Red Hat Ceph Storage

Release Notes

9.0

Contents

Release notes for 9.0 3

New features and enhancements 3

`cephadm` utility 3

Ceph Dashboard 4

Ceph File System (CephFS) 4

Ceph Object Gateway 4

Multi-site Ceph Object Gateway 7

RADOS 7

Ceph Block Device (RBD) mirroring 8

Deprecated functionality 8

Technology Preview 9

Ceph File System (CephFS) 9

Ceph Object Gateway 9

RADOS 10

Bug fixes 10

Ceph Dashboard 10

Ceph File System (CephFS) 10

Ceph Object Gateway 10

Multi-site Ceph Object Gateway 14

Known issues 14

`cephadm` utility 14

Ceph build 15

Ceph Dashboard 16

Ceph File System (CephFS) 16

Ceph Object Gateway 17

Ceph Object Gateway multi-site 18

RADOS 18

Ceph Block Device (RBD) 19

Sources 19

Asynchronous updates 20

Release notes for 9.0z1 20

Bug fixes 20

All bug fixes 21

Security fixes 21

Known issues 22

Release notes for 9.0z2 22

New features and enhancements 22

Bug fixes 23

All bug fixes 25

Security fixes 26

Known issues 26

Release notes for 9.0z3 26

New features and enhancements 26

Bug fixes 27

All bug fixes 30

Security fixes 32

Release notes for 9.0

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 `cephadm` certificate lifecycle management for improved Ceph cluster security

`cephadm` certificate lifecycle management was previously available as limited release. This enhancement provides full availability for new and existing customers in production environments.

With this enhancement, `cephadm` now has certificate lifecycle management in the certmgr subsystem. This feature provides a unified mechanism to provision, rotate, and apply TLS certificates for Ceph services, supporting both user-provided and automatically generated `cephadm`-signed certificates. As part of this feature, certmgr periodically checks the status of all certificates managed by `cephadm` and issues health warnings for any that are nearing expiration, misconfigured, or invalid. This improves Ceph cluster security and simplifies certificate management through automation and proactive alerts.

(BZ#2397793)

Multiple container registries can now be defined in registry credentials

Previously, only a single container registry credential could be configured. However, users may have different registries for different service containers.

With this enhancement, registry credentials can now define multiple container registries. To store multiple registry credentials, use the following command:

```
cephadm -v registry-login --registry-json registry.json cat registry.json
cat registry.json
{
  "registry_credentials": [
    {
      "url": "registry1",
      "username": "user1",
      "password": "xxx"
    },
    {
      "url": "registry2",
      "username": "user2",
      "password": "xxx"
    }
  ]
}
```

(BZ#2338350)

Enhanced `config` parameter to set the maximum number of OSDs to upgrade in parallel

With this enhancement, the **`config`** parameter sets the maximum number of OSDs that can be upgraded in parallel. The default value is 16.

For example,

```
[ceph: root@ceph-node-0 ceph]# ceph config get mgr mgr/cephadm/
max_parallel_osd_upgrades16
[ceph: root@ceph-node-0 ceph]#
[ceph: root@ceph-node-0 ceph]# ceph config set mgr mgr/cephadm/max_parallel_osd_upgrades
32
```

(BZ#2390040)

Ceph Dashboard

New support for managing Ceph Object Gateway accounts

Previously, managing Ceph Object Gateway accounts was only possible through the command-line interface (CLI) using radosgw-admin commands.

With this enhancement, you can now view account details, create new accounts, manage quotas, and link users and buckets to an account directly from the Ceph Dashboard.

As a result, Ceph Object Gateway environments align more closely with AWS-style account and IAM semantics, improving usability, scalability, and security governance.

(BZ#2315832)

New migration from Promtail to Grafana Alloy for centralized logging

Previously, centralized logging relied on Promtail, which is now deprecated and no longer recommended for new deployments.

With this enhancement, Red Hat Ceph Storage uses Grafana Alloy for log scraping and forwarding. Grafana Alloy provides a unified, modern, and more efficient agent for log collection, processing, and forwarding.

Grafana Alloy simplifies configuration management across clusters and improves performance and reliability. As a result, centralized logging reduces maintenance overhead, improves observability performance, and aligns the monitoring stack with current Grafana best practices.

For more information, see Viewing centralized logs of the Ceph cluster.

(BZ#2398027)

Ceph File System (CephFS)

Case sensitivity and Unicode normalization can now be configured during subvolume group creation

Previously, it was possible to configure Unicode normalization and case sensitivity when creating a subvolume, but not when creating a subvolume group. To apply these settings, users had to run additional commands after the group was created.

With this enhancement, new command arguments allow users to configure Unicode normalization and case sensitivity directly during subvolume group creation, eliminating the need for extra steps.

(BZ#2359805)

Source information of clone subvolumes is now preserved

Previously, after cloning was completed, the source information (subvolume or snapshot) of the clone was removed from the `.meta` file. As a result, when users ran the **`subvolume info`** command for a clone subvolume, they could not view details about its source.

With this enhancement, source information for a clone subvolume is now preserved even after cloning is complete. This allows the **`subvolume info`** command to include details about the source subvolume in its output, making it easier for users to find and view the origin of a clone.

(BZ#2349154)

Now supports monitoring subvolume-level metrics

CephFS now provides performance metrics at the subvolume level, including IOPS, throughput, and latency. These metrics help administrators monitor IO allocations for applications and protocol gateways that use CephFS subvolumes. Metrics are available through Prometheus, the Ceph Manager stats module, and the Ceph Dashboard.

For more information, see Viewing subvolume metrics for CephFS metadata server clients .

Ceph Object Gateway

Bucket logging support for Ceph Object Gateway with bug fixes and enhancements

Bucket logging was previously available as limited release. This enhancement provides full availability for new and existing customers in production environments.

Bucket logging provides a mechanism for logging all access to a bucket. The log data can be used to monitor bucket activity, detect unauthorized access, get insights into the bucket usage and use the logs as a journal for bucket changes. The log records are stored in objects in a separate bucket and can be analyzed later.

Bucket logging includes support for source and destination buckets across different tenants, suffix/prefix-based key filtering, and standardized AWS operation names in log records.

For more information, see Bucket logging.

(BZ#2308169, BZ#2341711)

Restore objects transitioned to remote cloud endpoint back into Ceph Object gateway using the `cloud-restore` feature

The `cloud-restore` feature was previously available as limited release. This enhancement provides full availability for new and existing customers in production environments.

This feature allows users to restore objects transitioned to remote cloud endpoint back into Ceph Object gateway, using either S3 restore-object API or by re-hydrating using read-through options.

For more information, see Using the radosgw-admin cli for cloud restore operations.

(BZ#2293539)

New support for updating the restoration period for archived objects

With this enhancement, you can now update the expiry date of a restored object by reissuing the restore-object API request with a new restoration period. The updated period is calculated from the current time, allowing you to retain data longer or expire it sooner without re-downloading from the remote cloud endpoint.

For more information, seeRestoring objects from S3 cloud-tier storage.

(BZ#2312937)

New CLI commands introduced to help monitor and debug restore operations

Previously, administrators had limited visibility into object restore operations, which made monitoring and debugging difficult.

With this enhancement, the system introduces two new CLI commands:

`radosgw-admin restore list`

Lists the restore status of objects in a bucket.

`radosgw-admin restore status`

Displays restore attributes for a specific object.

The bucket statistics also include restore-related information for easier monitoring.

For more information, see Using the radosgw-admin cli for cloud restore operations.

(BZ#2345487)

New Ceph Object Gateway S3 support for IBM COS accelerated archive

Previously, restore operations for the S3-Cloud-Glacier type with IBM COS failed because the system passed Glacier-specific parameters intended for AWS, which were not applicable to other cloud vendors.

With this enhancement, the system introduces a new `glacier_restore_tier_type` value, `NoTier`, to handle these scenarios and enable successful restores for IBM COS accelerated storage class (equivalent to AWS Glacier).

For more information, see Compatibility matrix for Red Hat Ceph Storage 9.0 and Transitioning data to IBM Cloud Object Store (COS) service.

(BZ#2365095)

Improved CLI output for topic management

The radosgw-admin topic list command has been enhanced for better usability. The output format is now consistent across v1 and v2 topics and excludes the `topics` section, reducing complexity for automation and scripting.

(BZ#2360425)

Enhanced conditional operations

This enhancement introduces support for conditional `PUT` and `DELETE` operations, including bulk and multi-delete requests. These conditional operations improve data consistency for some workloads.

> **Note:** The conditional `InitMultipartUpload` is not implemented in this release.

(BZ#2375000, BZ#2350732)

Flushed object name now emitted

Previously, users had no direct way to identify the last object that was flushed. This made it harder to determine the correct starting point when traversing log objects in the log bucket.

With this enhancement, the system now replies with the name of the last flushed object. As a result, users can easily identify the most recent object and streamline log traversal operations.

(BZ#2364399)

Reduced client impact during bucket resharding

With this enhancement, bucket resharding now does most of its processing before it starts to block write operations. This should significantly reduce the client-visible impact of resharding on large buckets.

(BZ#2303488)

Committed objects now added to log buckets even without pending records

Previously, when committing an object, it was not added to the log bucket if there were no log records pending. This made it harder for consumers to reliably determine the last committed object when listing log bucket contents.

With this enhancement, committed objects are now added to the log bucket even if no log records are pending. As a result, consumers can easily identify the last committed object and traverse log objects more efficiently.

(BZ#2394062)

Clear error propagation for logging failures in journal mode

Previously, when logging failed in journal mode, the customer received generic or misleading error messages. For example, a customer performing a regular S3 operation could see a 403 error if permissions were missing on the log bucket, even though permissions were correct on the target bucket.

With this enhancement, the system now propagates a clear error message indicating that the failure occurred during logging, not the primary operation. As a result, customers can quickly identify and resolve logging-related issues without confusion.

(BZ#2395210)

Automatic permission setting for D3N cache directory

Previously, configuring the RGW D3N cache directory required manual steps to set permissions, such as running **`chmod a+rwx rgw_d3n_l1_datacache_persistent_path`**. This added complexity and increased setup time.

With this enhancement, the correct permissions are automatically applied when the D3N cache directory is created. As a result, customers experience fewer manual configuration steps, improving setup efficiency and overall usability.

(BZ#2239586)

New support for AWS S3 *GetAccountSummary*

Previously, AWS S3 *GetAccountSummary* was not supported, which limited certain workloads that require account-level information, such as Terraform-based automation.

With this enhancement, AWS S3 *GetAccountSummary* is now supported.

(BZ#2381576)

New support for AWS *STS GetCallerIdentity*

Previously, AWS *STS GetCallerIdentity* was not supported, limiting the ability to validate user identities and enforce access policies before creating or modifying policies. This gap impacted workflows that rely on identity verification, such as Terraform-based automation.

With this enhancement, AWS *STS GetCallerIdentity* is now supported. As a result, customers can securely validate identities and access policies, enabling more robust policy management and seamless integration with Terraform workflows.

(BZ#2381577)

Multi-site Ceph Object Gateway

Improved reliability for multi-site replication data log delivery

Previously, in rare cases, replication data logs could lose updates, which created the appearance of stalled replication even though data consistency was not affected.

With this enhancement, the multi-site replication process is hardened to prevent such occurrences. As a result, replication performance is smoother, and log reduction happens more promptly, improving overall system responsiveness.

(BZ#2053348)

Cleanup added for index segments of replicated buckets

Previously, dynamic resharding with multi-site replication had a long-standing limitation: old index segments were not cleaned up due to simultaneous access to old and new index shards during replication. This resulted in persistent space leakage.

With this enhancement, cleanup for index segments of replicated buckets has been added. As a result, the space leakage issue is resolved, improving storage efficiency and overall system health.

(BZ#2400114)

Aligned operation names with AWS for consistent log integration

Previously, operation names in Ceph logs were inconsistent with the operation types used by AWS. This required different approaches for log consumption depending on whether Ceph or AWS logs were being processed.

With this enhancement, operation names in Ceph logs now match the names used in AWS logs. This alignment simplifies integration and makes log consumption more consistent across systems.

(BZ#2372311)

RADOS

Enhanced support for moving stretch mode to normal mode

Previously, Ceph clusters operating in stretch mode could not be reverted to normal mode without manual intervention.

With this enhancement, Ceph introduces a command that allows users to gracefully exit stretch mode.

```
ceph mon disable_stretch_mode CRUSH_RULE --yes-i-really-mean-it
```

Users can optionally specify a CRUSH rule to which all pools should be migrated. If no rule is provided, Ceph automatically selects a default replicated CRUSH rule.

(BZ#1892474)

Enhanced detection of network partitions under connectivity election strategy

Previously, monitors operating under the connectivity election strategy did not provide user-facing alerts when network partitions occurred.

With this enhancement, monitors can detect network partitions between themselves. The elected leader monitor evaluates connectivity scores shared by its peers to identify partitioned connectivity groups. When a netsplit is detected, monitors emit health warnings.

Example of complete location-level partitions warning

```
Netsplit detected between dc1 and dc2
```

Example of individual monitor disconnections warning

```
Netsplit detected between mon.a and mon.d
```

(BZ#2318936)

New ISA plugin support for erasure coded pools

Previously, erasure coded pools only supported Jerasure plugins.

With this enhancement, ISA plugin is now the default and both plugins are now supported.

General enhancements for RADOS and RADOS BlueStore

This version provides several enhancements for RADOS and RADOS BlueStore. These enhancements include the following:

BlueStore discard optimization

Actively triggers block device discards to prevent excessive queue growth on SSDs and improves performance on lower-grade drives.

Faster device scanning

`ceph-volume` scans devices up to 100-times faster, streamlining day-one cluster setup operations.

Improved write latency

Uses a single consolidated `fdatasync` call in the WAL to reduce latency and improve overall write performance in BlueStore.

RADOS OMAP iteration

Optimizes object map (OMAP) iteration to reduce latency during large-scale operations and improve responsiveness in complex workloads.

Erasure coding ratio support enhancements

This release introduces new support and qualification for 5+2 and 6+2 erasure coding ratios. These configurations deliver an optimal balance of performance, scalability, and cost efficiency, making them ideal for clusters that require high storage utilization and robust data protection.

For more information, see Erasure code profiles.

Ceph Block Device (RBD) mirroring

Improved tracking of mirror group snapshot states

Previously, `rbd-mirror` tracked the progress of a mirror group snapshot without distinguishing between a snapshot that was created and one that was fully synced.

With this enhancement, a new internal field (`complete`) is integrated into the `GroupSnapshotNamespaceMirror` structure. This field determines whether a snapshot is completely synced. The existing state field of creating and created continues to indicate whether teh snapshot has been created. Together, these fields provide a more precise distinction between snapshots that are created (metadata available) and those that are fully synced.

As a result, mirror group snapshot status tracking is more accurate and consistent, improving compatibility and robustness in the `rbd-mirror` process. The user-facing output of the **`rbd group snap ls`** command is also updated to reflect clearer state names: `creating` and `created` instead of `incomplete` and `complete`. A mirror group snapshot is completely synced on the secondary cluster when the `NAMESPACE` column shows as `copied`, and still syncing when it shows `not copied`.

(BZ#2396583)

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 Secure Token Service.

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.

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 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 Viewing Ceph Object Gateway per-user and per-bucket performance counters.

(BZ#2036531)

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 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 Track the data availability score of a cluster (Technology Preview).

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.

Ceph Dashboard

Ceph Object Gateway page now loads after a multi-site configuration

Previously, the Ceph Object Gateway page does not load because the dashboard could not find the correct access key and secret key for the new realm during multi-site configuration.

With this fix, the Ceph Object Gateway page can find the correct access and secret key for the new realm and loads as expected.

(BZ#2231072)

Ceph File System (CephFS)

`enctag` value length is now restricted to 255 characters

Previously, `enctag` could be stored with values longer than 255 characters. However, operations such as **`enctag get`** only supported displaying values up to 255 characters. If an `enctag` with a longer value was stored, the system returned a general `Unexpected error` output.

With this fix, the system now enforces a maximum `enctag` value length of 255 characters. Only valid `enctag` are accepted and stored, allowing operations such as **`enctag get`** to display the `enctag` successfully.

(BZ#2359400)

Unsupported encryption algorithms now return an error on CephFS

Previously, the CephFS userspace did not validate encryption algorithms when setting up `fscrypt`. Only `AES-256-XTS` and `AES-256-CTS` were supported, but if a different algorithm was requested, CephFS silently used the default supported algorithm without notifying the user.

With this fix, a validation check ensures that only supported encryption algorithms are allowed when setting up `fscrypt` on CephFS. If an unsupported algorithm is supplied, the system returns an `EINVAL` error code.

(BZ#2362686)

Ceph Object Gateway

Cloud-S3 restore requests no longer lost after Ceph Object Gateway restart

Previously, if the Ceph Object Gateway (RGW) service restarted while a restore request for the cloud-s3 cloud tier was in progress, the request state was lost because it was not stored persistently. As a result, the restore operation was not retried.

With this fix, the system now stores the state of restore requests persistently. If the Ceph Object Gateway service restarts during processing, the request is resumed automatically, ensuring continuity without manual intervention.

For more information, seeRestoring objects from S3 cloud-tier storage.

(BZ#2312933)

Local reads now work as expected

Previously, local reads were sometimes unavailable for recently created or modified RADOS objects due to protocol limitations.

With this fix, eligible reads can now be performed locally, improving consistency and reliability in all environments.

(BZ#2309383)

`max_objs_per_shard` no longer goes below a safe minimum

Previously, in versioned buckets, the **`max_objs_per_shard`** value was reduced by a factor of three to account for the additional index entries created by object versioning. In cases, such as debugging, this value could artificially be set low to trigger early resharding. With these two items in combination, these adjustments could result in a **`max_objs_per_shard`** value of `0`, leading to a division by 0 crash.

With this fix, the value can not be reduced below a safe minimum.

(BZ#2228033)

Ceph Object Gateway STS now supports encryption keys larger than 1024 bytes

Previously, the Ceph Object Gateway (RGW) STS implementation did not support encryption keys larger than 1024 bytes. Users had to manually adjust Keycloak settings by lowering the priority of the `rsa-enc-generated` provider and reducing the **`keySize`** to `1024`.

With this fix, RGW STS now supports encryption keys larger than 1024 bytes without requiring manual configuration changes in Keycloak. This improves security and simplifies setup.

(BZ#2276931)

Checksum type and checksum algorithm now display with uncompleted multipart uploads

Previously, when listing parts for uncompleted multipart uploads, the checksum type and checksum algorithm were missing because the logic to extract these fields was not implemented.

With this fix, the logic to extract the checksum type and checksum algorithm has been added, so these fields now appear when listing parts for uncompleted multipart uploads.

(BZ#2324147)

`radosgw-admin` no longer crashes by non-positive values

Previously, when running the **`radosgw-admin bucket reshard`** command, using a non-positive **`--num-shards`** value, such as a zero or a negative number, would cause `radosgw-admin` to crash.

With this fix, the **`--num-shards`** value is checked an error message is emitted if a non-positive value is provided. As a result, `radosgw-admin` reshard commands run as expected, and are not able to create a crash.

(BZ#2401174)

Empty string in the *HTTP_X_AMZ_COPY_SOURCE* header no longer causes crashing

Previously, the *HTTP_X_AMZ_COPY_SOURCE* header could have an empty string output, rather than NULL. When the empty string was passed to `RGWCopyObj::parse_copy_location()` the empty name would cause a crash.

With this fix, a check is in place that the header contains a valid string.

(BZ#2412218)

Multipart upload completion logs now include object size

Previously, the log record for multipart upload completion did not include the object size.

With this fix, the object size is now included in the completion log record.

(BZ#2365697)

Operation names in log records now match AWS naming conventions

Previously, the name of the operation in Ceph Object Gateway log records did not match the name used in AWS. As a result, consumers that could parse AWS generator records failed when processing records generated by Ceph Object Gateway.

With this fix, the operation name in log records is now consistent with AWS naming conventions. The same consumer can be used for records from both sources.

(BZ#2365931)

Bucket logging configuration changes no longer cause data loss

Previously, a race condition occurred when the bucket logging configuration was changed while the system was running. This caused log objects to be garbage collected, and log records were lost after the garbage collector ran.

With this fix, the race condition has been resolved. Users can now safely change bucket logging configurations without risking data loss.

(BZ#2393440)

Copied objects from versioned to non-versioned buckets are now accessible

Previously, when copying an object from a versioned bucket to a non-versioned bucket, some versioning attributes were mistakenly copied to the destination object. This could make the copied object inaccessible.

With this fix, versioning attributes are removed when copying to a non-versioned bucket and the copied object is now accessible.

(BZ#2390658)

Data transition to AWS non-default regions is now supported

Previously, the cloud tier module did not handle the **`location_constraint`** parameter required by AWS when creating a bucket in a non-default region. As a result, data transition to an AWS cloud endpoint failed if the target bucket was in a non-default region.

With this fix, a new parameter, **`location_constraint`**, has been added to the **`tier_config`** configuration. This parameter must be set or updated along with region when using AWS non-default regions. Data can now be transitioned to AWS non-default regions successfully.

(BZ#2402662)

Tenant user policy and role-based permissions now work as expected after upgrade

Previously, some policy or role-based permissions involving legacy tenant users behaved differently after upgrading to releases that support IAM accounts. As a result, expected access grants would fail.

With this fix, a configuration option has been introduced to allow backward compatibility with previous version behavior.

(BZ#2416754)

Predefined ACLs are now correctly matched

Previously, reversed logic in the comparison functor for predefined ACL matching caused all predefined ACLs to be rejected.

With this fix, calls to `compare()` have been replaced with `operator==` and predefined ACLs now match correctly.

(BZ#2344639)

Permission checks for multipart upload initialization are now correct

Previously, the permission check for InitMultipart incorrectly used the bucket Amazon Resource Name (ARN) instead of the object ARN. This could cause *PutObject* requests to fail unexpectedly.

With this fix, permission checks now use the object ARN as intended. Multipart upload initialization and subsequent object operations work correctly.

(BZ#2362318)

Improved bulk delete performance in versioned buckets

Previously, the Ceph Object Gateway object deletion logic was inefficient, particularly in how it invoked `update_olh`. This caused high latency and could eventually lead to system lockups when processing bulk deletes of up to 1,000 object versions per request in versioned buckets.

With this fix, `update_olh` is now limited to a single invocation per bulk-delete request. This change significantly improves system behavior under heavy bulk-delete workloads.

(BZ#2387764)

ACL checks after `AssumeRole` are now correctly enforced

Previously, incorrect logic failed to verify ACLs after an `AssumeRole` operation. As a result, checks for explicit ACL grants failed incorrectly.

With this fix, `RoleApplier::get_perms_from_aclspec()` now calls `rgw_perms_from_aclspec_default_strategy()` to check for matching ACL grants. Additionally, missing `RoleApplier` support has been added to grant access based on ACLs.

(BZ#2406837)

Log records now correctly indicate ACL-based authorization

Previously, the `aclRequired` field in the log record would display as with a hyphen (`-`), even when the request was authorized by an ACL. This was misleading because it suggested that the operation was authorized by a bucket policy.

With this fix, the field is set to `Yes` whenever a request is authorized by an ACL.

(BZ#2371110)

Log records now correctly indicate authentication type for unauthenticated requests

Previously, the `AuthenticationType` field in the log record was incorrectly set to `QueryString` for unauthenticated requests.

With this fix, the field is set to hyphen (`-`) for unauthenticated requests.

(BZ#2371109)

IAM policy now recognizes `AbortMultipartUpload` `Deny` requests

Previously, a session policy incorrectly used the `AbortMultipartUpload` action. As a result, a `Deny` statement for `AbortMultipartUpload` in the IAM policy was not respected when `PutObject` was allowed.

With this fix, the action in the IAM policy was corrected. The `Deny` for `AbortMultipartUpload` in the IAM policy is now properly enforced.

(BZ#2302541)

Delete bucket policy now returns correct status code

Previously, the delete bucket policy operation returned HTTP status code `204`, instead of the correct `200` code.

With this fix, the HTTP status code was corrected, and delete bucket policy now returns `200`, as expected.

(BZ#2343728)

`api_name` field now initializes during zone group rename

Previously, the **`api_name`** field in the zone group map was not initialized during a zone group rename because the variable assignment was missing.

With this fix, the **`api_name`** variable is now assigned. The **`api_name`** field is now correctly initialized during zone group rename.

(BZ#2366182)

Multipart object decryption now works for `partNumber` requests

Previously, if a multipart object was encrypted using SSE-C or SSE-S3, a get object request with `partNumber` did not decrypt the part.

With this fix, the logic was updated to attach the saved crypt prefix, if present, when the get action is a get-part operation. This enables Ceph Object Gateway to decrypt the part for get object requests with `partNumber`.

(BZ#2315856)

Multi-site Ceph Object Gateway

Rate limit configurations now synchronize across all Ceph Object Gateway multi-site zones

Previously, user and bucket quota rate limit configurations set on a secondary Ceph Object Gateway site were not synchronized back to the primary site in a multi-site setup. This caused configuration inconsistencies, resulting in different rate-limiting behaviors between zones.

With this fix, the metadata synchronization process for Ceph Object Gateway multi-site has been improved. Rate limit updates from secondary zones are now correctly propagated to the primary zone, validated, and applied. All user and bucket rate limit configurations now synchronize across all zones, ensuring consistent behavior throughout the Ceph Object Gateway multi-site cluster.

(BZ#2393477)

Ceph Object Gateway multi-site now automates cleanup of deleted bucket instances and index objects

Previously, deleted bucket instances and related index objects were retained and not trimmed during bilog trimming to allow multi-site sync to finish processing deletions on other zones.

With this fix, a mechanism has been introduced to automatically clean up deleted bucket instances and index objects across all zones as part of the bucket index log trimming process. Additionally, the *DeleteBucket* API now returns `409 BucketNotEmpty` errors until the bucket is empty on all zones when a sync policy is enabled.

(BZ#1696875)

Object lock configuration rule is now synchronized to the secondary zone

Previously, when object lock was enabled on a bucket, the rule failed to replicate to other zones, causing object lock inconsistencies between zones.

With this fix, multi-site replication now synchronizes the object lock configuration to all zones, ensuring consistent buckets across zones.

(BZ#2317768)

`RGWBucketFullSyncCR` no longer spins indefinitely when the source bucket has been deleted

Previously, the coroutine `RGWBucketFullSyncCR` reused the `bucket_list_result` member without clearing its prior state. Stale entries and the **`is_truncated`** flag from a previous iteration could persist, causing the loop to continue even after the bucket was deleted.

With this fix, the constructor of `RGWBucketFullSyncCR` clears the provided `bucket_list_result` at the start. This ensures that each listing begins with a clean state and accurately reflects the current remote bucket contents.

(BZ#2412220)

Known issues

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

`cephadm` utility

Grafana certificate does not migrate during upgrade

When you upgrade from Red Hat Ceph Storage 8.1 to 9.0, the existing user-signed Grafana certificate is not migrated. Instead, Grafana switches to a `cephadm`-signed certificate. As a result, duplicate certificate entries may appear, and certificate-related health warnings can persist. Manual reconfiguration is required if you want to use custom TLS certificates.

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

To work without custom TLS certificates, you can continue using the `cephadm`-signed certificate.

As a workaround to use custom TLS certificates, complete the following steps:

1. Change the Grafana specification to use `certificate_source: reference`.
2. Use `certmgr` to upload a valid user-signed certificate and key for each host.
3. Run the **`ceph orch reconfig grafana`** command.

(BZ#2414999)

Management gateway does not open HTTPS port during deployment

When the management gateway (`mgmt-gateway`) is deployed with default settings and `firewalld` is active, the default HTTPS port (443) is not opened in `firewalld`. The gateway listens on port 443 and is reachable locally, but remote access to the dashboard fails until the firewall is manually adjusted.

As a workaround, use one of the following options:

- Explicitly configure a port for `mgmt-gateway` by using the **`--port`** option or setting `spec.port`. This ensures that `cephadm` opens the correct port in `firewalld`.
- Manually open HTTPS (443) in `firewalld`. For example,

```
firewall-cmd --add-service=https
firewall-cmd --add-port=443/tcp
```

(BZ#2417683)

`cephadm` operations may fail when interactive shell aliases are present

In Red Hat Ceph Storage 7.1, `cephadm` uses the shell `mv` command on remote hosts. If the `cephadm` SSH user has interactive aliases such as `` `mv='mv -i' `` (and similar for `rm` or `cp`), these aliases trigger prompts and block `cephadm` operations. As a result, commands like **`ceph orch upgrade`**, **`cephadm bootstrap`**, or adding hosts may hang or fail because `mv` waits for user confirmation instead of running non-interactively.

Currently there is no workaround. To avoid this issue, remove or disable interactive aliases for `mv`, `rm`, and `cp` for the `cephadm` SSH user. For example, comment them out in `.bashrc` or define them only for interactive shells, then rerun the `cephadm` operation.

(BZ#2360008)

Promtail image remains visible after migration to Alloy

During the transition from Promtail to Alloy, `cephadm` 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. The behavior is intentional to prevent breaking log collection on clusters that have not fully migrated.

No workaround is required. Ignore the Promtail image entry during the supported transition phase. If log collection has fully migrated to Alloy and is verified, you can optionally remove legacy Promtail daemons and images manually. This cleanup is not required for cluster operation.

(BZ#2418617)

Ceph build

QAT cannot be used for TLS offload or acceleration mode together with SSL set

Enabling QAT on HAProxy with SSL enabled injects legacy OpenSSL engine directives. The legacy OpenSSL engine path breaks the TLS handshake, emitting the `tlsv1 alert internal error` error. With the TLS handshake broken, the TLS termination fails.

As a workaround, disable the QAT at HAProxy in order to keep the TLS handshake. Set the configuration file specifications as follows:

- `haproxy_qat_support: false`
- `ssl: true`

As a result, QAT is disabled and the HAProxy TLS works as expected.

> **Note:** Under heavy connection rates higher CPU usage may be seen versus QAT-offloaded handshakes.

(BZ#2373189)

HAProxy deployment fails when QAT is enabled with ingress

Deploying HAProxy with the QAT feature enabled fails on Red Hat Ceph Storage 9 container images when using the ingress feature. This occurs because HAProxy no longer supports `ssl_engine` in default builds. In addition, newer OpenSSL versions have removed the legacy engine used by QAT, making them incompatible. Attempts to use older OpenSSL versions or build a QAT provider for newer versions also lead to compatibility issues. As a result, HAProxy cannot run with QAT enabled, and deployment fails.

There is no way to enable QAT with HAProxy. To continue using HAProxy without QAT, update the HAProxy configuration file (typically located at `/var/lib/haproxy/haproxy.cfg`) as follows:

```
haproxy_qat_support: false
ssl: true
```

(BZ#2406166)

Ceph Dashboard

Active alert displays even when Prometheus module is active

In some cases, the Ceph Dashboard shows an active alert for `CephMgrPrometheusModuleInactive` even though the Prometheus module is enabled. This can happen due to a cluster misconfiguration that causes the Ceph target to go down, falsely triggering the alert. The alert remains visible unless silenced, even when the Prometheus module is functioning correctly.

As a workaround, to suppress the alert, from the Ceph Dashboard, select the `CephMgrPrometheusModuleInactive` alert and create a silence.

Observability › Alerts › CephMgrPrometheusModuleInactive › Create Silence

For more information, see Managing alerts.

(BZ#2187272)

Dashboard cannot delete non-default zone groups or zones

Users cannot delete non-default zone groups or zones from the Ceph Dashboard. Attempts to delete them fail.

As a workaround, delete non-default zone groups and zones through the command-line interface by using the appropriate **`radosgw-admin`** commands.

(BZ#2406519)

Ceph File System (CephFS)

Subvolume operations delayed due to GIL contention during asynchronous cloning

When the asynchronous cloner in the volumes module (`mgr`/volumes) uses the CephFS Python binding, it invokes the Ceph client library API while holding the Python Global Interpreter Lock (GIL). During asynchronous clone operations, the GIL remains locked for an extended period, which prevents other CephFS subvolume operations such as create and delete from acquiring the GIL in time. As a result, customers may experience delayed responses when performing subvolume operations.

As a workaround, temporarily pause cloning to allow other subvolume operations to proceed.

> **Important:** This workaround is not practical in most production environments and should be used only in exceptional cases.

(BZ#2429623)

Ceph Object Gateway

Lifecycle processing stuck in PROCESSING state for a given bucket

If a Ceph Object Gateway server is unexpectedly restarted when the lifecycle processing is in progress for a given bucket, that bucket does not resume processing lifecycle work for at least two scheduling cycles and is stuck in PROCESSING state.

This is an expected behavior as it is intended to avoid multiple Ceph Object gateway instances or threads from processing the same bucket simultaneously, especially when the debugging is in progress in production.

Currently there is no workaround.

(BZ#2401203)

Ceph Object Gateway services down after upgrade

After upgrading, Ceph Object Gateway services may fail to start. The service fails to start as the `rgw` service now enforces the `rgw_realm` configuration but no realm exists in the Ceph Object Gateway configuration. As a result, the following outputs occur:

- The Ceph Object Gateway logs show the following error:

  ```
  rgw main: failed to load zone: (2) No such file or directory
  ```

- The **`ceph orch ps | grep rgw`** output displays Ceph Object Gateway in an error state.
- Ceph Object Gateways are missing from **`ceph versions`**.

As a workaround, removing the `rgw_realm` entry and restart all Ceph Object Gateway services.

1. Verify if the Ceph Object Gateways are configured with no realm indicated while the Ceph configuration database specifies a realm.
   a. Check the Ceph Object Gateway realm list.

   ```
   radosgw-admin realm list
   ```

   The following is an example with an empty realm list:

   ```
   [ceph: root@host01 /]# radosgw-admin realm list
   {
                              "default_info": "",
                              "realms": []
   }
   ```

   b. Check the Ceph configuration database.

   ```
   ceph config dump | egrep "^WHO|rgw_realm”
   ```

   For example,

   ```
   [ceph: root@host01 /]# ceph config dump | egrep "^WHO|rgw_realm
   WHO                                MASK      LEVEL     OPTION       VALUE
   xxxxx.yyyyy                                  advanced  rgw_realm    default
   ```

   If step 1a matches step 1b, continue to step 2 to remove the `rgw_realm` from the Ceph configuration database.

   If the two steps do not match, contact Red Hat Support.

2. Remove the `rgw_realm` from the Ceph configuration database.

   ```
   ceph config rm xxxxx.yyyyy rgw_realm
   ```

3. Restart all Ceph Object Gateway services.

   ```
   ceph orch restart rgw
   ```

(BZ#2365888)

Ceph Object Gateway multi-site

Sync failure occurs after renaming a zone or zone group

Renaming a zone or zone group in the Primary zone in the `master_zonegroup` can cause sync failures. When sync failures occur, the following sync status error is may be emitted and further sync operations are affected:

```
failed to retrieve sync info: (2200) Unknown error 2200
```

As a workaround, before renaming a zone or zone group in the `master_zonegroup`, remove the old zone or zone group name from the Ceph configuration file. For more information, see Removing a zone from a zone group and Renaming a zone group.

(BZ#2210695)

Secondary site continues to display old zone group name after rename

In some cases, when a zone group is renamed on the Primary site, the Secondary site may still display the old zone group name. This occurs because the old name is not removed from the `.rgw.root` pool after the rename operation. As a result, both the old and new zone groups appear under the **`radosgw-admin zonegroup list`** command, and sync operations may be impacted.

As a workaround, complete the following steps.

1. Verify that the new zone group name exists.

```
radosgw-admin zonegroup list
```

2. List the `.rgw.root` pool and locate the old zone group name.

```
rados -p .rgw.root ls
```

   The old name appears in the format:
   `zonegroups_names.OLD_ZONEGROUP_NAME`

3. Remove the old zone group name from the pool.

```
rados -p .rgw.root rm zonegroups_names.OLD_ZONEGROUP_NAME
```

Removing the old zone group name restores normal sync operations.

(BZ#2416993)

Multi-site lifecycle expiration does not clean OLH entries in versioned buckets

Multi-site lifecycle expiration may fail to remove object log header (OLH) entries in versioned buckets. The system leaves stale data in the bucket index. This issue occurs when lifecycle expiration runs on multi-site deployments for versioned buckets. As a result, stale OLH entries remain after object deletion. This causes bucket index bloat and may impact bucket operations for customers.

As a workaround, administrators can manually detect and repair the affected buckets.

1. Detect stale entries.

```
radosgw-admin bucket check olh --dump-keys --bucket=BUCKET_NAME --hide-progress
```

2. Repair the bucket index.

```
radosgw-admin bucket check olh --fix --bucket=BUCKET_NAME
```

After repair, stale entries are purged.

(BZ#2430821)

RADOS

Placement groups are not scaled down in `upmap-read` and `read` balancer modes

Currently, `pg-upmap-primary` entries are not properly removed for placement groups (PGs) that are pending merge. For example, when the bulk flag is removed on a pool, or any case where the number of PGs in a

pool decreases. As a result, the PG scale-down process gets stuck and the number of PGs in the affected pool do not decrease as expected.

As a workaround, remove the `pg-upmap-primary` entries in the OSD map of the affected pool. To view the entries, run the **`ceph osd dump`** command and then run **`ceph osd rm-pg-upmap-primary PG_ID`** for reach PG in the affected pool. After using the workaround, the PG scale-down process resumes as expected.

(BZ#2302230)

Ceph Block Device (RBD)

Kernel client does not support `pg-upmap-primary`

The kernel client currently does not support the `pg-upmap-primary` feature. As a result, users may encounter issues when attempting to mount images or filesystems using the kernel client in environments where `pg-upmap-primary` is configured.

If issues occur during mounting with the kernel client, verify that the issue is due to this support issue.

1. Confirm that your cluster contains `pg-upmap-primary` mappings.

```
ceph osd dump | grep "pg_upmap_primary"
```

2. Check the kernel log for the following error message.

```
$ dmesg | tail

[73393.901029] libceph: mon2 (1)10.64.24.186:6789 feature set mismatch, my
2f018fb87aa4aafe < server's 2f018fb8faa4aafe, missing 80000000
[73393.901037] libceph: mon2 (1)10.64.24.186:6789 missing required protocol
features
```

These error details confirm that the cluster is using features that the kernel client does not currently support.

- If this error message is not emitted, contact Red Hat Support.
- With this error message continue by removing the related mappings.

As a workaround, remove the related `pg-upmap-primary` mappings.

1. If using the balancer module, change the mode back to one that does not use `pg-upmap-primary`.

```
ceph balancer mode upmap
```

2. Remove all `pg-upmap-primary` mappings.

```
ceph osd rm-pg-upmap-primary-all
```

(BZ#2304313)

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/

Asynchronous updates

This section describes the bug fixes, known issues, and enhancements for asynchronous updates.

Release notes for 9.0z1

An update is now available for Red Hat Ceph Storage 9.0z1 which provides bug fixes and a new known issue.

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.

Ceph File System (CephFS)

Improved xattr dump handling to prevent MDS crash

Previously, the MDS could crash when handling xattrs in `CInode.cc` due to empty `bufptr` values being dumped.

With this fix, the code now checks whether the buffer contains data before dumping it, and explicitly dumps an empty string when the buffer length is zero. This prevents spurious empty buffer entries and ensures safe handling of xattr values. As a result, xattr dumps are cleaner and more accurate, and the MDS no longer crashes in this scenario.

(IBMCEPH-12883)

Subvolume operations no longer blocked during asynchronous clone

Previously, the CephFS Python binding used by the asynchronous cloner in the volumes module invoked the client library API while holding the Python Global Interpreter Lock (GIL). Because the GIL was held for an extended duration, other subvolume operations in the volumes module were blocked while waiting to acquire the lock.

With this fix, the CephFS client API is now invoked without holding the GIL. As a result, subvolume operations in the volumes module can progress normally even when an asynchronous clone operation is running.

(IBMCEPH-12760)

MDS crash due to NULL pointer dereference prevented

Previously, the MDS could crash when a NULL MDRequestRef pointer was dereferenced. With this fix, the logic now returns early when the MDRequestRef is NULL, instead of attempting to dereference it.

As a result, crashes caused by this condition are prevented, improving overall MDS stability.

(IBMCEPH-12900)

Ceph Object Gateway

Object unlink handling updated for zero-shard configuration

Previously, the system did not correctly handle object unlink operations in specific zero-shard configurations.

With this fix, the code has been updated to ensure proper handling when both **`bucket_index_max_shards`** and the bucket's **`num_shards`** are set to `0`. As a result, object unlink operations now succeed in this scenario.

(IBMCEPH-12915)

Updated processing logic to ensure all topics are handled

Previously, only the first 1,000 topics were repeatedly processed, preventing the remaining topics from being handled as expected.

With this fix, the system now processes topics in batches of 1,000, ensuring that all topics are eventually processed rather than cycling over only the initial set. As a result, bucket notifications are now sent for all topics, and topic queues no longer fill up or block service operation.

(IBMCEPH-12914)

All bug fixes

This section lists a complete listing of all bug fixes in this release of Red Hat Ceph Storage 9.0z1.

Table 1: List of all bug fixes

Issue key	Severity	Summary
IBMCEPH-12760	Critical	Shallow Clone does not work as expected when an RWX clone is in progress.
IBMCEPH-12841	Critical	pg_autoscaler is calculating correctly but not implementing PG counts and changes, due to high threshold
IBMCEPH-10918	Important	Allow Ingress service to expose the metrics via HTTPS
IBMCEPH-12770	Important	unix attributes stored on objects appear not to be persistent (as required)
IBMCEPH-12827	Important	Unexpected error getting (earmark|encryption tag): error in getxattr: No data available [Errno 61]
IBMCEPH-12883	Important	MDS crashed executing asok_command: dump tree with assert ceph::__ceph_assert_fail(char const*, char const*, int, char const*)
IBMCEPH-12900	Important	ceph-mds crashed - mds-rank-fin
IBMCEPH-12903	Important	ceph-crash not authenticating with cluster correctly
IBMCEPH-12906	Important	COT get attribute command fails for BlueFS ENOSPC OSD
IBMCEPH-12914	Important	notification code will go into infinite loop when there are more than 1K topics
IBMCEPH-12981	Important	RBD Group mirror snapshots remain in "created" / "not copied" state after rbd-mirror daemon stop and kill
IBMCEPH-12982	Important	Group resync does not recover snapshots stuck in ‘not copied’ state
IBMCEPH-12999	Important	RPMInspect fails on executable stack
IBMCEPH-13113	Important	radosgw-admin 'bucket rm --bypass-gc' ignores refcount (can lead to DL)
IBMCEPH-12786	Moderate	Fix AdminOps API GetAccount() and DeleteAccount()
IBMCEPH-12853	Moderate	ECDummyOp memory leak in Fast EC
IBMCEPH-12915	Moderate	When `bucket_index_max_shards` is set to `0` in the zone group and the bucket has `num_shards` set to `0`, the object unlink operation fails

Security fixes

This section lists security fixes from this release of Red Hat Ceph Storage 9.0z1.

For details about each CVE, see CVE Records.

- CVE-2021-23358

- CVE-2024-51744
- CVE-2024-55565
- CVE-2025-22868
- CVE-2025-26791
- CVE-2025-66418
- CVE-2025-66471
- CVE-2025-7783

Known issues

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

Ceph Object Gateway multi-site

Bucket index shows stale metadata after lifecycle expiration in versioned buckets

In rare cases, when lifecycle expiration removes objects from versioned buckets, some omap entries in the bucket index might remain even though the objects have already been removed.

As a result, some omap entries may remain in the bucket index. In cases that many leftover keys accumulate, the following error is emitted: `(27) File too large`. This inconsistency can affect tools or processes that depend on accurate bucket index listings.

As a workaround:

1. Scan the bucket for leftover keys.

```
radosgw-admin bucket check olh --bucket=testbucket --dump-keys --hide-progress
```

2. Remove the leftover omap entries.

```
radosgw-admin bucket check olh --bucket=testbucket --fix
```

(IBMCEPH-12980)

Release notes for 9.0z2

An update is now available for Red Hat Ceph Storage 9.0z2 which provides new features, enhancements, and bug fixes. There are also new known issues in this release.

New features and enhancements

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

Ceph build

More cluster deployment support options

ARM-based platform cluster deployment support was previously available as limited release. This enhancement provides full availability for new and existing customers in production environments.

Red Hat Ceph Storage clusters now support cluster deployments on ARM 64 (aarch64) CPUs. This enhancement enables you to consume Ceph storage from cost-effective AWS Graviton instances and other ARM-based platforms.

(ISCE-1400)

PKCE enforcement for OAuth2 authorization code flow

PKCE (Proof Key for Code Exchange) enforcement has been added to python-oauthlib for the OAuth2 authorization code flow. This enhancement strengthens OAuth2 authentication by providing protection against man-in-the-middle (MITM) attacks.

With this change, applications that use OAuth2 through python-oauthlib can enforce PKCE as part of the authorization process, improving overall authentication security.

(IBMCEPH-12630)

Ceph Object Gateway

Improved Object Gateway Lifecycle Processing Performance

This release enhances Ceph Object Gateway lifecycle (LC) processing performance for buckets with multiple, overlapping lifecycle rules.

The improvement groups lifecycle rules that share the same prefix and object tag conditions, allowing Object Gateway to enumerate objects and fetch tags only once instead of performing multiple passes for each rule. This reduces unnecessary I/O and improves concurrency during lifecycle execution.

As a result, lifecycle operations complete faster and more predictably for large buckets with complex lifecycle configurations, improving overall efficiency in high-scale environments.

(IBMCEPH-13353)

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

Optional installation of service dependencies during `cephadm` prepare-host

An issue that prevented administrators from controlling the installation of service dependencies when running the `cephadm` prepare-host command has been fixed.

Previously, the command always installed required system packages and dependencies, even in environments where hosts were already preconfigured or managed by external tools.

With this fix, `cephadm` prepare-host now supports more flexible host preparation behavior, allowing administrators to better align dependency management with their existing provisioning and automation workflows.

(IBMCEPH-13460)

`cephadm`-ansible playbooks correctly resolve the `ceph_config` module

An issue that caused `cephadm`-ansible playbooks to fail due to an unresolved `ceph_config` module has been fixed.

Previously, playbooks that relied on the `ceph_config` module could not run successfully, even though the same playbooks worked as expected in earlier releases. This regression prevented automation workflows and tests from retrieving or updating Ceph configuration values through Ansible.

With this fix, the `ceph_config` module is resolved correctly, and `cephadm`-ansible playbooks that depend on this module now run successfully as expected.

(IBMCEPH-13542)

Alertmanager redeployment no longer fails during upgrade when using legacy custom templates

Previously, during an upgrade from Red Hat Ceph Storage 6.1z9 to 8.1z1, the `cephadm`-managed Alertmanager service failed to redeploy if a custom `alertmanager.yml` template referenced the `default_webhook_urls` variable. Template rendering failed with an undefined variable error because this variable is no longer guaranteed in the rendering context.

With this fix, Alertmanager redeploys successfully during upgrade, allowing the upgrade process to continue without interruption and preventing the cluster from entering a `HEALTH_WARN` or incomplete upgrade state.

(IBMCEPH-12929)

Ceph build

gRPC reflection support for the new gRPC service

An issue that prevented gRPC reflection from being available for the new gRPC service has been fixed.

Previously, gRPC clients that rely on reflection were unable to discover service descriptors automatically, which caused requests to fail unless the service definition was provided explicitly.

With this fix, the gRPC reflection capability is available as expected. gRPC clients can now discover service descriptors dynamically, enabling standard tooling and workflows to interact with the service without requiring manual proto file configuration.

(IBMCEPH-13627)

Ceph version string correctly reflects production build after upgrade to RHCS 9

After upgrading a Red Hat Ceph Storage (RHCS) 8 cluster to RHCS 9, the Ceph version output could display the string `rc - RelWithDebInfo`.

Previously, this caused confusion by suggesting that a release candidate or debug build was running, even though the cluster was using a supported production image.

With this fix, the Ceph version has been fixed. The Ceph version string now correctly reflects the production build status after upgrading to RHCS 9.

(IBMCEPH-13276)

Ceph File System (CephFS)

Listing encrypted case-insensitive CephFS directories

An issue that caused errors when listing the contents of encrypted, case-insensitive CephFS directories has been fixed.

Previously, directory listing operations could fail when both encryption and case-insensitive directory features were enabled, even though the data existed and permissions were correctly configured.

With this fix, directory contents can now be listed successfully when encryption and case-insensitive directory support are enabled together.

(IBMCEPH-13452)

Ceph Object Gateway

Intermittent HTTP 409 errors for Ceph Object Gateway S3 requests

An issue that caused Ceph Object Gateway to intermittently return HTTP 409 `ConcurrentModification` errors for S3 requests has been fixed.

Previously, these errors could occur during normal S3 operations, including GET requests, even when no conflicting client activity was expected.

With this fix, Ceph Object Gateway handles concurrent access scenarios more reliably, reducing unexpected request failures and improving consistency for S3 client operations.

(IBMCEPH-13756)

GET requests for multipart-uploaded objects return correct results

An issue that caused GET requests to fail with HTTP 404 for multipart-uploaded objects, even though HEAD requests succeeded, has been fixed.

This inconsistency could lead applications to incorrectly assume that objects were unavailable after validating their existence.

With this fix, GET requests for multipart-uploaded objects now behave consistently, ensuring reliable access to objects uploaded using multipart upload.

(IBMCEPH-13618)

Ceph Object Gateway successfully starts after upgrade

An issue that caused the Object Gateway service to fail to start after upgrading to Red Hat Ceph Storage 9.0 and 9.0z1 has been fixed.

Previously, Ceph Object Gateway could enter an error state during startup, preventing object storage services from becoming available after the upgrade.

With this fix, Ceph Object Gateway initializes successfully following the upgrade, restoring object storage availability and normal administrative operations.

(IBMCEPH-13877)

All bug fixes

This section lists a complete listing of all bug fixes in this release of Red Hat Ceph Storage 9.0z2.

Table 2: List all bug fixes

Issue key	Summary
IBMCEPH-10839	Squid deployed OSDs are crashing with !ito->is_valid()in BlueStore::Blob::copy_extents_over_empty()
IBMCEPH-11376	QoS Bandwidth limit shows PerClient behavior when set as PerShare behavior
IBMCEPH-11414	IOPS,throughput and latency perf panels do not show data when filtered by time picker
IBMCEPH-12396	Cluster QoS default port number modification and corresponding firewall rule
IBMCEPH-12630	oauthlib: Missing PKCE enforcement in OAuth2 Authorization Code Flow in oauthlib
IBMCEPH-12758	Replication and application resources stuck during application deletion
IBMCEPH-12787	NFS crashes when the ingress mode is configured as haproxy-protocol
IBMCEPH-12823	Namespace creation error help message has wrong command for namespace list
IBMCEPH-12916	Allow formatting availability score in JSON
IBMCEPH-12929	Upgrade RHCS 6.1z9 to RHCS 8.1z1 failed due to custom alertmanager.yml
IBMCEPH-12935	enable_cluster_qos field is not displayed when qos is enabled
IBMCEPH-13232	Removing the cluster qos disable command from cqos
IBMCEPH-13276	Upgrading RHCS 8 cluster to RHCS 9 shows version as(rc-RelWithDebInfo)
IBMCEPH-13302	OSD deployment and startup fails on nodes with high OSD count
IBMCEPH-13353	Backport request for PR#66367 to 9.0 for GCHQ
IBMCEPH-13361	CEPHADM_STRAY_DAEMON health warning
IBMCEPH-13369	Bandwidth does not get distributed when qos bandwidth control is enabled at cluster and export level
IBMCEPH-13434	Proxy module improvements
IBMCEPH-13452	cephfs: Error listing encrypted case-insensitive directory contents
IBMCEPH-13453	Add support to update read and write bandwidth parameters in MB/sec
IBMCEPH-13460	Allow optional installation of service dependencies during host preparation
IBMCEPH-13493	Bandwidth control limit set at export level does not work as expected
IBMCEPH-13542	cephadm-ansible playbook fails with couldn't resolve module/action ceph_config
IBMCEPH-13568	Upgrading cluster from dashboard upgrades to the upstream versions
IBMCEPH-13618	HEAD returns the object but GET requests fail with code 404 for multi-part upload
IBMCEPH-13627	The new gRPC service is missing the corresponding reflections library
IBMCEPH-13692	ceph orch daemon add osd ignores db_devices passed via command for OSD deployment
IBMCEPH-13694	cephadm shell ceph-volume inventory errors out when Thin LVS are created
IBMCEPH-13726	Failed to instantiate ACL
IBMCEPH-13756	HTTP 409 error for Ceph RGW S3 requests

Issue key	Summary
IBMCEPH-13847	PerClient limiter is not working as expected in CQOS(ops and bandwidth control)in cloud build
IBMCEPH-13868	libcephfs proxy daemon crashes when configuring fscrypt key
IBMCEPH-13877	fails to start after upgrade to Tentacle due to inability to decode current period object(.3)from .rgw.root pool

Security fixes

This section lists security fixes from this release of Red Hat Ceph Storage 9.0z2.

For details about each CVE, see CVE Records.

- CVE-2023-48795
- CVE-2024-11831
- CVE-2025-59436
- CVE-2025-59437

Known issues

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

Ceph Object Gateway multi-site

Bucket index shows stale metadata after lifecycle expiration in versioned buckets

In rare cases, when lifecycle expiration removes objects from versioned buckets, some bucket index metadata might not be cleaned up correctly. As a result, stale index entries can remain even though the corresponding objects have already been removed.
If many stale entries accumulate, operations that depend on accurate bucket index listings can be affected. In some cases, tools that read the bucket index might report errors, such as `(27) File too large`.

This issue can impact administrative operations or background processes that rely on consistent bucket index metadata after lifecycle expiration runs on versioned buckets.

As a workaround, administrators can use available Object Gateway tooling to scan for and remove leftover bucket index entries.

(IBMCEPH-12980)

Release notes for 9.0z3

An update is now available for Red Hat Ceph Storage 9.0z3 which provides new features, enhancements, and bug fixes. There are also new known issues in this release.

New features and enhancements

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

`ceph-mgr`

Balancer logs are reduced while preserving key information

Previously, the balancer module generated repeated info-level log messages in quick cycles, which cluttered logs and made them harder to analyze.

With this enhancement, the balancer module summarizes repetitive log entries while retaining all important information. As a result, logs are more concise and easier to review without losing relevant details.

(IBMCEPH-12153)

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

ceph-exporter log rotation is correctly handled

Previously, the post logrotate template did not include ceph-exporter, so its logs were not written to a new file after rotation.

This fix adds ceph-exporter to the post logrotate signal list, ensuring that its logs are correctly written to the new log file after rotation.

(IBMCEPH-15536)

`cephadm` deployments no longer fail due to concurrent daemon-reload operations

Previously, `cephadm` could trigger multiple `systemctl daemon-reload` commands concurrently during parallel deployments, as threads attempted to operate on individual daemons' systemd units at the same time. This behavior could lead to intermittent deployment failures, where systemd temporarily reported that it could not find a daemon's unit.

With this fix, `cephadm` separates parallel deployment steps to prevent overlapping `systemctl daemon-reload` operations, ensuring more reliable handling of systemd units and reducing deployment failures.

(IBMCEPH-14297)

OSDs are now registered under the correct OSD specification

Previously, OSDs could be registered under an incorrect OSD specification after creation. As a result, commands such as `ceph orch ls` displayed OSDs under the wrong specification.

With this fix, a validation check ensures that the OSD specification service ID matches the OSD's affinity before registration, ensuring that OSDs are now correctly associated with their intended specification.

(IBMCEPH-13187)

`cephadm`-ansible playbook now runs successfully with corrected repository URL

Previously, the Red Hat repository URL in `cephadm`-ansible contained unintended whitespace, which caused incorrect parsing of the repository path and resulted in an invalid repository path. As a result, package and repository access operations failed.

With this fix, the extraneous whitespace has been removed from the repository URL in the configuration. The `cephadm`-ansible playbook now runs successfully without repository or path-related errors.

(IBMCEPH-14237)

Ceph Manager (Ceph Orchestrator)

`cephadm` export output now correctly formats multi-line strings

Previously, the YAML interpreter did not correctly process multi-line strings using the "|" symbol. As a result, when using the **`ceph orch ls --export`** command, multi-line string formatting was incorrect, preventing customers from reusing the exported output for new specifications without manual modification.

With this fix, support for multi-line string interpretation has been added to the YAML interpreter, ensuring that **`ceph orch ls --export`**now correctly formats multi-line strings using the "|" symbol.

(IBMCEPH-14046)

Ceph monitoring

PG imbalance alerts are now evaluated per device class

Previously, the alert expression calculated the average placement group (PG) count across all OSDs globally, without considering CRUSH device classes. As a result, OSDs from different device classes (for example, HDD and SSD) were compared against the same baseline, leading to inaccurate PG imbalance alerts.

With this fix, the alert expression now calculates the average PG count separately for each device class and evaluates each OSD against the average of its own class. As a result, PG imbalance alerts are more accurate, eliminating false positives caused by cross-class comparisons in heterogeneous storage environments.

(IBMCEPH-13281)

Ceph Dashboard

Improved error message when creating a local storage class with an existing tier-type name

Previously, when attempting to create a local storage class using a name already assigned to a tier-type storage class, such as s3 or s3-glacier, the system returned the generic `404 - Not Found` error message.

With this fix, the error handling has been improved to provide a clearer and more informative message when a storage class name conflict occurs, helping customers quickly identify and resolve the issue.

(IBMCEPH-13086)

Ceph File System (CephFS)

Updated subvolume removal workflow to prevent inconsistent states

Previously, removing a subvolume in a full-cluster condition could leave the subvolume in an invalid state.

With this fix, the subvolume removal workflow has been updated so that metadata is now updated before the UUID directory is moved to the .trash directory. This change ensures that any `ENOSPC` error is detected during the metadata update, allowing the operation to fail safely and preventing inconsistent state.

As a result, the system no longer leaves subvolumes in a partially removed or invalid state, and subsequent subvolume operations complete successfully.

(IBMCEPH-15168)

MDS no longer crashes due to improper `mds_lock` usage

Previously, due to a bug in the MDS, certain code paths could run without the required `mds_lock` being held. This unsafe behavior caused the MDS to crash unpredictably.

With this fix, the `mds_lock` is always held whenever it is required. As a result, the MDS now operates as expected without experiencing arbitrary crashes.

(IBMCEPH-14355)

MDS service stability during journal replay improved

Previously, the MDS service became unavailable during the journal replay phase, interrupting file system operations until the service was manually restarted.

With this fix, the MDS service completes journal replay successfully and remains available.

(IBMCEPH-14090)

Subvolume snapshot visibility flag now correctly reported after cluster upgrade

Previously, after a cluster upgrade, the subvolume snapshot visibility flag (**`SNAPDIR_VISIBILITY`**) was incorrectly reported as `0` (visibility disabled) instead of `1`. Although this did not affect access by default, enabling `client_respect_subvolume_snapshot_visibility` could cause snapshots to appear inaccessible.

With this fix, the snapshot visibility flag is now correctly set and reported after upgrades, ensuring consistent and expected behavior when snapshot visibility controls are in use.

(IBMCEPH-12365)

Subvolume deletion supports deep directory hierarchies

Previously, deleting a subvolume with a deep file hierarchy could result in a maximum recursion limit error due to recursive processing in the underlying code.

This fix replaces the recursive logic with an iterative approach to handle deep directory structures more reliably. With this update, subvolumes with deep hierarchies can be deleted successfully without encountering recursion limit errors.

(IBMCEPH-11604)

CephFS client no longer crashes during cache trimming

Previously, a missing reference pin on directory dentries opened with `O_DIRECTORY` allowed the client metadata cache to treat them as evictable while still in use. As a result, the CephFS client could crash during cache trimming under heavy workloads or during unmount, leaving the cluster in a health warning state.

With this fix, the missing dentry pin has been added for opened directories, along with a safety guard in the cache eviction path to prevent use-after-free scenarios. As a result, the CephFS client no longer crashes during cache trimming, and missing pins are now detected by assertion instead of causing silent memory corruption.

(IBMCEPH-14336)

Ceph Object Gateway

Lifecycle delete markers are now correctly identified and expired

Previously, a required key commit to support lifecycle delete-marker expiration was not included. As a result, the lifecycle workflow did not correctly identify delete markers and skipped removing them.

With this fix, a new mechanism has been introduced to report delete markers using an object handle or its state. This mechanism is also used in GET and HEAD operations to correctly report objects as delete markers, including for tools such as `awscli`.

As a result, lifecycle delete-marker expiration now works correctly, including for objects created in versioning-suspended buckets, and commands such as `awscli list-object-versions` correctly identify delete markers.

(IBMCEPH-13829)

S3 GET requests now correctly authorized for versioned bucket objects

Previously, the version ID was populated during a simple GET request for an object in a versioned bucket. As a result, the `s3:GetObject` action in IAM policies did not allow the request, even when it was explicitly permitted.

With this fix, the version ID is no longer populated during simple GET requests for objects in versioned buckets, ensuring that the `s3:GetObject` action correctly authorizes the request.

(IBMCEPH-13754)

Ceph Object Gateway no longer crashes when admin assumes a non-existent role

Previously, Ceph Object Gateway could crash when an admin user sent an `AssumeRole` request with an invalid or non-existent `RoleArn`. This occurred because the request was not properly validated before processing.

With this fix, invalid or non-existent roles are handled gracefully, preventing crashes and ensuring stable Ceph Object Gateway behavior when processing `AssumeRole` requests.

(IBMCEPH-13532)

SNS topics with invalid names can now be modified after upgrade

Previously, strict SNS topic name verification was introduced to ensure specification compliance. As a result, topics created before this change that contained invalid characters could not be modified after an upgrade, because name validation failed.

With this fix, a configuration option has been added to relax topic name verification, allowing existing topics with invalid characters to be modified successfully.

(IBMCEPH-14209)

Kafka notifications are no longer rejected due to oversized batches after recovery

Previously, after a Kafka cluster recovered from an outage, a large number of notifications were batched together and resent to the broker for performance reasons. As a result, if the broker was configured with maximum message size or batch size limits, these large batched messages could be rejected.

With this fix, a configuration option has been added to control the maximum batch size on Ceph Object Gateway. When sending a large number of notifications, such as after Kafka recovery, messages are split into smaller batches based on this setting, preventing rejections.

(IBMCEPH-14055)

Bucket listing now completes when encountering filtered entries

Previously, bucket listing operations could fail to make progress when processing sequences of filtered-out entries. As a result, bucket listings could be incomplete.

With this fix, the listing logic has been improved to ensure continued progress through filtered entries, allowing bucket listings to complete successfully.

(IBMCEPH-14042)

Ceph Object Gateway multi-site

Data sync failures during full sync are no longer reported as successful

Previously, when a full data sync was in progress and failures such as `-EBUSY` or `-EIO` occurred (for example, during a primary Ceph Object Gateway restart), the sync operation could silently fail while the status incorrectly indicated that it was caught up.

This fix ensures that such failures are properly detected and reported during the sync process instead of being ignored. With this update, sync status accurately reflects failures, preventing false “caught up” states.

(IBMCEPH-12583)

RADOS

`ceph-mon` shutdown no longer causes iterator-related crashes

Previously, during `ceph-mon` shutdown, the monitor store was closed before Monitor destruction. This caused RocksDB iterators held by `SyncProvider` to be destroyed after the underlying database was already closed, leading to instability.

This fix ensures that Monitor objects are properly destroyed before the monitor store is closed, preventing invalid access to RocksDB iterators. With this update, shutdown completes cleanly without iterator-related errors or crashes.

(IBMCEPH-13740)

All bug fixes

This section lists a complete listing of all bug fixes in this release of Red Hat Ceph Storage 9.0z3.

Table 3: List of all bug fixes

Issue key	Severity	Summary
IBMCEPH-14090	Critical	Ceph MDS crashes during journal replay
IBMCEPH-13281	Important	CephPGImbalance alerts incorrectly on heterogeneous clusters
IBMCEPH-13285	Important	cephadm ansible generates incorrect repo for Ceph 9
IBMCEPH-13532	Important	RGW crash on admin AssumeRole with invalid RoleArn
IBMCEPH-13739	Important	ceph mon shutdown no longer causes iterator related crash
IBMCEPH-13864	Important	Edit client config option remains stuck in loading state
IBMCEPH-14055	Important	Kafka notifications produce oversized batch error after recovery
IBMCEPH-14135	Important	Dashboard missing MDS observability metrics
IBMCEPH-14140	Important	Dashboard reports incorrect CephFS capacity usage

Issue key	Severity	Summary
IBMCEPH-14207	Important	Dashboard pool manager role cannot create pools due to access denied
IBMCEPH-14209	Important	Unable to re create topic with invalid characters
IBMCEPH-11604	Moderate	Switch to non recursive implementation for directory tree removal
IBMCEPH-12365	Moderate	Snapshot visibility subvolume option is incorrect after upgrade
IBMCEPH-12583	Moderate	Full sync of objects with special character fails to sync to archive site
IBMCEPH-12908	Moderate	Unable to see old metrics using loki query after upgrade
IBMCEPH-13187	Moderate	Ceph orch ls misreports count for OSD services
IBMCEPH-13555	Moderate	GUI OSD delete popup lacks preserve OSD ID option
IBMCEPH-13750	Moderate	Dashboard empty state illustration missing image asset
IBMCEPH-13829	Moderate	Lifecycle error on expiration of delete marker with null versionId
IBMCEPH-14047	Moderate	CephFS metrics MDS crash during update_rank0
IBMCEPH-14051	Moderate	Root squash not working as expected with multifs caps
IBMCEPH-14171	Moderate	Apply zstd upgrade to 9.0
IBMCEPH-14188	Moderate	Unable to restart stop RGW service through dashboard
IBMCEPH-14269	Moderate	Service apply fails after upgrade due to validation
IBMCEPH-14297	Moderate	OSD upgrade fails trimming old cgroups
IBMCEPH-14336	Moderate	Ceph MGR crashing with segmentation fault
IBMCEPH-15168	Moderate	Clone operations loop and fail continuously
IBMCEPH-15193	Moderate	Documentation URL incorrect in dashboard
IBMCEPH-13740	Low	MON crash during shutdown fixed
IBMCEPH-13754	Low	403 response for HeadObject with GetObject permission
IBMCEPH-13811	Low	GUI OSD delete popup lacks preserve OSD ID option
IBMCEPH-14042	Low	Treat EFBIG as advance and retry in unordered listing
IBMCEPH-14046	Low	Orch export ingress service ssl_cert formatting issue

Issue key	Severity	Summary
IBMCEPH-14149	Low	Orch export ingress service prevents apply
IBMCEPH-14237	Low	Repo URL whitespace bug causes DNF failure
IBMCEPH-14291	Low	Orchestrator cannot redeploy DB devices
IBMCEPH-14329	Low	Add APIs to set and clear MOTD
IBMCEPH-14332	Low	Add API and CLI to manage remote_write
IBMCEPH-14355	Low	MDS segfault with log trim
IBMCEPH-15017	Low	Create storage class UI page broken
IBMCEPH-15536	Low	logrotate does not signal ceph exporter

Security fixes

This section lists security fixes from this release of Red Hat Ceph Storage 9.0z3.

For details about each CVE, see CVE Records.

- CVE-2019-10790
- CVE-2023-25153
- CVE-2024-55565
- CVE-2025-47913
- CVE-2025-47914
- CVE-2025-58181
- CVE-2025-59343
- CVE-2025-64718
- CVE-2025-64756

Published date: 2026-06-