi?id=1964097

Data sync logging experienced delays in processing

Previously, data sync logging could be subject to delays in processing large backlogs of log entries.

With this release, data sync includes caching for bucket sync status. The addition of the cache speeds the processing of duplicate datalog entries when a backlog exists.

Multisite sync logging can now use FIFO to offload logging to RADOS data objects

Previously, multisite metadata and data logging configurations used OMAP data logs. With this release, FIFO data logging is available. To use FIFO with green field deployments, set the config option rgw_default_data_log_backing to fifo.

> **NOTE**
>
> Configuration values are case-sensitive. Use **fifo** in lowercase to set config options.

To change the data log backing that sites use, use the command **radosgw-admin --log-type fifo datalog type**.

3.7. RADOS

Ceph messenger protocol revised to msgr v2.1.

With this release, a new version of Ceph messenger protocol, msgr v2.1, is implemented, which addresses several security, integrity and potential performance issues that were with the previous version, msgr v2.0. All Ceph entities, both daemons and clients, now default to msgr v2.1.

Ceph health details are logged in the cluster log

Previously, the cluster log did not have the Ceph health details , so it was difficult to conclude on the root cause of the issue. With this release, the Ceph health details are logged in the cluster log which enables the review of the issues that might arise in the cluster.

Improvement in the efficiency of the PG removal code

Previously,the code was inefficient as it did not keep a pointer to the last deleted object in the placement group (PG) in every pass which caused an unnecessary iteration over all the objects each time. With this release,there is an improved PG deletion performance with less impact on the client I/O. The parameters **osd_delete_sleep_ssd** and **osd_delete_sleep_hybrid** now have the default value of 1 second.

3.8. RADOS BLOCK DEVICES (RBD)

New option **-o noudev** to run commands from a custom network namespace on rbd kernel client

Previously, the commands like **rbd map** and **rbd unmap** from a custom network namespace on the rbd kernel client would hang until manual intervention. With this release, adding the option **o -noudev** to commands like **rbd map -o noudev** and **rbd unmap -o noudev** works as expected. This is particularly useful when using Multus instead of the default OpenShift SDN for networking in OCP.

CHAPTER 4. BUG FIXES

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

4.1. THE CEPH ANSIBLE UTILITY

The **/targets** URL is showing Prometheus as down

Previously, the Prometheus target URL was configured with the **localhost** value. The **localhost** value was causing the Prometheus service not to listen on the **localhost** address when the target status was considered down. With this release the Prometheus configuration file was updated to use the IP address for the target URL value. As a result, the Prometheus target status is reporting correctly.

(BZ#1933560)

ceph-volume might cause metadata corruption when creating OSDs during a Ceph deployment

Previously, when **ceph-volume** issued LVM commands like creating volume groups, logical volumes, setting tags, etc it could result in corrupted metadata when creating OSDs during a Ceph deployment. With this release, users can conditionally disable the **lvmetad** service by setting the **lvmetad_disabled** parameter to **true** in the **group_vars/all.yml** file on the host thereby the metadata corruption is avoided.

(BZ#1955040)

The role **ceph-dashboard** in **ceph-ansible** enforces the common name of the self-signed certificate to **ceph-dashboard**

Previously, when using the self-signed certificates generated by **ceph-ansible**, it enforced the common name (CN) to **ceph-dashboard** thereby causing applications like Prometheus to error out due to the mismatch in the hostname of the node sending certificate to the clients.

With this release, **ceph-ansible** sets the CN with proper values and Prometheus works as expected.

(BZ#1978869)

Rolling upgrade fails when Ceph containers are collocated

The **rolling_update.yml** Ansible playbook fails when the Ceph Monitor and Ceph Object Gateway daemons are collocated with containers, and when the multi-site Ceph Object Gateway is enabled. This failure was caused by the **radosgw-admin** commands not able to execute because of the Ceph Monitor container is stopped during the upgrade process. With this release, the multi-site Ceph Object Gateway code within the **ceph-handler** role is skipped during the upgrade process. As a result, the **rolling_update.yml** Ansible playbook runs successfully.

(BZ#1984880)

Ceph Monitor quorum check fails when switching to a containerized daemons

A regression bug was introduced in the **switch-from-non-containerized-to-containerized-ceph-daemons.yml** Ansible playbook. This regression bug caused the Ceph Monitor quorum check to fail because the current node's host name was not tested. With this release, the **ansible_hostname** fact from the current Ceph Monitor node is used correctly. As a result, the Ceph Monitor quorum check is successful.

(BZ#1990733)

Adding a new Ceph Object Gateway instance when upgrading fails

The **radosgw_frontend_port** option did not consider more than one Ceph Object Gateway instance, and configured port **8080** to all instances. With this release, the **radosgw_frontend_port** option is increased for each Ceph Object Gateway instance, allowing you to use more than one Ceph Object Gateway instance.

(BZ#1859872)

Ceph Ansible removes any socket file and enables redeployment of cluster

Previously, *.asok files were left on completion of purge playbook causing failure during a redeployment of a cluster. With this update, **ceph-ansible** removes any socket file that may be present and the clusters can be redeployed safely.

(BZ#1861755)

Added support for log rotation for the tcmu-runner process in containerized Red Hat Ceph Storage deployments

Previously, when Red Hat Ceph Storage with iSCSI was deployed in containers, there was no log rotation for the tcmu-runner process thereby consuming all the space in the containers. With this release, log rotation support is added to the tcmu-runner process and the log files are periodically rotated resulting in lesser space consumption.

(BZ#1873915)

The FileStore to BlueStore migration process can fail for OSD nodes that have a mix of FileStore OSDs and BlueStore OSDs

Previously, if deployments running Red Hat Ceph Storage versions earlier than 3.2 never had **osd_objectstore** explicitly set in either **group_vars**, **host_vars**, or **inventory**, the deployment had FileStore OSDs. FileStore was the default prior to Red Hat Ceph Storage 3.2.

After upgrading the deployed storage cluster to Red Hat Ceph Storage 3.2, new OSDs added to an existing OSD node would use the BlueStore backend because it became the new default. This resulted in a mix of FileStore and BlueStore OSDs on the same node. In some specific cases, a FileStore OSD might share a journal or DB device with a BlueStore OSD. In such cases, redeploying all the OSDs causes **ceph-volume** errors, either because partitions cannot be passed in **lvm batch** or because of the GPT header.

With this release, there are two options for migrating OSDs with a mix of FileStore and BlueStore configurations:

- Set the extra variable **force_filestore_to_bluestore** to **true** when running the **filestore-to-bluestore.yml** playbook. This setting forces the playbook to automatically migrate all OSDs, even those that already use BlueStore.

- Run the **filestore-to-bluestore.yml** playbook without setting **force_filestore_to_bluestore** (the default is **false**). This causes the playbook to automatically skip the migration on nodes where there is a mix of FileStore and BlueStore OSDs. It will migrate the nodes that have only FileStore OSDs. At the end of the playbook execution, a report displays to show which nodes were skipped.

Before upgrading from Red Hat Ceph Storage 3 to 4, manually examine each node that has been skipped in order to determine the best method for migrating the OSDs.

(BZ#1875777)

Special characters can be set in the Docker registry password

Previously, special characters set in the Docker registry password were not handled correctly. With this release, the Ansible playbook does not fail when special characters are set in the Docker registry password. Special characters can now be used in the Docker registry password and the Ansible playbook works as expected

(BZ#1880252)

ceph-volume Ansible module reports correct information on logical volumes and volume groups

Previously, when applying the command **ceph-volume lvm zap --destroy** command on a Red Hat Enterprise Linux 7 host with Red Hat Enterprise Linux 8 based containers on an OSD, the lvm cache was not refreshed for the host and still reported the logical volumes and volume groups that are present. With this release,**ceph_volume** Ansible module triggers a command on the host to ensure the lvm cache is refreshed and reports correct information on logical volumes and volume groups.

(BZ#1886534)

Ceph container logs can be viewed by using the journalctl command

Previously, Ceph container logs were not present in journald as Podman used Kubernetes file as the default log driver when running containers in detached mode and systemd type forking. With this release, the Ceph containers are configured with journald log driver and the logs are available using the **journalctl** command.

(BZ#1890439)

Ceph Ansible sets file and directory ownership values to nobody:nobody

Previously, Ceph Ansible set the file and directory ownership values to **root:root**. This caused permissions issues for **alertmanager** and **prometheus** services.

With this release, Ceph Ansible sets ownership to **nobody:nobody**. This eliminates the permissions issues.

(BZ#1901543)

4.2. THE COCKPIT CEPH INSTALLER

Ceph installation through the Cockpit no longer fails

Previously, when a user provides **localhost** as the name of the host in the host selection page, the Ceph installation through the Cockpit failed with an error **Systemd must be present**. With this release, the host page UI will display a proper error message and deny the user from using **localhost** as a hostname to continue.

(BZ#1828246)

Cockpit installer produces incorrect values for number of buckets in rgws.yml

Previously, using the Cockpit installer to generate the **rgws.yml** file produced incorrect values for **defaults.rgw.buckets.data:pgnum** and **rgw_override_bucket_index_max_shards**.

With this release, the Cockpit installer creates correct values in **rgws.yml**.

(BZ#1855148)

4.3. CEPH MANAGEMENT DASHBOARD

The prometheus query is fixed to report real-time metrics on CPU usage on the dashboard

Previously, the hosts screen on the Red Hat Ceph Storage dashboard displayed incorrect data about the CPU usage for the nodes in the cluster due to an inaccurate prometheus query. With this release, the prometheus query was fixed and the data on CPU usage gives near real-time metrics.

(BZ#1830375)

Removed options which expect future data from Grafana dashboards embedded in Red Hat Ceph Storage dashboard

Previously, when the option "this week" was selected from "Historical Data" in the overall performance graph, the metrics were not displayed as few historical data options in the Grafana metrics expected future data. With this release,the historical data options that expect future data have been removed and the metrics are displayed as expected.

(BZ#1868638)

Cherrypy no longer reveals its versions in header and error pages

Previously, cherrypy revealed its version in header and error pages. Exposing this information caused a potential security vulnerability. With this release, the headers for both dashboard and Prometheus servers now show Ceph-Dashboard and Ceph-Prometheus instead of the cherrypy version. This change eliminates the security vulnerability.

(BZ#1927719)

4.4. CEPH FILE SYSTEM

Ceph fs status command no longer fails with AttributeError` exception

Previously, Ceph **fs status** command failed with **AttributeError** exception, caused by incorrect handling of metadata during the rejoin. With this release, Ceph **fs status** command returns expected status as unknown objects are now accepted if NoneType is the metadata object type.

(BZ#1884463)

4.5. CEPH MANAGER PLUGINS

The alerts module of Ceph manager works as expected with no health warnings

Previously, Ceph manager would override some module specific config options with default values resulting in the failure of the module with **ALERTS_SMTP_ERROR unable to send alert email** as it continued to use the default value **smtp_ssl=true**. With this release, the default value handling of the Ceph manager is fixed and the alerts module works as expected with no health warnings.

(BZ#1878145)

The restful module API endpoints are now accessible

Previously, the restful module used **dict.iteritems** which is no longer available in python 3. As a result a number of restful module API endpoints were not accessible. With this release, the restful module has been updated to use **dic.items** and the API endpoints are accessible.

(BZ#1897995)

IPv6 addresses were not handled properly in ceph-mgr Prometheus exporter

Previously, some metadata metrics in the ceph-mgr Prometheus exporter output showed incomplete or malformed IPv6 addresses. With this release, the exporter now handles IPv6 addresses correctly.

(BZ#1929064)

4.6. THE CEPH VOLUME UTILITY

Executing the `ceph-volume` command sent debugging output to /var/log/ceph/ceph-volume.log

Previously, executing the **ceph-volume** command always sent debugging-level output to /**var/log/ceph/ceph-volume.log** regardless of the level set in the **--log-level** option. With this release, executing the **ceph-volume** command sends output at the level that the **--log-level** option specifies.

(BZ#1867717)

ceph volume lvm batch reports SSD devices correctly and deploys correct configuration

Previously, **ceph-volume lvm batch** command caused a race condition with **udev** resulting in SSD devices incorrectly reported as HDD, causing unexpected deployment configuration. With the code update in this release, **ceph-volume lvm batch** command reports SSDs correctly and deploys the expected configuration of the cluster.

(BZ#1878500)

Stack update fails after adding new SSD OSDs to the storage cluster

Previously, using stack update to add new SSD OSDs to the same volume group (VG) in the storage cluster caused stack update to fail. This occurred because stack update incorrectly viewed the new SSDs as belonging to different logical volumes (LVs), instead of belonging to the same LV. With this release, stack update views newly added OSDs as belonging to the same LV, and it no longer fails.

(BZ#1892441)

4.7. CEPH OBJECT GATEWAY

Set-lifecycle and delete-lifecycle actions against upgraded OSDs now work as expected

Previously, during an upgrade from Red Hat Ceph Storage 3 to Red Hat Ceph Storage 4.2z2, installing a legacy lifecycle policy against upgraded OSDs yielded a structure decoding error although the set-lifecycle action would succeed. With this release, the changes required to decode bucket lifecycle status entries are fixed and the upgrade of daemons works as expected.

(BZ#1993365)

The --reset-stats option updates buckets in groups for users with large numbers of buckets

Previously, the **radosgw-admin** user **--reset-stats** option simultaneously updated the stats for all buckets owned by a user. For users with very large numbers of buckets, the time required to make the updates could exceed the length of the associated RADOS operation. This could cause Ceph to mark OSDs as down, and could cause the OSDs to flap.

With this release, the **--reset-stats** option updates the stats in groups of 1000 buckets. This allows large numbers of buckets to update without resulting in OSD flapping.

(BZ#1859257)

gc perf counters increments when gc entries are purged from the system

Previously, gc perf counters did not increment when gc entries were purged from the system. With this code update, the correct value of gc perf counter is observed in accordance with the number of gc entries that have been deleted from the system.

(BZ#1882484)

Listing of entries in the last GC object does not enter a loop

Previously, the listing of entries in the last GC object entered a loop because the marker was reset every time for the last GC object. With this release, the truncated flag is updated which does not cause the marker to be reset and the listing works as expected.

(BZ#1884023)

The Ceph Object Gateway syncs bucket cache information during the bucket creation process

Previously, the Ceph Object Gateway did not sync the cache of the bucket information upon bucket creation. This led to the condition when a user tried to access a bucket that did not exist from one RGW and then created that bucket from another Ceph Object Gateway, accessing the bucket from the first Ceph Object Gateway caused a 404 error, stating that the bucket did not exist, although it did. With this update, RGW syncs the cache during the bucket creation process, so that each Ceph Object Gateway can access the bucket.

(BZ#1901973)

KafkaConnect sends objects from a Kafka topic to the RGW S3 bucket

Previously, sending objects from a Kafka topic to the RGW S3 bucket failed because the chunked-encoding object signature was not calculated correctly.

This produced the following error in the RADOS Gateway log:

20 AWSv4ComplMulti: ERROR: chunk signature mismatch

With this release, the chunked-encoding object signature is calculated correctly, allowing KafkaConnect to send objects successfully.

(BZ#1912538)

4.8. MULTI-SITE CEPH OBJECT GATEWAY

Bucket index do not collect entries after syncing on a bucket has been disabled

Previously, the use of **radosgw-admin bucket check --fix …** variable on a bucket in which multi-site syncing has been disabled, would set an incorrect flag indicating syncing has not been disabled. Data would be added to bucket index logs that would not be used or trimmed, thereby consuming more storage over time. With this release, the syncing flag is now copied correctly when running the **radosgw-admin bucket check --fix …** command. Bucket index logs do not collect entries after syncing on a bucket is disabled.

(BZ#1894702)

4.9. RADOS

The Progress module is no longer stuck for an indefinite time

Previously, the progress vents in Ceph status were stuck for an indefinite time. This was caused by the Progress module checking the PG state early and not syncing with the epoch of the OSDMap. With this release, progress events now pop up as expected.

(BZ#1826224)

Progress module causing Ceph Monitor crashes

During backfill and recovery operations, the progress module can generate negative progress events. With large storage clusters, too many negative progress events can lead to large memory allocation on the Ceph Monitor nodes, causing cascading Ceph Monitor crashes. With this release, the code ensures that progress events are not negative. As a result, the progress module does not cause the Ceph Monitors to crash.

(BZ#1997838)

Ceph Monitors now show caution when handling forwarded OSD failure reports

Previously, Ceph Monitors would incorrectly report slow operations to administrators and logging systems. With this release, Ceph Monitors now handle forwarded OSD failure reports with caution, so many fewer inaccurate slow operations will occur.

(BZ#1866257)

Monitor trims osdmaps appropriately

Previously, the monitor failed to trim stale osdmaps although they were used by "out" OSDs because the monitor considered both "in" and "out" OSDs when trimming osdmaps. With this release, the "out" OSDs are not considered and the osdsmaps are trimmed appropriately.

(BZ#1875628)

BlueStore and FileStore OSDs list objects in the same order in a mixed cluster

Previously, in a cluster with both BlueStore and FileStore OSDs, deep scrub and backfill could report missing objects due to an inconsistency of the sorting mechanism in the backend. With this upgrade, a feature flag **OSD_FIXED_COLLECTION_LIST** has been added to ensure that the **collection_list** method in BlueStore lists objects in the same order as FileStore.

(BZ#1880188)

Log files were created with incorrect permissions

Previously, a code addition changed the order in which relevant functions were called. This caused some daemons to create log files with incorrect permissions. With this release, functions are called in the correct order, and the daemons create log files with the correct permissions.

(BZ#1884469)

Enabling **bluefs_buffered_io** prevents performance degradation

Previously, the option **bluefs_buffered_io** was disabled which led to slow operations of RocksDB and OMAP interactions in certain scenarios.With this release the option **bluefs_buffered_io** is set to **True** , thereby preventing performance degradation.

(BZ#1930264)

Memory growth because of onodes is now controlled

Previously, the default value for the option **bluestore_cache_trim_max_skip_pinned** was 64 which was very low for large clusters. Since this option controlled the rate of trimming onodes and with current default value, it could have led to a buildup of onodes causing memory growth. With this release, the default value of bluestore_cache_trim_max_skip_pinned` is 1000 and the memory growth is controlled.

(BZ#1947215)

4.10. RADOS BLOCK DEVICES (RBD)

Applications using librbd works when client-side QoS throttling is enabled

Previously, when client-side QoS throttling was enabled in librbd, it could crash because the data path was not properly protected with a lock. With this release, the missing lock is added and the applications using librbd for IO works as expected when client-side QoS throttling is enabled.

(BZ#1878268)

4.11. NFS GANESHA

All file layouts in READDIR return results

Previously, some file layouts caused a loop in READDIR and never returned results. With this update, READDIR works as expected and returns the results correctly.

(BZ#1845501)

CHAPTER 5. KNOWN ISSUES

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

5.1. THE CEPH ANSIBLE UTILITY

Deploying the placement group autoscaler does not work as expected on CephFS related pools only

To work around this issue, the placement group autoscaler can be manually enabled on CephFS related pools after the playbook has run.

(BZ#1836431)

Ceph OSD fails to use `osd_max_markdown_count` parameter as the systemd unit template enforces the parameter `Restart=always`

The systemd unit templates of OSD Daemons enforce the parameter **Restart=always** thereby preventing the use of **osd_max_markdown_count** parameter resulting in the restarting of services. To workaround this issue, use the **ceph_osd_systemd_overrides** variable to override the **Restart=** parameter in the OSD systemd template, that is,

```
[osds]
osd0 ceph_osd_systemd_overrides="{'Service': {'Restart': 'no'}}"
```

(BZ#1860739)

The `filestore-to-bluestore` playbook does not support the `osd_auto_discovery` scenario

Red Hat Ceph Storage 4 deployments based on **osd_auto_recovery** scenario can't use the **filestore-to-bluestore** playbook to ease the **BlueStore** migration.

To work around this issue, use **shrink-osd** playbook and redeploy the shrinked OSD with **osd_objectstore: bluestore**.

(BZ#1881523)

The upgrade process does not automatically stop the `ceph-crash` container daemons.

The upgrade process issues a call to the role **ceph-crash**, but the call only starts the **ceph-crash** service. If the **ceph-crash** container daemons are still running during the upgrade process, they will not be restarted when upgrade is complete.

To work around this issue, manually restart the **ceph-crash** containers after upgrading.

(BZ#1943471)

5.2. THE CEPH VOLUME UTILITY

When users run `osd.yml` or `site.yml` playbook, ceph-ansible does not create OSDs on the new devices

When users explicitly pass a set of db devices,**--db-devices** or wal devices **--wal-devices** where one is unavailable to the **ceph-volume lvm batch**, it is then filtered out and the results are different then what is expected. Current implementation of **ceph-volume lvm batch** does not allow the adding of new OSDs

in a non-interactive mode if one of the passed db or wal device is unavailable to prevent an expected OSD topology. Due to this ceph-volume limitation, ceph-ansible is unable to add new OSDs in the batch scenario of devices and dedicated_devices.

(BZ#1896803)

5.3. MULTI-SITE CEPH OBJECT GATEWAY

Objects fail to sync in a Ceph Object Gateway multisite set-up

Some objects may fail to sync and might have a status mismatch when users run **radosgw-admin sync status** command in a Ceph Object Gateway multisite set-up.

Currently, there is no workaround for this issue.

(BZ#1905369)

CHAPTER 6. 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.

Ubuntu is no longer supported

Installing a Red Hat Ceph Storage 4 cluster on Ubuntu is no longer supported. Use Red Hat Enterprise Linux as the underlying operating system.

Configuring iSCSI gateway using **ceph-ansible** is no longer supported

Configuring the Ceph iSCSI gateway by using the **ceph-ansible** utility is no longer supported. Use **ceph-ansible** to install the gateway and then use the **gwcli** utility of the to configure the Ceph iSCSI gateway. For details, see the *The Ceph iSCSI Gateway* chapter in the *Red Hat Ceph Storage Block Device Guide*.

ceph-disk is deprecated

With this release, the **ceph-disk** utility is no longer supported. The **ceph-volume** utility is used instead. For details, see the *Why does* **ceph-volume** *replace* `ceph-disk` section in the *Administration Guide* for Red Hat Ceph Storage 4.

FileStore is no longer supported in production

The FileStore OSD back end is now deprecated because the new BlueStore back end is now fully supported in production. For details, see the *How to migrate the object store from FileStore to BlueStore* section in the *Red Hat Ceph Storage Installation Guide*.

Ceph configuration file is now deprecated

The Ceph configuration file (**ceph.conf**) is now deprecated in favor of new centralized configuration stored in Ceph Monitors. For details, see the *The Ceph configuration database* section in the *Red Hat Ceph Storage Configuration Guide*.

6.1. THE CEPH VOLUME UTILITY

Deploying an OSD on a device setup for **dm-cache** is unsupported

Deploying an OSD with **dm-cache** is no longer supported. Use the Bluestore backend instead of **dm-cache**.

For more information, see the *How to migrate the object store from FileStore to BlueStore* section in the *Red Hat Ceph Storage Installation Guide*.

CHAPTER 7. SOURCES

The updated Red Hat Ceph Storage source code packages are available at the following location:

- For Red Hat Enterprise Linux 7:
  http://ftp.redhat.com/redhat/linux/enterprise/7Server/en/RHCEPH/SRPMS/

- For Red Hat Enterprise Linux 8:
  http://ftp.redhat.com/redhat/linux/enterprise/8Base/en/RHCEPH/SRPMS/