Red Hat Ceph Storage 1.3 Release Notes

Ceph Storage v1.3 release notes.

Red Hat Customer Content Services

Red Hat Ceph Storage 1.3 Release Notes

Ceph Storage v1.3 release notes.

Legal Notice

Abstract

The Release Notes document describes the major features and enhancements implemented in Red Hat Ceph Storage and the known issues in this v1.3 release.

Table of Contents

CHAPTER 1. ACKNOWLEDGMENTS

Red Hat Ceph Storage v1.3 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, and
- SanDisk.

CHAPTER 2. OVERVIEW

Red Hat Ceph Storage v1.3 is the second release of Red Hat Ceph Storage. New features for Ceph Storage include:

2.1. PACKAGING

For organizations that require highly secure clusters, Red Hat Ceph Storage ships with an ISO-based installation so that you can deploy Ceph without a connection to the internet. For organizations that allow the Ceph cluster to connect to the internet, Red Hat Ceph Storage supports a CDN-based installation (RHEL only).

For RHEL 7 both ISO-based and CDN-based installation are available.

For Ubuntu 14.04 currently only ISO-based installation is available. The first point release of Red Hat Ceph Storage v1.3 for Ubuntu 14.04 will introduce an online repository based installation.

Red Hat Ceph Storage v1.3 for RHEL 7 ships with two Stock Keeping Units (SKUs).

- **Red Hat Ceph Storage for Management Nodes**: The repositories for this SKU provide access to the installer, Calamari and Ceph monitors. You may use this SKU on up to six physical nodes.
- **Red Hat Ceph Storage**: The repository for this SKU provides access to OSDs. You will need one SKU for each node containing Ceph OSDs.

For CDN-based installations, you will need to attach pools for for these SKUs. See the Installation Guide for details.

Red Hat Ceph Storage v1.3 for RHEL 7 has the following repositories:

- **`rhel-7-server-rhceph-1.3-calamari-rpms`** contains the Calamari repository.
- **`rhel-7-server-rhceph-1.3-installer-rpms`** contains the **`ceph-deploy`** repository.
- **`rhel-7-server-rhceph-1.3-mon-rpms`** contains the **`ceph-mon`** daemon.
- **`rhel-7-server-rhceph-1.3-osd-rpms`** contains the **`ceph-osd`** daemon.
- **`rhel-7-server-rhceph-1.3-tools-rpms`** contains the **`ceph`** CLI tools, Ceph Block Device kernel for RHEL 7.1 and higher, and the Ceph Object Gateway.

You will need to enable these repositories on various hosts. For details about installation on RHEL 7, see RHCS v1.3 Installation Guide for RHEL (x86_64).

For details about installation on Ubuntu 14.04, see RHCS v1.3 Installation Guide for Ubuntu (x86_64).

2.2. CEPH STORAGE CLUSTER

The Ceph Storage Cluster has a number of new features and improvements.

- **Monitor Performance:** Ceph monitors now perform writes to the local data store asynchronously, improving overall responsiveness.
- **Cache Tiering (Tech Preview):** Cache tiering has some of the same overhead as the underlying storage tier, so it performs best under certain conditions. A series of changes have been made in the cache tiering code that improve performance and reduce latency; namely,

objects are not promoted into the cache tier by a single read; instead, they must be found to be sufficiently *hot* before getting promoted to the cache tier.

- **New Administrator Commands:** The **`ceph osd df`** command shows pertinent details on OSD disk utilizations. The **`ceph pg ls ...`** command makes it much simpler to query PG states while diagnosing cluster issues.
- **Local Recovery Codes (Tech Preview):** the OSDs now support an erasure-coding scheme that stores some additional data blocks to reduce the IO required to recover from single OSD failures.
- **Degraded vs Misplaced:** The Ceph health reports from **`ceph -s`** and related commands now make a distinction between data that is degraded (there are fewer than the desired number of copies) and data that is misplaced (stored in the wrong location in the cluster). The distinction is important because the latter does not compromise data safety.
- **Recovery Tools:** The **`ceph-objectstore-tool`** allow you to mount an offline OSD disk, retrieve PGs and objects and manipulate them for debugging and repair purposes. Red Hat Ceph Storage support personnel are the heaviest users of this tool. Consult Red Hat Ceph Storage support before using the **`ceph-objecstore-tool`**.
- **CRUSH Improvements:** We have added a new **`straw2`** bucket algorithm that reduces the amount of data migration required when changes are made to the cluster.
- **OSD SSD Optimization:** Ceph Storage v1.3 provides some optimization which results in less CPU overhead per operation and thus more operations. This improvement is relevant for fast hardware such as SSDs. If you experienced CPU bound operations with SSDs before, you should see an improvement. However, this will not address network bottlenecks.
- **Time-scheduled Scrubbing:** Ceph Storage v1.3 supports **`osd_scrub_begin_hour`** and **`osd_scrub_end_hour`** time ranges as allowable hours for scrubbing. Ceph ignores these settings if the OSD exceeds **`osd_scrub_max_interval`**.

Erasure-coding and cache tiering are tech previews only and are not supported for production clusters.

2.3. CEPH BLOCK DEVICE

- **Mandatory Exclusive Locks:** The mandatory locking framework (disabled by default) adds additional safeguards to prevent multiple clients from using the same image simultaneously. See the **`rbd --help`** interface for additional usage and the Ceph Architecture Guide for architecture details.
- **Copy-on-Read Cloning:** Copy-on-read for image clones improves performance for some workloads. For example, when used with OpenStack the copy happens when you read data, and thereby happens a bit faster, reducing flattening time and allowing graduating to a parent clone. Copy-on-read is disabled by default, but you may enable it by setting **`rbd_clone_copy_on_read=true`** in your Ceph configuration.
- **Object Maps:** Ceph Block Device now has an object map function that tracks which parts of the image are actually allocated (i.e., block devices are thin provisioned, so this shows the index of objects that actually exist). Object maps are valuable with clones that get objects form a parent, as they improve performance for clones when re-sizing and importing, exporting or flattening. Object maps are off by default, but you can enable them in your Ceph configuration by specifying **`rbd default format = 2`** and **`rbd default features = X`**, where **`X`** is the sum of the feature bits.

- **Read-ahead**: Ceph Block Device supports read ahead, which provides a small difference in improvement for virtio (e.g., 10%), but it doubles the improvement for IDE.
- **Allocation Hinting:** The allocation hints are to prevent fragmentation on the filesystem beneath the OSD. The block device knows the size of its objects, so it sends that size as an allocation hint with write operations so the OSD reserves that amount of space for the object. With additional writes, the object will be sequential when it is fully written in the filesystem. This prevents performance degradation from fragmentation. Allocation hinting is on by default.
- **Cache Hinting:** Ceph Block Device supports cache hinting, which makes more efficient use of the client side cache or cache tiering by making import/export slightly more efficient in Red Hat Ceph Storage v1.3.

2.4. CEPH OBJECT GATEWAY

- **Civetweb Installation:** The **`ceph-deploy`** tool now has a new **`ceph-deploy rgw create <HOST>`** command that quickly deploys an instance of the S3/Swift gateway using the embedded Civetweb server (defaulting to port **`7480`**). The new installation method dramatically simplifies installation and configuration compared to Apache and FastCGI. Presently, it only supports HTTP. To use HTTPS, you may use a proxy server.
- **S3 API Object Versioning:** Instead of deleting previous versions of S3 objects, the gateway will maintain a history of object versions.
- **Bucket Sharding:** When buckets contain an extraordinary number of objects (e.g., 100k-1M+), bucket index performance degraded in previous releases. Bucket sharding dramatically improves performance in those scenarios.
- **Swift API Placement Policies:** The Swift API now allows you to create a bucket and specify a placement pool key (e.g., mapping a bucket and its object to high performance pools, such as SSD backed pools).

CHAPTER 3. FIXED ISSUES

Bug ID	Component	Status	Summary
1207328	Ceph	ON_QA	Cache tiering [tech preview]
1207344	Ceph	ON_QA	RGW: Swift Storage Policy
1217668	Installer	VERIFIED	ceph-deploy is not aware of RH Ceph mon/osd package split
1218962	Installer	VERIFIED	"RFE: ""ceph-deploy mds create"" should print a helpful error that CephFS is not supported"
1219294	Installer	VERIFIED	Update ice-setup to reflect new ISO/CDN channel structure
1219344	Build	VERIFIED	"1.3.0: ""ceph-deploy calamari connect <node>"" fails"
1221830	Installer	VERIFIED	ceph-deploy tries radosgw which is ceph-radosgw in 1.3.0 (No Match for argument: radosgw)
1223475	Installer	VERIFIED	all ice-setup tests are failing
1192424	Calamari	VERIFIED	calamari-server - is not stripped of DWARF data on x86_64
1194814	Build	VERIFIED	calamari-server: rpm -V fails
1210415	Build	VERIFIED	ceph-dencoder links against libtcmalloc

Bug ID	Component	Status	Summary
1211310	Build	VERIFIED	rebase calamari to 1.3
1188878	Distribution	VERIFIED	calamari-minions dependencies really should be separate repo
1206745	Calamari	VERIFIED	Multi-cluster UI
1206746	Calamari	VERIFIED	Crushmap management in Calamari API
1206747	Calamari	VERIFIED	Role support, read/only and read/write
1207341	Ceph	VERIFIED	RGW: Bucket sharding
1222094	Ceph	VERIFIED	rgw: broken manifest when resending part
1222095	Ceph	VERIFIED	rgw: broken multipart upload when resending parts
1215802	Calamari	VERIFIED	Diamond package fails to install on minion nodes during state.highstate
1225172	Ceph	VERIFIED	librbd: aio calls may block
1192022	Build	VERIFIED	rbd udev rules should be in ceph-common

Bug ID	Component	Status	Summary
1194156	Distribution	VERIFIED	[ceph-1.3] change ISO structure to match variants
1197734	Build	VERIFIED	ceph - Library files not compiled with RELRO
1199257	Distribution	VERIFIED	Missing product certificates on installed system
1207323	Ceph	VERIFIED	OSD with SSD
1207324	Ceph	VERIFIED	More robust rebalancing
1207327	Ceph	VERIFIED	Local/pyramind erasure codes (Tech Preview)
1207329	Ceph	VERIFIED	Time-scheduled scrubbing
1207331	Ceph	VERIFIED	Degraded Object improvements: minsize changes
1207337	Ceph	VERIFIED	IPv6 OSD support
1207339	Ceph	VERIFIED	RGW: Object Versioning
1207348	Ceph	VERIFIED	RBD: Allocation hinting

Bug ID	Component	Status	Summary
1207353	Ceph	VERIFIED	RBD: Cache hinting
1207354	Ceph	VERIFIED	RBD: Copy on Read
1207356	Ceph	VERIFIED	RBD: Mandatory exclusive locks
1207357	Ceph	VERIFIED	RBD: Object map
1207358	Ceph	VERIFIED	RBD: Read-ahead
1207359	Ceph	VERIFIED	RBD:Local client cache enabled by default
1207361	Ceph	VERIFIED	RBD: import/export parallelization
1210037	Distribution	VERIFIED	rebase ceph to 0.94.1
1210038	Distribution	VERIFIED	rebase ceph-deploy to 1.5.25
1211304	Build	VERIFIED	ceph-objectstore-tool should be in ceph-osd subpackage
1214518	Build	VERIFIED	"rgw attempts to start using ""apache"" UID"

Bug ID	Component	Status	Summary
1217893	Distribution	VERIFIED	ISO contains different packages than puddle
1219296	Distribution	VERIFIED	add ice-setup back into builds, installer channel
1219322	Build	VERIFIED	ceph-test package for downstream
1217903	Distribution	VERIFIED	ISO - missing README, EULA, GPL, GPG, cert
1213723	Ceph	VERIFIED	Compensate for pg removal bug from firefly and earlier when upgrading to hammer
1222505	Installer	VERIFIED	"1.3.0: ""ceph-deploy install"" with custom cluster name fails"
1223149	Installer	VERIFIED	ceph-deploy install --repo does not handle MON vs OSD
1231990	Installer	VERIFIED	Missing public key
1181915	Ceph	VERIFIED	Request for object over 512K using range header fails when using swift api.
1187821	Ceph	VERIFIED	.rgw pool contains extra objects
1207346	Ceph	VERIFIED	RGW: IPv6

Bug ID	Component	Status	Summary
1213986	Ceph	VERIFIED	RGW swift API: Response header of COPY request for object does not contain certain headers
1213989	Ceph	VERIFIED	RGW Swift API: lack of mandatory ETag header in response for COPY/PUT with X-Copy-From
1214000	Ceph	VERIFIED	rgw: keystone token cache does not work correctly
1214007	Ceph	VERIFIED	rgw: civetweb number of threads is limited
1214073	Ceph	VERIFIED	rgw: shouldn't need to disable rgw_socket_path if frontend is configured
1214826	Ceph	VERIFIED	rgw: object set attrs clobbers object removal bucket index update
1232953	Ceph	VERIFIED	rgw: multipart objects starting with underscore are incompatible with older versions
1186544	Calamari	VERIFIED	Update logo to Red Hat
1215850	Calamari	VERIFIED	De-hardcode 7.0 in calamari
1222153	Installer	VERIFIED	ceph-deploy rgw create command is broken
1219559	Distribution	VERIFIED	cinder-volume keeps opening Ceph clients until the maximum number of opened files reached

Bug ID	Component	Status	Summary
1227351	Documentation	VERIFIED	[GSS] Upgrade procedure from radosgw apache implementation to radosgw CivetWeB implementation
1225209	Distribution	VERIFIED	CVE-2015-3010 ceph-deploy: keyring permissions are world readable in ~ceph [ceph-1.3]
1225214	Distribution	VERIFIED	CVE-2015-4053 ceph-deploy: ceph-deploy admin command copies keyring file to /etc/ceph which is world readable [ceph-1.3]
1209975	Ceph	VERIFIED	Civetweb
1262976	Ceph	VERIFIED	upstart: make config less generous about restarts. This issue was specific to Ubuntu only.

CHAPTER 4. KNOWN ISSUES

Bug ID	Component	Status	Summary
1222509	Ceph	Assigned	Monitor fails to come up. It dies as soon as it's started.
1223335	Calamari	Assigned	Calamari UI: calamari UI→ Graph → Selecting a mon from host list does not display any graph.
1223656	Calamari	Assigned	GUI: Manage→ClusterSettings→Update button is not disabled when check box item is unchecked and clicking on Update leaves button unusable later.
1225222	Ceph	Assigned	OSD crash in **`release_op_ctx_locks`** with rgw and pool snaps.
1229976	Ceph	Assigned	When used with OpenStack cinder, **`rbd_max_clone_depth`** doesn't enforce a flatten on the **`rbd`** volume after the depth is reached.
1230679	Calamari	Assigned	Unable to start Calamari post RHEL upgrade.
1231203	Ceph	Assigned	Running **`ceph-deploy mon add <mon2>`** failed to complete in 300 sec. After interrupting the command(**`ctrl-c`**), the ceph commands time out.
1232036	Ceph	Assigned	**`radosgw-agent`** can't use IPv6 destination.
1250042	Ceph	New	When the writeback process is blocked by I/O errors, Ceph Block Device terminates unexpectedly after force shutdown in the Virtual Manager.
1269048	Ceph	New	The number of restarts before the upstart saturation for different kill intervals is not consistent. This issue is specific to Ubuntu only.