Red Hat Ceph Storage 1.3.3 Release Notes

Release notes for Red Hat Ceph Storage 1.3.3

Red Hat Ceph Storage Documentation Team

Red Hat Ceph Storage 1.3.3 Release Notes

Release notes for Red Hat Ceph Storage 1.3.3

Legal Notice

Abstract

The Release Notes document describes the major features and enhancements implemented in Red Hat Ceph Storage and known issues and notable bug fixes in this 1.3.3 release.

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.

CHAPTER 2. ACKNOWLEDGMENTS

This version of Red Hat Ceph Storage 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.

"ceph" rebased to 0.94.9

The **ceph** packages have been updated to the upstream version 0.94.9, which provides a number of bug fixes and enhancements over the previous version.

The "radosgw-agent fully sync" mode now copies only changed objects

With this update, the **radosgw-agent** in **full sync** mode copies only objects that have changed, instead of all of them, when synchronizing objects between two Ceph Object Gateways after a failover.

The default quota for a user is set when creating the user

With this update, when creating a new Ceph Object Gateway user, the default quota is set for the user. Previously, the default quota for a user was not set until the user performed an action in the Ceph Object Gateway.

"ceph-deploy" rebased to 1.5.36

The **ceph-deploy** package has been updated to the upstream version 1.5.36, which provides a number of bug fixes and enhancements over the previous version.

"ceph-deploy" now has its own ceph-deploy-${cluster_name}.log file

The **ceph-deploy** utility has now its own **ceph-deploy-${cluster-name}.log** file. Previously, the **${cluster-name}.log** file was used to log **ceph-deploy** entries, which could lead to confusion because this file is also used to log entries related to the Ceph cluster.

Transition to Phase-2 Production support

Red Hat will deliver continued support of Ceph Storage 1.3 for 36 months since its original release, until June 30th, 2018. Having met our original commitment to provide 12 months of Production Phase 1 support, this release marks the transition of Red Hat Ceph Storage 1.3 to Production Phase 2 support.

The details of the Red Hat Ceph Storage's life cycle are available at https://access.redhat.com/articles/1372203. If you need any technical assistance, contact our Global Support team by using the Red Hat Customer Portal.

CHAPTER 4. KNOWN ISSUES

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

The "diamond" service sometimes does not start

After installing Red Hat Ceph Storage and connecting the Ceph nodes to the Calamari server, in some cases, the **diamond** service does not start on certain Ceph nodes despite the **diamond** package being installed. As a consequence, graphs in the Calamari UI are not generated for such nodes.

To work around this issue, run the following command from the administration node as root or with **sudo**:

```
salt '*' state.highstate
```

(BZ#1378826)

Various race conditions occur when using "ceph-disk"

The process of preparing and activating disks using the **ceph-disk** utility involves also usage of the **udev** device manager and the **systemd** service manager. Consequently, various race conditions can occur causing the following problems:

- Mount points for OSDs are duplicated
- **ceph-disk** fails to activate a device even though the device has been successfully prepared with the **ceph-disk prepare** command
- Some OSDs are not activated at boot time

To work around these issues, perform the steps below:

1. Manually remove the **udev** rules by running the following command as **root**:

   ```
   # rm /usr/lib/udev/rules.d/95-ceph-osd.rules
   ```

2. Prepare the disks:

   ```
   $ ceph-disk prepare
   ```

3. Add the "ceph-disk activate-all" string to the **/etc/rc.local** file. Run the following command as **root**:

   ```
   # echo "ceph-disk activate-all" | tee -a /etc/rc.local
   ```

4. Reboot the system or activate the disks by running the following command as **root**:

   ```
   # ceph-disk activate-all
   ```

   (BZ#1300703)

Graphs for monitor hosts are not displayed

Graphs for monitor hosts are not displayed in the Calamari server GUI when selecting them from the **Graphs** drop-down menu. (BZ#1223335)

Deleting large RBD images with "object map" can cause OSDs to crash

An attempt to delete a large RBD image with the **`object map`** feature enabled can cause the OSD nodes to trigger the **`suicide_timeout`** and self-terminate.

To work around this issue, remove the image manually instead of using the **`rbd rm`** command. To avoid this issue, do not enable **`object map`** on images that are hundreds of terabytes in size. (BZ#1325322)

Missing Calamari graphs

After installing a new Ceph cluster and initializing the Calamari server, Calamari graphs can be missing. The graphs are missing on any node connected to Calamari after the **`calamari-ctl initialize`** command was run. To work around this issue, run the **`calamari-ctl initialize`** and **`salt '*' state.highstate`** commands after connecting additional Ceph cluster nodes to Calamari. These commands can be run multiple times without any issues. (BZ#1273943)

Missing password field in Calamari

An attempt to create a new user in the Calamari UI fails with the **`password field is required`** error because the password field is missing from the Calamari UI. To work around this issue, use the Raw data form and create the new user by using the JSON format, for example:

```
{
    "username": "user",
    "email": "user@example.com",
    "password": "password"
}
```

(BZ#1305478)

Upstart cannot stop or restart the initial`ceph-mon` process on Ubuntu

When adding a new monitor on Ubuntu, either manually or by using the **`ceph-deploy`** utility, the initial **`ceph-mon`** process cannot be stopped or restarted using the Upstart init system. To work around this issue, use the **`pkill`** utility or reboot the system to stop the **`ceph-mon`** process. Then, it is possible to restart the process using Upstart as expected. (BZ#1255497)

"ice_setup" returns various errors when installing Calamari on Ubuntu

When running the **`ice_setup`** utility on Ubuntu, the utility returns various warnings, similar to the ones in the output below:

```
apache2_invoke: Enable configuration javascript-common
invoke-rc.d: initscript apache2, action "reload" failed.
apache2_invoke: Enable module wsgi
Adding user postgres to group ssl-cert
Creating config file /etc/logrotate.d/postgresql-common with new
```

```
version
Building PostgreSQL dictionaries from installed myspell/hunspell
packages...
Removing obsolete dictionary files:
 * No PostgreSQL clusters exist; see "man pg_createcluster"
ERROR: Module version does not exist!
salt-master: no process found
```

These warnings do not affect the installation process of Calamari and can be safely ignored. (BZ#1305133)

The "Update" button is not disabled when a check box is cleared

In the Calamari web UI on the **Manage** > **Cluster Settings** page, the **Update** button is not disabled when a check box is cleared. Moreover, further clicking on the **Update** button displays an error dialog box, which leaves the button unusable. To work around this issue, reload the page as suggested in the error dialog. (BZ#1223656)

Creating encrypted OSD sometimes fails

When creating encrypted OSD nodes by using the **`--dmcrypt`** option with the **`ceph-deploy`** utility, the underlying **`ceph-disk`** utility fails to create second journal partition on same journal device that is used by another OSD. (BZ#1327628)

CHAPTER 5. NOTABLE BUG FIXES

This section describes bugs fixed in this release of Red Hat Ceph Storage that have significant impact on users.

Calamari now correctly handles manually added OSDs that do not have "ceph-osd" running

Previously, when OSD nodes were added manually to the Calamari server but the **`ceph-osd`** daemon was not started on the nodes, the Calamari server returned error messages and stopped updating statuses for the rest of the OSD nodes. The underlying source code has been modified, and Calamari now handles such OSDs properly. (BZ#1360467)

OSDs no longer reboot when corrupted snapsets are found during scrubbing

Previously, Ceph incorrectly handled corrupted snapsets that were found during scrubbing. This behavior caused the OSD nodes to terminate unexpectedly every time the snapsets were detected. As a consequence, the OSDs rebooted every few minutes. With this update, the underlying source code has been modified, and OSDs no longer reboots in the described situation. (BZ#1273127)

OSD now deletes old OSD maps as expected

When new OSD maps are received, the OSD daemon marks the unused OSD maps as **`stale`** and deletes them to keep up with the changes. Previously, an attempt to delete stale OSD maps could fail for various reasons. As a consequence, certain OSD nodes were sometimes marked as **`down`** if it took too long to clean their OSD map caches when booting. With this update, the OSD daemon deletes old OSD maps as expected, thus fixing this bug. (BZ#1291632)

%USED now shows correct value

Previously, the **`%USED`** column in the output of the **`ceph df`** command erroneously showed the size of a pool divided by the raw space available on the OSD nodes. With this update, the column correctly shows the space used by all replicas divided by the raw space available on the OSD nodes. (BZ#1330643)

SELinux no longer prevents "ceph-mon" and "ceph-osd" from accessing /var/lock/ and /run/lock/

Due to insufficient SELinux policy rules, SELinux denied the **`ceph-mon`** and **`ceph-osd`** daemons to access files in the **`/var/lock/`** and **`/run/lock/`** directories. With this update, SELinux no longer prevents **`ceph-mon`** and **`ceph-osd`** from accessing **`/var/lock/`** and **`/run/lock/`**. (BZ#1330279)

The QEMU process no longer hangs when creating snapshots on images

When the RADOS Block Device (RBD) cache was enabled, creating a snapshot on an image with active I/O operations could cause the QEMU process to become unresponsive. With this update, the QEMU process no longer hangs in the described scenario. (BZ#1316287)

"ceph-deploy" now correctly removes directories of manually added monitors

Previously, an attempt to remove a manually added monitor node by using the **ceph-deploy mon destroy** command failed with the following error:

```
UnboundLocalError: local variable 'status_args' referenced before
assignment"
```

The monitor was removed despite the error, however, **ceph-deploy** failed to remove the monitor configuration directory located in the **/var/lib/ceph/mon/** directory. With this update, **ceph-deploy** removes the monitor directory as expected. (BZ#1278524)

The least used OSDs are selected for increasing the weight

With this update, the least used OSD nodes are now selected for increasing the weight during the **reweight-by-utilization** process. (BZ#1333907)

OSDs are now selected properly during "reweight-by-utilization"

During the **reweight-by-utilization** process, some of the OSD nodes that met the criteria for reweighting were not selected. The underlying algorithm has been modified, and OSDs are now selected properly during **reweight-by-utilization**. (BZ#1331764)

OSDs no longer receive unreasonably large weight during "reweight-by-utilization"

When the value of the **max_change** parameter was greater than an OSD weight, an underflow occurred. Consequently, the OSD node could receive an unreasonably large weight during the **reweight-by-utilization** process. This bug has been fixed, and OSDs no longer receive large weight in the described situation. (BZ#1331523)

OSDs no longer crash when using "rados cppool" to copy an "omap" object

The **omap** objects cannot be stored in an erasure-coded pool. Previously, copying the **omap** objects from a replicated pool to an erasure-coded pool by using the **rados cppool** command caused the OSD nodes to terminate unexpectedly. With this update, the OSD nodes return an error message instead of crashing in the described situation. (BZ#1368402)

Listing versioned buckets no longer hangs

Due to a bug in the bucket listing logic, the **radosgw-admin bucket list** and **radosgw-admin bucket stats** commands could become unresponsive while attempting to list versioned buckets or get their statistics. This bug has been fixed, and listing versioned buckets no longer hangs in the described situation. (BZ#1322239)

Ceph Object Gateway now properly uploads files to erasure-coded pools

Under certain conditions, Ceph Object Gateway did not properly upload files to an erasure-coded pool by using the SWIFT API. Consequently, such files were broken and an attempt to download them failed with the following error message:

```
ERROR: got unexpected error when trying to read object: -2
```

The underlying source code has been modified, and Ceph Object Gateway now properly uploads files to erasure-coded pools. (BZ#1369013)

The `ceph osd tell` command now prints correct error message

When the deprecated **`ceph osd tell`** command was executed, the command returned a misleading error message. With this update, the error message is correct. (BZ#1193710)

"filestore_merge_threshold" can be set to a negative value as expected

If the **`filestore_merge_threshold`** parameter is set to a negative value, merging of subdirectories is disabled. Previously, an attempt to set **`filestore_merge_threshold`** to a negative value by using the command line failed and an error message similar to the following one was returned:

```
"error": "error setting 'filestore_merge_threshold' to '-40': (22)
Invalid argument"
```

As a consequence, it was not possible to disable merging of subdirectories. This bug has been fixed, and **`filestore_merge_threshold`** can now be set to a negative value as expected. (BZ#1284696)

"radosgw-admin region-map set" output includes the bucket quota

Previously, the output of the **`radosgw-admin region-map set`** command did not include the bucket quota, which led to confusion if the quota was properly set. With this update, the **`radosgw-admin region-map set`** output includes the bucket quota as expected. (BZ#1349484)

The form of the "by-parttypeuuid" term is now correct

The **`ceph-disk(8)`** manual page and the **`ceph-disk`** python script now include the correct form of the **`by-parttypeuuid`** term. Previously, they included **`by-parttype-uuid`** instead. (BZ#1335564)

Index files are removed as expected after deleting buckets

Previously, when deleting buckets, the buckets' index files remained in the **`.rgw.buckets.index`** file. With this update, the index files are removed as expected. (BZ#1340496)

"ceph df" now shows proper value of "MAX AVAIL"

When adding a new OSD node to the cluster by using the **`ceph-deploy`** utility with the **`osd_crush_initial_weight`** option set to **`0`**, the value of the **`MAX AVAIL`** field in the output of the **`ceph df`** command was **`0`** for each pool instead of the proper numerical value. As a consequence, other applications using Ceph, such as OpenStack Cinder, assumed that there is no space available to provision new volumes. This bug has been fixed, and **`ceph df`** now shows proper value of **`MAX AVAIL`** as expected. (BZ#1306842)

The columns in the "rados bench" command output are now separated correctly

This update ensures that the columns in the **`rados bench`** command output are separated correctly. (BZ#1332470)

OSDs now obtain PID files properly during an upgrade

After upgrading from Red Hat Ceph Storage 1.2 to 1.3, some of the OSD daemons did not obtain PID files properly. As a consequence, such OSDs could not be restarted or stopped by using SysVinit commands and therefore could not be upgraded to the newer version. This update ensures that OSDs obtain PID files properly during an upgrade. As a result, OSDs are upgraded to newer versions as expected. (BZ#1299409)

The default value of "osd_scrub_thread_suicide_timeout" is now 300

The **`osd_scrub_thread_suicide_timeout`** configuration option ensures that poorly behaving OSD nodes self-terminate instead of running in degraded states and slowing traffic. Previously, the default value of **`osd_scrub_thread_suicide_timeout`** was set to 60 seconds. This value was not sufficient when scanning data for objects on extremely large buckets. This update increases the default value of **`osd_scrub_thread_suicide_timeout`** 300. (BZ#1300539)

PG collection split no longer produces any orphaned files

Due to a bug in the underlying source code, a placement group (PG) collection split could produce orphaned files. Consequently, the PG could be incorrectly marked as inconsistent during scrubbing, or the OSD nodes could terminate unexpectedly. The bug has been fixed, and PG collection split no longer produces any orphaned files. (BZ#1334534)

The bucket owner is now properly changed

Previously, the bucket owner was not properly changed by using the **`radosgw-admin bucket unlink`** and **`radosgw-admin bucket link`** commands. As a consequence, the new owner was not able to access the bucket. The underlying source code has been modified, and the bucket owner is now properly changed as expected. (BZ#1324497)

The monitor nodes exit gracefully after authenticating with an incorrect keyring

When a new cluster included monitor nodes that were previously a part of an another cluster, the monitor nodes terminated with a segmentation fault when attempting to authenticate with an incorrect keyring. With this update, the monitor nodes exit gracefully instead of crashing in the described scenario. (BZ#1312587)

CHAPTER 6. SOURCES

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

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