Red Hat Ceph Storage

Release Notes

9.2

October 2026

Contents

Release notes for 9.2 3
New features and enhancements 3
Deprecated functionality 7
Technology Preview 8
Bug fixes 10
Known issues 16
Sources 20

Release notes for 9.2

Red Hat Ceph Storage is a hardened, qualified, secure, and supported enterprise software curated from the Ceph open-source project and delivered by Red Hat.

New features and enhancements

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

`cephadm` utility

New support for configurable CephFS client logging for NFS services

Previously, administrators could not enable CephFS client logging for NFS services, limiting visibility into CephFS client operations when troubleshooting NFS services that use the CephFS backend.

With this enhancement, administrators can enable CephFS client logging by setting the enable_cephfs_client_log parameter to `true` in the NFS service specification. When enabled, `cephadm` mounts a host log directory into the NFS Ganesha container and configures the `libcephfs` client to write debug logs. Additional parameters, including cephfs_client_log_level and cephfs_client_log_dir, allow administrators to control logging verbosity and specify the log file location. This enhancement provides access to CephFS client logs for diagnosing NFS and CephFS connectivity, performance, and operational issues.

(IBMCEPH-8697)

New SSL/TLS support for the Prometheus Manager module

Previously, the Prometheus Manager module exposed metrics over unencrypted HTTP connections.

With this enhancement, the Prometheus Manager module supports SSL/TLS, enabling the metrics endpoint to serve content over HTTPS. This enhancement improves security by helping protect metrics data during transmission.

(IBMCEPH-17574)

Ceph Dashboard

Enhanced Monitor status visibility in the dashboard

Previously, the dashboard displayed two separate quorum-related tables that overlapped with the status table, reducing readability and making Monitor status information harder to view.

With this enhancement, the quorum-related information is consolidated into a single Monitors table and displayed in a vertical layout alongside the status table. An additional In Quorum column indicates quorum participation for each monitor. This enhancement improves dashboard usability and provides clearer visibility into monitor status.

(IBMCEPH-14296)

New support for complete CephFS snapshot mirroring management

Previously, the Ceph Dashboard did not provide controls to remove snapshot mirroring peers or disable snapshot mirroring for CephFS file systems. Administrators had to use the command-line interface (CLI) to run the following commands:

```
ceph fs snapshot mirror peer_remove
```

```
ceph fs snapshot mirror disable
```

With this enhancement, the Ceph Dashboard includes options to remove mirroring peers and disable snapshot mirroring directly from the user interface. As a result, administrators can perform common CephFS snapshot mirroring management tasks directly in the Ceph Dashboard without using the CLI.

(IBMCEPH-16969)

Enhanced virtual-host style S3 domain and certificate configuration in the Ceph Dashboard

Previously, configuring virtual-hosted-style bucket access and zonegroup hostnames required administrators to manually update service specifications or use the CLI. Wildcard certificates for bucket subdomains also had to be configured manually.

With this enhancement, the Ceph Dashboard provides integrated controls for configuring virtual-host-style bucket access and S3 domain settings. Administrators can perform the following tasks directly from the dashboard:

Configure S3 hostnames
: Define one or more hostnames for virtual-host-style bucket access.

Enable virtual-host-style addressing
: Configure bucket access by using hostnames such as `bucket.s3.example.com`.

Manage zonegroup hostnames
: Configure zonegroup hostname settings through the graphical interface.

Generate wildcard certificates
: Automatically generate wildcard Subject Alternative Name (SAN) certificates for bucket subdomains when using `cephadm`-signed certificates.

As a result, administrators can configure S3 domain settings through a simplified user interface, standardize bucket addressing for S3 applications, and help prevent SSL certificate mismatches for bucket subdomains.

(ISCE-3164)

New dashboard support for configuring and managing CephFS snapshot mirroring
: Previously, configuring CephFS snapshot mirroring required multiple command-line steps to establish remote cluster connections, deploy mirror daemons, configure replication schedules, and monitor synchronization status.

  With this enhancement, the Ceph Dashboard provides a dedicated user interface for configuring and managing CephFS snapshot mirroring across clusters. The interface includes guided setup for connecting remote peer clusters, mirror daemon deployment and management, bootstrap token generation and import, replication configuration for subvolumes and directories, and centralized monitoring of replication status, connection health, synchronization checkpoints, and mirror activity.

  As a result, administrators can configure and monitor CephFS snapshot mirroring more easily, reducing deployment complexity, minimizing configuration errors, and improving visibility into disaster recovery replication workflows.

  (ISCE-2752)

Ceph Block Devices

Enhanced Ceph Object Gateway live migration with direct source cluster connection and credential configuration
: Previously, import-only Ceph Object Gateway live migrations from a remote Ceph cluster required administrators to distribute source cluster configuration files and keyrings to all migration client nodes.

  With this enhancement, the rbd migration prepare --import-only command and API support specifying source cluster connection information directly in the migration specification. Administrators can provide source monitor endpoints and authentication credentials in the source-spec JSON file or reference credentials stored securely in the Monitor configuration key store. Source cluster connection metadata is also persisted in the destination cluster, enabling clients and services that consume Ceph Block Device images to automatically inherit the required source cluster configuration.

  This enhancement simplifies cross-cluster live migrations, reduces configuration management overhead, and improves integration with downstream consumers of migrated images.

  (ISCE-4345)

Ceph File System

Memory-bounded journal recovery is now available in `cephfs-journal-tool`
: Previously, recovering dentries from file systems with very large MDS journals could consume unbounded memory. In some cases, the `cephfs-journal-tool` process was terminated by the Out-of-Memory (OOM) killer before metadata recovery completed.

  With this enhancement, the `cephfs-journal-tool event recover_dentries` command supports memory-bounded batching through the `--max-rss` option.

  As a result, administrators can perform offline journal metadata recovery on large journals more reliably by limiting memory usage, helping prevent OOM terminations and ensuring successful dentry recovery.

  (IBMCEPH-14394)

NFS services

Enhanced security for NFS service gRPC communication

Previously, the server and client certificates required for gRPC communication were not generated automatically.

With this enhancement, `cephadm` automatically generates the server and client certificates required for gRPC support, enabling gRPC communication for the NFS service to be secured using certificates that are generated and managed by `cephadm`.

(IBMCEPH-16155)

Ceph Object Gateway

New runtime heap profiling and memory inspection support for Ceph Object Gateway

Previously, runtime heap profiling and memory inspection commands were available for Object Storage Daemon (OSD), Ceph Monitor (MON), and Ceph Metadata Server (MDS) daemons but were not supported for Ceph Object Gateway. As a result, administrators had limited ability to diagnose memory usage and investigate memory-related issues on running Ceph Object Gateway daemons.

With this enhancement, Ceph Object Gateway supports the ceph daemon SOCKET heap administrative socket commands for runtime heap inspection and tcmalloc profiling. Administrators can view heap statistics, start and stop heap profiling, generate heap dumps, and manage memory release behavior without restarting the daemon. This enhancement provides memory diagnostics capabilities that are consistent with other Ceph daemons.

(IBMCEPH-13029)

New support for unique service identifiers in performance counter data

Previously, performance counter data did not include a service-specific identifier that could be used to distinguish between services reporting metrics to monitoring and management tools.

With this enhancement, administrators can specify a unique service identifier by using the --service_unique_id option when starting supported Ceph services. When configured, a `service_unique_id` section is added to the performance counter output and includes the specified identifier. This enhancement enables monitoring and management tools to associate performance data with specific services more easily.

(IBMCEPH-10722)

Bucket notifications are now available for cloud restore events

With this enhancement, the Ceph Object Gateway can send bucket notifications for restore operations on objects that were moved to a cloud tier.

You can set up bucket notifications for the following events:

s3:ObjectRestore:Post
: Sent when a restore request starts for an object in a cloud tier.

s3:ObjectRestore:Completed
: Sent when the restore finishes.

As a result, applications can track cloud restore operations through their notification endpoints, such as Kafka.

(IBMCEPH-14125)

Ceph Object Gateway multi-site

Full sync now preserves metadata log entries

Previously, Ceph Object Gateway multi-site metadata full synchronization did not copy metadata log (`mdlog`) entries from the primary zone. As a result, if a newly synchronized zone was promoted to the metadata master during a failover, secondary zones that were still performing incremental synchronization could miss metadata updates because the required `mdlog` entries were not available.

With this enhancement, metadata full synchronization now copies `mdlog` entries from the primary zone along with metadata objects.As a result, newly synchronized zones retain the metadata log history required to support incremental synchronization. Secondary zones can continue to receive metadata updates successfully if the synchronized zone is promoted to metadata master during a failover.

(IBMCEPH-7621)

RADOS

Monitor election strategy changes are now restricted in stretch mode

Previously, Monitor election strategies could be changed while stretch cluster mode was enabled. Because stretch mode depends on connectivity-based elections to maintain quorum and tiebreaker behavior across sites, changing the election strategy could lead to election failures or cluster instability.

With this enhancement, IBM Storage Ceph prevents changes to the Monitor election strategy while stretch mode is enabled. As a result, stretch cluster deployments maintain the required election behavior for stable operation. To change the election strategy, you must first disable stretch mode.

(IBMCEPH-8302)

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.

> **Important:** Deprecated functionality continues to be supported until the end of life of Red Hat Ceph Storage 9.x. Deprecated functionality will likely not be supported in future major releases of this product and is not recommended for new deployments. For the most recent list of deprecated functionality within a particular major release, refer to the latest version of release documentation.

Deprecated method of configuring OIDC federation and IAM roles at the tenant level

All OIDC resources are now managed as resources within a Ceph Object Gateway account. These OIDC resources include providers, roles, and polices, As a result, all OIDC operations that target a tenant, including the global or empty tenant, are considered deprecated. The deprecated operations incldue creating providers, creating roles, and assuming roles.

With the newer per-account model, federated users are directly associated with the account and Ceph Object Gateway no longer creates shadow users (for example, `TENANT$USER_NAMESPACE`) upon role assumption. The account itself tracks all resources and identities.

Tenant-based OIDC federation users should migrate their configurations to the new Ceph Object Gateway per-account model, before feature removal.

For more information, see Red Hat Ceph Storage Ceph Object Gateway > Authenticating with Ceph Object Gateway > Secure Token Service.

(ISCE-3162)

Technology Preview

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

`cephadm` utility

New `cephadm` agent

The new `cephadm` agent runs on each host in a Red Hat Ceph Storage cluster and collects host and daemon metadata locally before sending updates to the Ceph Manager over TLS-secured HTTPS connections. By default, the Ceph Manager retrieves host information through SSH connections.

When the `cephadm` agent is enabled, host metadata is collected locally and communicated directly to the Ceph Manager, reducing SSH communication overhead and providing near real-time host and daemon state information. This approach can improve orchestration responsiveness and scalability for larger clusters. If an agent stops sending updates, the cluster raises a CEPHADM_AGENT_DOWN health warning. In this situation, the Ceph Manager automatically falls back to SSH-based communication for the affected host.

The `cephadm` agent is disabled by default. To enable the feature, set the `mgr/cephadm/use_agent` configuration option.

For more information, see Administering Red Hat Ceph Storage > Administration > cephadm operations > cephadm agent (Technology Preview).

(ISCE-4095)

Ceph File System (CephFS)

Support for CephFS snapshot mirroring

With this feature you can replicate a CephFS to a remote CephFS on another Ceph storage cluster. Snapshot synchronization copies snapshot data to a remote Ceph File System, and creates a new snapshot on the remote target with the same name. You can configure specific directories for snapshot synchronization.

> **Note:** To use CephFS snapshot mirroring, both the source and the target storage clusters must be running the same Red Hat Ceph Storage version.

For more information, see Red Hat Ceph Storage File Systems > CephFS snapshot mirroring (Technology Preview).

Ceph Object Gateway

New per-user and per-bucket usage counters in Prometheus

Ceph Object Gateway now exports per-user and per-bucket usage counters via performance counters automatically collected by the ceph-exporter and made available in Prometheus. This provides low-overhead, real-time visibility into the following:

- Per-bucket metrics: used bytes, utilized bytes, and number of objects
- Per-user metrics: used bytes and number of objects
- Cache performance metrics: cache hits, misses, updates, and evictions

> **Note:** These metrics are disabled by default. To enable them, configure the appropriate settings in your Ceph Object Gateway configuration.

For more information, see Red Hat Ceph Storage Dashboard > Managing Ceph Object Gateway > Viewing Ceph Object Gateway per-user and per-bucket performance counters.

(IBMCEPH-2898)

RADOS

Balanced primary placement groups can now be observed in a cluster

Previously, users could only balance primaries with the offline `osdmaptool`.
With this enhancement, autobalancing is available with the `upmap` balancer. Users can now choose between either the `upmap-read` or `read` mode. The `upmap-read` mode offers simultaneous `upmap` and `read` optimization. The `read` mode can only be used to optimize reads.

For more information, see Administering Red Hat Ceph Storage > Operations > Managing manager daemons > Using the Ceph Manager balancer module.

(BZ#1870804)

Now supports tracking data availability score of a cluster

This release introduces a feature that tracks the data availability score of a Ceph cluster over time. The score represents how accessible your data is at any given moment, based on factors such as OSD health, placement group states, and redundancy policies.

By monitoring this metric, administrators gain a fact-based view of cluster reliability and can validate availability percentages (for example, 99.99%) against service-level objectives. This capability provides actionable insight into operational resilience and helps ensure confidence in Ceph as a storage platform for critical workloads.

For more information, see Administering Red Hat Ceph Storage > Monitor Red Hat Ceph Storage clusters > Monitor a Ceph cluster with the CLI > High-level monitoring > Track the data availability score of a cluster (Technology Preview).

(ISCE-2144)

Bug fixes

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

`cephadm` utility

Cluster upgrades now complete faster and with less disruption

Previously, during upgrades that required CephX keyring rotation, such as upgrades to Red Hat Ceph Storage 9.1, `cephadm` performed keyring rotation as a separate post-upgrade task after daemon upgrades completed. As a result, upgrades could take significantly longer, prolong PG degradation, reduce I/O performance, and trigger unnecessary manager failovers and monitor restarts during staggered upgrades.

With this fix, `cephadm` performs keyring rotation as part of the daemon upgrade process and eliminates redundant key rotation cycles across staggered upgrade phases. As a result, cluster upgrades complete faster, reduce unnecessary daemon restarts, and minimize upgrade-related service disruption.

(IBMCEPH-17825)

The `cephadm`-agent version is now reported correctly in orchestrator output

Previously, the version detection logic did not support `cephadm`-agent because the agent runs as a native Python process instead of a containerized daemon. As a result, the ceph orch ps command displayed the agent version as <unknown>, making it difficult to validate deployments and upgrades.

With this fix, dedicated version handling for `cephadm`-agent reports the associated `cephadm` version for agent daemons. As a result, ceph orch ps --daemon-type agent now displays the correct `cephadm`-agent version, simplifying version verification and upgrade validation.

(IBMCEPH-15691)

NFS deployments no longer fail when grace service initialization requires additional time

Previously, NFS deployments could fail with CEPHADM_APPLY_SPEC_FAIL when the `ganesha-rados-grace` service did not initialize within the configured timeout period, causing temporary cluster health warnings.

With this fix, the `ganesha-rados-grace` timeout is configurable and defaults to 30 seconds. As a result, administrators can increase the timeout when needed, preventing transient deployment failures and improving NFS deployment reliability.

(IBMCEPH-14352)

`OSD_FLAGS` health warning text now accurately describes affected flags

Previously, the `OSD_FLAGS` health warning used ambiguous syntax that could be misread. The warning read: [WRN] OSD_FLAGS: 1 OSDs or CRUSH {nodes, device-classes}have {NOUP,NODOWN,NOIN,NOOUT} flags set.

With this fix, the warning text is updated to use clearer language. As a result, the health warning now reads: [WRN] OSD_FLAGS: 1 OSDs or CRUSH nodes/device-classes have one or more of these flags set: NOUP, NODOWN, NOIN, NOOUT.

(IBMCEPH-11614)

`cephadm` host checks now support FIPS-enabled environments

Previously, `cephadm` generated `ED25519` SSH keys by default during bootstrap. In FIPS-enabled environments, these keys were not supported and could prevent `cephadm` from connecting to managed hosts. As a result, host checks could fail, managed hosts could be reported as offline, and CEPHADM_HOST_CHECK_FAILED health warnings could be raised.

With this fix, `cephadm` automatically generates FIPS-compliant `RSA-4096` SSH keys in FIPS-enabled environments and provides actionable guidance when incompatible SSH keys are detected on existing clusters. As a result, `cephadm` can successfully connect to managed hosts and complete host checks in FIPS-enabled environments without generating unnecessary host check failures.

(IBMCEPH-16272)

OSD service specifications now require a service ID

Previously, the ceph orch apply osd -i command accepted OSD service specifications that did not include a `service_id`. Although the specification was accepted, OSD deployment subsequently failed because

the OSD containers and systemd services were not started.

With this fix, the ceph orch apply osd -i command validates OSD service specifications and rejects specifications that do not include a `service_id`. As a result, invalid OSD service specifications are identified immediately with a clear error message, preventing failed OSD deployments and providing behavior that is consistent with ceph orch apply -i.

(IBMCEPH-14354)

`cephadm` no longer crashes when draining hosts

Previously, a race condition could occur between draining a host and `cephadm` applying service specifications in the serve loop. In this scenario, the serve loop could attempt to access a host that was no longer present in the inventory, causing the `cephadm` module to crash.

With this fix, hosts that are no longer present in the inventory during service specification processing are skipped. As a result, the `cephadm` module no longer crashes due to this race condition when hosts are being drained.

(IBMCEPH-19376)

Ceph build

OSDs no longer abort on startup on non-4K page size systems

Previously, an assertion failure could occur in BlueFS envelope header alignment during append and flush operations on hosts where the memory page size exceeded the BlueFS block size, such as `ppc64le` and 64 KB page-size `aarch64` systems.

With this fix, BlueFS envelope header alignment calculations correctly handle host architectures that use non-4K memory page sizes.As a result, OSDs start and run normally on non-4K page size systems without unexpected startup aborts.

(IBMCEPH-18524)

Ceph Ansible

Preflight checks now support bonded and bridged network interfaces

Previously, the cephadm-preflight.yml playbook could fail during interface validation when network interface names contained hyphens, such as bridge or bond interfaces. This issue caused undefined variable errors during preflight validation and prevented the checks from completing successfully.

With this fix, the interface validation logic correctly resolves Ansible facts for network interfaces with hyphenated names. As a result, preflight checks now complete successfully on systems that use bonded or bridged interfaces, improving the reliability of deployment validation.

(IBMCEPH-13396)

Ceph Dashboard

Target storage class information is now displayed in the Storage Class details view

Previously, the Storage Class details view did not display the target storage class because the target_storage_class field was missing from the form and details section.

With this fix, the target_storage_class field is added to both the Storage Class form and the details view. As a result, target storage class information is now displayed correctly, providing administrators with complete visibility into Storage Class configuration.

(IBMCEPH-17779)

Location constraint information is now displayed in the Storage Class details view

Previously, the Storage Class details view did not display the location constraint because the location_constraint field was missing from the form and details section.

With this fix, the location_constraint field is added to both the Storage Class details and the details view. As a result, location constraint information is now displayed correctly, providing administrators with complete visibility into Storage Class configuration.

(IBMCEPH-17778)

Storage Class labels now accurately reflect supported configuration
: Previously, the Glacier Configuration labels in the dashboard used terminology that implied support for AWS Glacier.

  With this fix, the labels are updated to use clearer and more accurate terminology that reflects the supported functionality. As a result, the dashboard no longer suggests AWS Glacier support and provides a more accurate representation of the available configuration options.

  (IBMCEPH-17777)

Realm field description now displayed correctly
: Previously, the Realm field on the Ceph Object Gateway service creation page displayed an incorrect description that did not accurately explain the field's purpose.

  With this fix, the description is updated to provide accurate guidance when configuring Ceph Object Gateway services.

  (IBMCEPH-16348)

Dashboard user creation now validates username length
: Previously, the dashboard allowed creation of users with usernames longer than 256 characters. Although the user was created successfully, login attempts failed because the username exceeded the supported length.

  With this fix, validation limits usernames to a maximum of 256 characters. When a username exceeds this limit, the dashboard displays a warning message and the username cannot be created. As a result, newly created dashboard users can authenticate successfully.

  (IBMCEPH-15230)

Ceph Object Gateway synchronization metrics now display correctly in Grafana dashboards
: Previously, Ceph Object Gateway synchronization metrics were exposed with zone-specific metric names instead of the generic metric name expected by Grafana. As a result, metrics such as `ceph_data_sync_from_zone_fetch_bytes_sum` were not collected correctly, and the RGW Sync Overview dashboard displayed no data in synchronization panels.

  With this fix, the metric reporting logic exposes synchronization metrics using the expected generic metric names. As a result, the RGW Sync Overview dashboard now collects the metrics correctly and displays synchronization data in Grafana panels as expected.

  (IBMCEPH-13197)

Ceph File Systems (CephFS)

`cephfs-mirror` directory synchronization no longer causes out-of-order RADOS operations
: Previously, concurrent directory synchronization operations in `cephfs-mirror` could issue writes to the same internal object without synchronization. This allowed RADOS operations to reach OSDs out of submission order, which could cause OSD crashes when debug operation ordering was enabled and might lead to stale synchronization metrics.

  With this fix, synchronization between concurrent directory synchronization operations ensures that updates are written in the correct order. As a result, `cephfs-mirror` processes directory synchronization statistics reliably, preventing OSD crashes related to out-of-order operations and improving metric consistency.

  (IBMCEPH-18029)

Subvolume quarantine access is now correctly enforced
: Previously, CephFS clients could incorrectly deny access to quarantined subvolumes due to client-side quarantine checks and stale cached state, preventing some recovery clients from accessing files and snapshots.

  With this fix, client-side quarantine enforcement is removed and access control relies on Ceph Metadata Server (MDS)-side enforcement. As a result, recovery clients with `q/Q` capabilities can access quarantined content as expected, and normal client access is restored immediately after quarantine is disabled.

  (IBMCEPH-17674)

MDS metrics collection no longer crashes during scaling and failover operations

Previously, MDS rank 0 could terminate unexpectedly during max_mds scaling operations or MDS failover when stale or duplicate client metrics removal messages were received for clients that were no longer tracked.

With this fix, handling of client metrics removal requests is improved by safely ignoring missing client and rank entries and protecting performance counter cleanup operations when metrics are no longer present. As a result, MDS rank 0 no longer crashes during scaling or failover events, improving the stability of metrics collection and administrative operations.

(IBMCEPH-16973)

MDS no longer crashes when processing large directory entries

Previously, when an individual directory entry or encoded inode exceeded the `max_dir_commit_size` limit, `CDir::_omap_commit_ops()` could trigger an OMAP commit before any updates, removals, or header information were staged. This caused the MDS to hit an assertion `(ceph_assert(header || !_set.empty() || !_rm.empty()))` and crash unexpectedly.

With this fix, `CDir::_omap_commit_ops()` no longer dispatches OMAP commits when the update, removal, and header sets are all empty. As a result, the MDS handles large directory entries safely without triggering empty OMAP commits, preventing daemon crashes and maintaining CephFS stability.

(IBMCEPH-17442)

Ceph Object Gateway

Cloud-tiered object categories are preserved during bucket resharding

Previously, when a bucket was manually or automatically resharded, some bucket index entries were updated in a way that reset the object category. As a result, cloud-tiered objects could incorrectly appear with the `Main` category instead of `CloudTiered`.

With this fix, the resharding process preserves object category information when bucket index entries are updated. As a result, cloud-tiered objects retain the correct category during manual and automatic bucket resharding operations.

(IBMCEPH-18034)

Lifecycle processing no longer crashes when cloud endpoints terminate uploads

Previously, a lifecycle worker could crash during object uploads if a cloud endpoint terminated a PUT request before the upload completed. Under this condition, request cleanup set the internal manager reference to `NULL`, but subsequent processing continued to access it, resulting in a crash.

With this fix, validation ensures the internal manager reference is present before it is accessed during request processing. As a result, lifecycle processing continues successfully when cloud endpoints end uploads unexpectedly, improving stability and reliability for cloud-tiering workloads.

(IBMCEPH-18030)

The --max-entries option now limits output correctly for radosgw-admin bi list

Previously, the radosgw-admin bi list command ignored the --max-entries option and returned all entries, regardless of the specified limit.

With this fix, the command honors the --max-entries setting when listing bucket index entries. As a result, radosgw-admin bi list now returns only the requested number of entries, improving usability when working with large bucket indexes.

(IBMCEPH-17241)

The x-amz-expiration response header now reports lifecycle expiration information correctly

Previously, the x-amz-expiration response header could incorrectly report `NoncurrentVersionExpiration` lifecycle rules for current object versions. As a result, GET and HEAD requests could return incorrect expiration dates and rule IDs, particularly in versioned buckets that used only `NoncurrentVersionExpiration` lifecycle rules.

With this fix, the lifecycle header processing logic uses only the current-version `Expiration` rule when generating the x-amz-expiration response header. In addition, requests for specific object versions no longer return the header, and current-version requests are correctly identified. As a result, the x-amz-expiration response header now reports accurate lifecycle expiration information.

(IBMCEPH-16346)

User creation date is now returned by the Admin Ops API

Previously, the user creation date was not included in Admin Ops API responses because the `dump_user_info()` function did not expose the create_date field. As a result, user creation date information was not visible when retrieving user details.

With this fix, the `dump_user_info()` function is updated to include the create_date field in API responses. As a result, the user creation date is now returned as part of the Get User API output, providing administrators with complete user account information.

(IBMCEPH-13193)

Authentication type is now logged correctly for presigned URLs that use SigV2

Previously, presigned URLs that used SigV2 were not handled correctly, which caused the Authentication Type field in log records to be set to – instead of the correct value.

With this fix, SigV2 presigned URLs are handled correctly. As a result, the Authentication Type field in log records now displays the correct authentication type.

(IBMCEPH-12412)

CORS requests now follow the correct processing path when no configuration exists

Previously, when no CORS configuration was found, the operation returned an invalid result. As a result, request processing followed an incorrect I/O path.

With this fix, the operation returns `0` when no CORS configuration is present.As a result, requests follow the correct I/O path and are processed as expected.

(IBMCEPH-16988)

CORS requests are now evaluated against all matching rules

Previously, CORS processing stopped at the first rule that matched the request origin without evaluating additional rules for the same origin. As a result, requests could be incorrectly rejected even when a valid matching rule existed.

With this fix, all applicable CORS rules for the matching origin are evaluated before making an access decision. As a result, CORS requests are processed correctly and accepted when a matching rule permits the operation.

(IBMCEPH-17318)

CORS rule validation now applied for all requests

Previously, CORS rule verification was run only for preflight requests. As a result, CORS requests that did not require a preflight check were not validated against the configured CORS rules.

With this fix, the CORS rule verification is applied to all CORS requests. CORS requests are now correctly validated against the configured rules.

(IBMCEPH-9464, IBMCEPH-13886)

RADOS

Improved client performance during EC backfill operations

Previously, during concurrent client workloads and backfill operations on Erasure Coded (EC) pools, EC recovery read operations could be prioritized ahead of client operations when an OSD was marked out for draining. As a result, client operations could experience increased tail latency and trigger slow operation warnings.

With this fix, EC recovery reads are scheduled through a background queue, recovery read costs are calculated more accurately, and mClock queue profiles are tuned to prevent background operations from delaying client workloads. As a result, client and backfill traffic are isolated more effectively, reducing client latency and slow operation warnings during EC recovery and backfill activities.

(IBMCEPH-16933)

PG `stuck peering` warnings are now reported more accurately

Previously, the `last_peered` value could become stale after a primary OSD change. As a result, the monitor could incorrectly report placement groups (PGs) as `stuck peering` even when peering completed within a few seconds, generating false `PG_AVAILABILITY` warnings during failovers, upgrades, or OSD instability.

With this fix, the `stuck peering` check uses the most recent of the `last_active` and `last_peered` timestamps, and replica OSDs continue to update PG state statistics when they later become primary. As a result, false `stuck peering` warnings are eliminated, and alerts are generated only when a placement group is genuinely stuck in the peering state.

(IBMCEPH-18447)

Placement group rebuild duration tracking is now more accurate across primary OSD failovers

Previously, rebuild progress was not tracked consistently when the primary OSD for a placement group (PG) changed. As a result, rebuild time was not fully recorded for PGs whose primary OSD was marked down during a rebuild operation.

With this fix, each primary OSD independently tracks rebuild operations that occur while it is responsible for the PG. As a result, rebuild duration metrics now reflect the combined rebuild time across primary OSD handovers, providing a more accurate view of the total rebuild duration for a placement group.

(IBMCEPH-17828)

NFS services

ACL settings and inheritance now work correctly for NFS shares

Previously, the NFS service did not correctly apply default ACLs when ACLs were first configured by using the nfs4_setfacl -s command. In addition, subdirectories created under directories with ACLs used overly restrictive permissions, which could prevent ACL inheritance for non-root users.

With this fix, the NFS service correctly applies default ACLs during initial configuration and uses permissive `777` permissions when creating subdirectories under directories that have ACLs configured. As a result, ACL settings and inheritance work correctly, ensuring that access permissions are applied as expected for users and subdirectories.

(IBMCEPH-19125)

Known issues

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

`cephadm` utility

Grafana certificate is not migrated during upgrade

When you upgrade from Red Hat Ceph Storage 8.1 to 9.0, the existing user-signed Grafana certificate is not migrated. This occurs because `cephadm` replaces the certificate with a `cephadm`-signed certificate during the upgrade process.

As a result, duplicate certificate entries can appear and certificate-related health warnings can persist. Manual reconfiguration is required to restore custom TLS certificates.

> **Note:** Data services remain unaffected.

As a workaround, complete the following steps:

1. Change the Grafana specification to use `certificate_source: reference`.
2. Upload valid certificates by using `certmgr`.
3. Run the following command:

```
ceph orch reconfig grafana
```

(IBMCEPH-13080)

Promtail image remains visible after migration to Alloy

When `cephadm` transitions from Promtail to Alloy, it continues to register the Promtail container image to maintain backward compatibility and ensure a smooth migration path.

As a result, Promtail still appears in the cephadm list-images output after upgrading, even though Alloy is the new default.

Currently, there is no workaround. Ignore the Promtail image entry during the supported transition phase.

(IBMCEPH-13162)

Applying an RDMA-enabled NFS spec to an existing service causes an error state

When you apply an RDMA-enabled NFS specification to an existing NFS service or cluster, the service enters an error state. This occurs because the RDMA configuration cannot be applied to a running service without a full redeployment.

As a result, the NFS service or cluster becomes unavailable and cannot function correctly.

As a workaround, redeploy the NFS service.

(IBMCEPH-14150)

NFS QoS configuration updates do not take effect when reapplying the spec

When the NFS specification is reapplied with updated cluster_qos_config values, the new QoS settings appear as updated in `cephadm` but do not take effect on the NFS client side.

As a result, NFS clients continue to operate with the previous QoS settings despite the new values being shown as set.

As a workaround, restart the NFS service or cluster after reapplying the specification.

(IBMCEPH-11821)

OSD upgrades can pause when daemon cleanup takes longer than expected

When you upgrade a cluster, system cleanup operations for an OSD can occasionally take longer than expected on smaller or heavily loaded systems. This occurs because `systemd` cgroup cleanup for the OSD can exceed the configured timeout during the upgrade.

As a result, the OSD restart can fail with a Device or resource busy error, causing `cephadm` to pause the upgrade and report the UPGRADE_REDEPLOY_DAEMON health warning.

As a workaround, wait for the affected OSDs to recover and for the cluster to return to a healthy state, and then run the following command to resume the upgrade:

```
ceph orch upgrade resume
```

(IBMCEPH-18039)

`cephadm` bootstrap can fail on FIPS-enabled hosts

When you run cephadm bootstrap on FIPS-enabled RHEL 9 or RHEL 10 hosts, bootstrap can fail because the SSH implementation attempts to use AES-GCM ciphers that are not supported under the OpenSSL FIPS provider. In addition, automatic license acceptance can fail because MD5-based operations used during initialization are blocked in FIPS mode.

As a result, cephadm bootstrap can fail during deployment, typically at the host addition stage or during license acceptance, causing cluster initialization to abort and the partially deployed cluster to be removed.

As a workaround, complete the following steps:

1. Create a custom SSH configuration file (for example, /root/cephadm-fips-ssh.conf) that restricts ciphers to AES-CTR:

```
Host *
    StrictHostKeyChecking no
    UserKnownHostsFile /dev/null
    Ciphers aes256-ctr,aes128-ctr
```

2. Run cephadm bootstrap with the --ssh-config /root/cephadm-fips-ssh.conf option and omit the --automatically-accept-license flag during initial deployment.

(IBMCEPH-17350)

The NFS service count in ceph -s might be reported incorrectly

When you deploy an NFS service by using `cephadm`, the generated ganesha.conf file might not include the required service registration settings in the `CEPH` block. This occurs because the service registration parameters are not populated automatically during deployment.

As a result, active NFS daemons might not register correctly with the Ceph service map, and the ceph -s command can display an incorrect number of active NFS daemons and hosts. For example, an NFS service deployed across multiple nodes can be reported as 1 daemon active (1 hosts) even though multiple NFS daemons are running.

As a workaround, complete the following steps:

1. Edit the ganesha.conf file for each NFS daemon.
2. Update the `CEPH` block to include the register_service and nodeid parameters. Ensure that the nodeid value matches the value configured in the `RADOS_KV` block:

```
CEPH {
    register_service = true;
    nodeid = "service-map-daemon-id";
}
```

3. Restart the NFS daemons so that the updated configuration is applied.
4. Verify that the NFS daemons are running by running one of the following commands:

```
ceph orch ps --daemon-type nfs
```

```
ceph orch ls
```

(IBMCEPH-13339)

Ingress daemons are not upgraded automatically during orchestrator upgrades

Ingress daemons (`haproxy` and `keepalived`) are not included in the `cephadm` upgrade workflow and upgrade status tracking logic. As a result, `ceph orch upgrade` can complete successfully while ingress

daemons continue running older container images. Ingress daemons might remain on older container images until they are manually redeployed after the upgrade.

As a workaround, run the following steps:

1. Manually redeploy the ingress service after the upgrade.

```
ceph orch redeploy ingress.SERVICE_ID
```

2. If the ingress service uses updated container images, set the new image references and redeploy the service.

```
ceph config set mgr mgr/cephadm/container_image_haproxy
ceph config set mgr mgr/cephadm/container_image_keepalived NEW_IMAGE
ceph orch redeploy ingress.SERVICE_ID
```

IBMCEPH-19197)

Ceph build

HAProxy deployment fails when QAT is enabled with the ingress feature

When you deploy HAProxy with the QAT feature enabled on Red Hat Ceph Storage 9.0 container images using the ingress feature, the deployment fails.

As a result, HAProxy cannot run with QAT enabled.

As a workaround, set haproxy_qat_support to `false` to continue using HAProxy without QAT.

(IBMCEPH-13100)

`ceph-common` installation can fail on RHEL 9.4 and RHEL 9.6 due to an OpenSSL dependency issue

When installing the `ceph-common-2:20.2.2-349.el9cp.x86_64` package on Red Hat Enterprise Linux (RHEL) 9.4 or 9.6, the installation can fail because the package requires the `OPENSSL_3.4.0` symbol, which is not available in the OpenSSL packages shipped with these operating system versions.

As a result, the `ceph-common` package installation fails and the package is not installed. If this occurs, the following error is emitted:

```
Error: Package: ceph-common-2:20.2.2-349.el9cp.x86_64
       Requires: libcrypto.so.3(OPENSSL_3.4.0)(64bit)
       You could try using --skip-broken to work around the problem
       You could try to run: rpm -Va --nofiles --nodigest
```

As a workaround, run the following steps:

1. Update the OpenSSL packages to versions that provide the required `libcrypto` symbols.
2. Download the required `openssl` and `openssl-libs` RPM packages (version 3.4.0 or later) from the Red Hat Customer Portal.
3. Run a local installation by using the dnf localinstall command.

For more information, see How to resolve OpenSSL dependency errors when installing or upgrading ceph-common on RHEL 9 systems? on Red Hat Customer Portal.

(IBMCEPH-20644)

Ceph Object Gateway

Bucket usage metrics can be overstated when multiple Ceph Object Gateway instances export performance counters

When rgw_enable_usage_perf_counters is set to `true`, each Ceph Object Gateway instance independently reads bucket statistics from RADOS and exports the same metrics through its local exporter. This occurs because each instance reports its own copy of the metrics without deduplication.

As a result, Prometheus receives identical bucket usage metrics from multiple Ceph Object Gateway instances. Queries such as `sum by (owner) (ceph_rgw_bucket_usage_bytes)` can report values that are multiplied by the number of active Ceph Object Gateway instances.

As a workaround, enable rgw_enable_usage_perf_counters on only one Ceph Object Gateway instance, or update Prometheus queries to use `max` instead of `sum` when aggregating bucket usage metrics across instances.

(IBMCEPH-13980)

Server-side copy operations can corrupt objects when `rgw_max_chunk_size` exceeds `rgw_get_obj_max_req_size`

When `rgw_max_chunk_size` is configured with a value greater than `rgw_get_obj_max_req_size`, server-side copy operations might not copy some objects correctly. This issue affects objects that fit within a single chunk but are larger than the configured request size.

As a result, data beyond the first `rgw_get_obj_max_req_size` bytes can be copied incorrectly. The copy operation completes successfully, returns no error, and reports the source object's ETag even though the copied object is corrupted.

Currently, there is no workaround.

As a workaround, do not configure `rgw_max_chunk_size` with a value greater than `rgw_get_obj_max_req_size`. Changing these settings from their default values is not recommended.

If objects were previously copied while this configuration was in use, verify the integrity of the copied objects.

(IBMCEPH-20062)

Ceph File Systems

Large file synchronization performance is reduced with multi-threaded CephFS mirroring

When multi-threaded CephFS mirroring is enabled and the workload contains large files, typically larger than 1 GB, synchronization performance can be reduced.

As a result, large file synchronization might complete more slowly than expected.

As a workaround, set cephfs_mirror_max_datasync_threads to `3` for workloads that primarily contain files larger than 1 GB.

(IBMCEPH-13885)

CephFS can enter an MDS crash loop after file system failover during cache trimming

When you run ceph fs fail on a multi-rank CephFS file system during periods of heavy Ceph Metadata Server (MDS) cache trimming, MDS rank 0 can encounter an inconsistent replica subtree state during journal replay and rejoin processing.

As a result, MDS daemons can repeatedly crash during the rejoin phase, preventing the file system from recovering and returning to an active state.

Currently, there is no workaround. To avoid this issue, do not run ceph fs fail while aggressive MDS cache trimming is in progress.

> **Important:** Running ceph fs reset can stop the crash loop but causes metadata loss, including damaged directory fragments and orphaned inodes.

(IBMCEPH-17527)

Ceph Object Gateway multi-site

Bucket statistics can be inaccurate after resharding cloud-transitioned objects

During bucket resharding, objects with pending index completion can incorrectly reset the object category of cloud-transitioned objects to Main. As a result, bucket statistics can be reported incorrectly.

To avoid this issue, pre-shard the bucket before performing cloud transition operations. As a result, cloud-transitioned objects can be accounted for incorrectly in bucket statistics after resharding.

(IBMCEPH-18034)

Sources

Use this information to attain Red Hat Ceph Storage source code packages.

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

- For Red Hat Enterprise Linux 9: http://ftp.redhat.com/redhat/linux/enterprise/9Base/en/RHCEPH/SRPMS/
- For Red Hat Enterprise Linux 10: https://ftp.redhat.com/redhat/linux/enterprise/10Base/en/RHCEPH/SRPMS/