



Red Hat Ceph Storage 7.1

リリースノート

Red Hat Ceph Storage 7.1 のリリースノート

Red Hat Ceph Storage 7.1 リリースノート

Red Hat Ceph Storage 7.1 のリリースノート

法律上の通知

Copyright © 2024 Red Hat, Inc.

The text of and illustrations in this document are licensed by Red Hat under a Creative Commons Attribution–Share Alike 3.0 Unported license ("CC-BY-SA"). An explanation of CC-BY-SA is available at

<http://creativecommons.org/licenses/by-sa/3.0/>

. In accordance with CC-BY-SA, if you distribute this document or an adaptation of it, you must provide the URL for the original version.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, the Red Hat logo, JBoss, OpenShift, Fedora, the Infinity logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux[®] is the registered trademark of Linus Torvalds in the United States and other countries.

Java[®] is a registered trademark of Oracle and/or its affiliates.

XFS[®] is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL[®] is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Node.js[®] is an official trademark of Joyent. Red Hat is not formally related to or endorsed by the official Joyent Node.js open source or commercial project.

The OpenStack[®] Word Mark and OpenStack logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation, or the OpenStack community.

All other trademarks are the property of their respective owners.

概要

本リリースノートでは、Red Hat Ceph Storage 7.1 製品リリース向けに実装された主要な機能、機能拡張、既知の問題、バグ修正を説明します。

目次

多様性を受け入れるオープンソースの強化	3
RED HAT CEPH STORAGE ドキュメントへのフィードバック (英語のみ)	4
第1章 概要	5
第2章 謝辞	6
第3章 新機能	7
3.1. CEPHADM ユーティリティー	7
3.2. CEPH DASHBOARD	10
3.3. CEPH ファイルシステム	12
3.4. CEPH OBJECT GATEWAY	13
3.5. マルチサイトの CEPH OBJECT GATEWAY	15
3.6. RADOS	15
3.7. RADOS ブロックデバイス (RBD)	16
第4章 バグ修正	17
4.1. CEPHADM ユーティリティー	17
4.2. CEPH ANSIBLE ユーティリティー	17
4.3. NFS GANESHA	17
4.4. CEPH DASHBOARD	17
4.5. CEPH ファイルシステム	19
4.6. CEPH OBJECT GATEWAY	21
4.7. マルチサイトの CEPH OBJECT GATEWAY	27
4.8. RADOS	27
4.9. RBD ミラーリング	30
第5章 既知の問題	31
5.1. CEPHADM ユーティリティー	31
5.2. CEPH OBJECT GATEWAY	31
第6章 ソース	33

多様性を受け入れるオープンソースの強化

Red Hat では、コード、ドキュメント、Web プロパティにおける配慮に欠ける用語の置き換えに取り組んでいます。まずは、マスター (master)、スレーブ (slave)、ブラックリスト (blacklist)、ホワイトリスト (whitelist) の 4 つの用語の置き換えから始めます。この取り組みは膨大な作業を要するため、用語の置き換えは、今後の複数のリリースにわたって段階的に実施されます。詳細は、[Red Hat CTO である Chris Wright のメッセージ](#) をご覧ください。

RED HAT CEPH STORAGE ドキュメントへのフィードバック (英語のみ)

Red Hat ドキュメントに対するご意見をお聞かせください。ドキュメントの改善点があればお知らせください。これを行うには、Bugzilla のチケットを作成します。

1. [Bugzilla](#) の Web サイトに移動します。
2. Component ドロップダウンメニューで、**Documentation** を選択します。
3. Sub-Component ドロップダウンで、適切なサブコンポーネントを選択します。
4. ドキュメントの適切なバージョンを選択します。
5. **Summary** および **Description** フィールドに、ドキュメントの改善に関するご意見を記入してください。ドキュメントの該当部分へのリンクも追加してください。
6. オプション: 添付ファイルを追加します (ある場合)。
7. **Submit Bug** をクリックします。

第1章 概要

Red Hat Ceph Storage は、非常にスケーラブルでオープンなソフトウェア定義のストレージプラットフォームであり、最も安定したバージョンの Ceph ストレージシステムと Ceph 管理プラットフォーム、デプロイメントユーティリティー、およびサポートサービスを組み合わせたものです。

Red Hat Ceph Storage ドキュメントは、https://access.redhat.com/documentation/ja-jp/red_hat_ceph_storage/7 から入手できます。

第2章 謝辞

Red Hat Ceph Storage バージョン 7.1 には、Red Hat Ceph Storage チームの数多くの貢献が反映されています。さらに Ceph プロジェクトでは、Ceph コミュニティーの個人や組織からの貢献の度合いが質と量の両面で大幅に拡大しています。Red Hat Ceph Storage チームの全メンバー、Ceph コミュニティーの個々の貢献者、および以下の組織を含むすべての方々の貢献に謝意を表します。

- Intel[®]
- Fujitsu[®]
- UnitedStack
- Yahoo[™]
- Ubuntu Kylin
- Mellanox[®]
- CERN[™]
- Deutsche Telekom
- Mirantis[®]
- SanDisk[™]
- SUSE[®]

第3章 新機能

本セクションでは、Red Hat Ceph Storage の今回のリリースで導入された主要な更新、機能拡張、新機能のリストを紹介します。

3.1. CEPHADM ユーティリティー

ユーザーは `idmap.conf` でさまざまな NFS オプションを設定可能に

この機能拡張により、`idmap.conf` で "Domain"、"Nobody-User"、"Nobody-Group" などの NFS オプションを設定できるようになりました。

[Bugzilla:2068026](#)

NFS の新しい haproxy プロトコルモードでクライアント IP 制限が可能に

以前は、NFS 経由の haproxy を使用するセットアップではクライアント IP 制限は機能しませんでした。

この機能拡張により、Cephadm でデプロイされた NFS は haproxy プロトコルをサポートするようになりました。ユーザーが Ingress と haproxy の両方の仕様に `enable_haproxy_protocol: True` を追加するか、`ceph nfs cluster create` コマンドに `--ingress-mode haproxy-protocol` を渡すと、NFS デーモンは haproxy プロトコルを使用します。

[Bugzilla:2068030](#)

ユーザーは Grafana API URL にアクセスするためにユーザー名とパスワードを入力する必要があります。

以前は、Grafana API URL に接続できる人なら誰でも、認証情報を必要とせずにアクセスできました。

この機能拡張により、Cephadm でデプロイされた Grafana には、ユーザーが Grafana API URL にアクセスするためのユーザー名とパスワードが設定されます。

[Bugzilla:2079815](#)

NFS バックエンドを備えた Ingress サービスは、HAProxy レイヤーを介さずに、keepalived のみを使用して NFS デーモンがバインドするための仮想 IP (VIP) を作成するようにセットアップできるようになりました。

この機能拡張により、NFS バックエンドを備えた Ingress サービスは、HAProxy レイヤーを介さずに、`keepalived` のみを使用して NFS デーモンがバインドする仮想 IP を作成するようにセットアップできます。これは、NFS デーモンが移動され、そのデーモンに接続するためにクライアントが別の IP を使用する必要がない場合に役立ちます。

Cephadm は、`keepalived` をデプロイして仮想 IP を設定し、NFS デーモンをその仮想 IP にバインドします。これは、`--ingress --ingress-mode keepalive-only --virtual-ip <VIP>` のフラグを使用し、`ceph nfs cluster create` コマンドを介して NFS モジュールを使用してセットアップすることもできます。

仕様ファイルは次のようになります。

```
service_type: ingress
service_id: nfs.nfsganasha
service_name: ingress.nfs.nfsganasha
placement:
  count: 1
  label: foo
```

```
spec:
  backend_service: nfs.nfsganesha
  frontend_port: 12049
  monitor_port: 9049
  virtual_ip: 10.8.128.234/24
  virtual_interface_networks: 10.8.128.0/24
  keepalive_only: true
```

これには **keepalive_ony: true** 設定が含まれます。

NFS 仕様は次のようになります。

```
networks:
  - 10.8.128.0/21
service_type: nfs
service_id: nfsganesha
placement:
  count: 1
  label: foo
spec:
  virtual_ip: 10.8.128.234
  port: 2049
```

これには、Ingress 仕様の仮想 IP と一致する **virtual_ip** フィールドが含まれます。

[Bugzilla:2089167](#)

HAProxy デーモンバインドは、付随する keepalived によって作成された仮想 IP 上のフロントエンドポートにのみ接続します。

この機能拡張により、HAProxy デーモンは、0.0.0.0 ではなく、付随する keepalived によって作成された仮想 IP 上でのみフロントエンドポートにバインドするようになります。Cephadm でデプロイされた HAProxy はフロントエンドポートを仮想 IP にバインドし、NFS デーモンなどの他のサービスが同じノード上の他の IP のポート 2049 にバインドできるようになります。

[Bugzilla:2176297](#)

Ingress サービスの Haproxy ヘルスチェック間隔がカスタマイズ可能になりました。

以前は、デフォルトの 2 秒のヘルスチェックの頻度が多すぎたため、不要なトラフィックが発生する場合があります。

この機能拡張により、Ingress サービスの HAProxy ヘルスチェック間隔をカスタマイズできるようになります。**health_check_interval** フィールドを含む Ingress 仕様を適用すると、サービスの各 HAProxy デーモンに対して Cephadm によって生成される HAProxy 設定に、ヘルスチェック間隔の値が含まれるようになります。

Ingress 仕様ファイル:

```
service_type: ingress
service_id: rgw.my-rgw
placement:
  hosts: ['ceph-mobisht-7-1-07lum9-node2', 'ceph-mobisht-7-1-07lum9-node3']
spec:
  backend_service: rgw.my-rgw
  virtual_ip: 10.0.208.0/22
```

```
frontend_port: 8000
monitor_port: 1967
health_check_interval: 3m
```

間隔の有効な単位は、**us**: マイクロ秒、**ms**: ミリ秒、**s**: 秒、**m**: 分、**h**: 時間 **d**: 日です。

[Bugzilla:2199129](#)

Grafana は、常に 0.0.0.0 にバインドするのではなく、ホスト上の特定のネットワーク内の IP にバインドするようになりました。

この機能拡張により、Grafana が IP にバインドするネットワークを含む "networks" セクションと、仕様の "spec" セクションに含まれる **only_bind_port_on_networks: true** の両方を含む Grafana 仕様ファイルを使用して、Cephadm は Grafana デーモンを 0.0.0.0 ではなくそのネットワーク内の IP にバインドするように設定します。これにより、ユーザーは、Grafana が別のサービスに使用するのと同じポートを、ホスト上の別の IP で使用できるようになります。すべてが移動されるわけではない仕様の更新の場合は、**ceph orch redeploy grafana** を実行して、設定の変更を取得できます。

Grafana 仕様ファイル:

```
service_type: grafana
service_name: grafana
placement:
  count: 1
networks:
  192.168.122.0/24
spec:
  anonymous_access: true
  protocol: https
  only_bind_port_on_networks: true
```

[Bugzilla:2233659](#)

すべてのブートストラップ CLI パラメーターが cephadm-ansible モジュールで使用できるようになりました。

以前は、ブートストラップ CLI パラメーターのサブセットのみが使用可能であり、モジュールの使用が制限されていました。

この機能拡張により、すべてのブートストラップ CLI パラメーターが **cephadm-ansible** モジュールで使用できるようになります。

[Bugzilla:2246266](#)

Prometheus スクレイプ設定が nfs-ganesha エクスポーターに追加されます。

この機能拡張により、Prometheus スクレイプ設定が nfs-ganesha エクスポーターに追加されます。これは、nfs-ganesha prometheus エクスポーターによって公開されたメトリクスを Ceph で実行されている Prometheus インスタンスにスクレイピングするためのもので、Grafana ダッシュボードによってさらに消費されます。

[Bugzilla:2263898](#)

Prometheus は、常に 0.0.0.0 にバインドするのではなく、ホスト上の特定のネットワーク内の IP にバインドするようになりました。

この機能拡張により、Prometheus が IP にバインドするネットワークを含む "networks" セクション

と、仕様の "spec" セクションに含まれる **only_bind_port_on_networks: true** の両方を含む Prometheus 仕様ファイルを使用して、Cephadm は Prometheus デーモンを 0.0.0.0 ではなくそのネットワーク内の IP にバインドするように設定します。これにより、ユーザーは、Prometheus が別のサービスに使用するのと同じポートを、ホスト上の別の IP で使用できるようになります。すべてが移動されるわけではない仕様の更新の場合は、**ceph orch redeploy prometheus** を実行して、設定の変更を取得できます。

Prometheus 仕様ファイル:

```
service_type: prometheus
service_name: prometheus
placement:
  count: 1
networks:
- 10.0.208.0/22
spec:
  only_bind_port_on_networks: true
```

[Bugzilla:2264812](#)

ユーザーはスナップショット (.snap ディレクトリー内のエクスポート) をマウントできるようになりました。

この機能拡張により、ユーザーはスナップショット (.snap ディレクトリー内のエクスポート) をマウントして RO モードで表示できるようになります。NFS MGR モジュールで作成された NFS エクスポートには、スナップショットをマウントできる **cmount_path** 設定 (これは設定できず、"/" のままにしておく必要があります) が含まれるようになりました。

[Bugzilla:2245261](#)

ceph rgw realm bootstrap... コマンドで提供される仕様ファイルを使用してゾーングループホスト名を設定できるようになりました

このリリースでは、Ceph Object Gateway マルチサイトセットアップの自動化を継続すると、ユーザーはブートストラップコマンド **ceph rgw realm bootstrap...** で渡された最初の仕様ファイルを使用してゾーングループのホスト名を設定できるようになりました。追加の手順は必要ありません。

以下に例を示します。

```
zonegroup_hostnames:
- host1
- host2
```

ユーザーが realm bootstrap コマンドで渡された Ceph Object Gateway 仕様ファイルの specification セクションに上記のセクションを追加すると、Cephadm は Ceph Object ゲートウェイモジュールがレルム/ゾーングループ/ゾーンの作成を完了した後に、仕様で定義されたゾーングループにこれらのホスト名を自動的に追加します。Cephadm モジュールが現在完了している他のアクティビティーに応じて、これが発生するまでに数分かかる場合があります。

3.2. CEPH DASHBOARD

Ceph Dashboard での CephFS スナップショットスケジュール管理

以前は、CephFS スナップショットスケジュールは、コマンドラインインターフェイスを使用しないと管理できませんでした。

この機能拡張により、CephFS スナップショットスケジュールを Ceph Dashboard からリスト表示、作成、編集、有効化、無効化、削除できるようになりました。

[Bugzilla:2264145](#)

Ceph Dashboard で NFSv3 ベースのエクスポートをサポートできるようになりました。

この機能拡張により、Ceph Dashboard で NFSv3 ベースのエクスポート管理のサポートが有効になります。

[Bugzilla:2267763](#)

CephFS の Ceph ユーザーを管理する機能が追加されました。

この機能拡張により、CephFS の Ceph ユーザーを管理する機能が追加されました。これにより、ファイルシステムビューからボリューム、サブボリュームグループ、サブボリュームに対するユーザーの権限を管理できるようになります。

[Bugzilla:2271110](#)

マルチサイト同期ステータス用の新しい API エンドポイントが追加されました。

以前は、マルチサイト同期ステータスは CLI コマンド経由でのみ利用可能でした。

この機能拡張により、Ceph Dashboard の API を介してマルチサイトステータスが追加されます。マルチサイト同期ステータスの新しい API エンドポイントは `api/rgw/multisite/sync_status` です。

[Bugzilla:2258951](#)

NVMe-oF ゲートウェイの監視が改善されました。

この機能拡張により、NVMe-oF ゲートウェイの監視を改善するために、発行されたメトリクスに基づいて NVMe-oF ゲートウェイのアラートが追加され、また、組み込みの Prometheus エクスポーターからのメトリクスが NVMe-oF ゲートウェイでスクレイピングされます。

[Bugzilla:2276038](#)

Ceph ダッシュボードでの CephFS クローン管理

この機能拡張により、Ceph ダッシュボードで CephFS クローン管理機能が提供されるようになりました。ユーザーは Ceph ダッシュボードを通じてサブボリュームのクローンを作成および削除できます。

[Bugzilla:2264142](#)

Ceph ダッシュボードでの CephFS スナップショット管理

この機能拡張により、Ceph ダッシュボードに CephFS スナップショット管理機能が提供されるようになりました。ユーザーは Ceph ダッシュボードを通じて、サブボリュームスナップショットを作成および削除できます。

[Bugzilla:2264141](#)

ユーザー/バケットごとのラベル付きパフォーマンスカウンター

この機能拡張により、ユーザーは Ceph Object Gateway ノードごとに実行されている操作に関する情報を取得できるだけでなく、Ceph ダッシュボードでユーザーごとおよびバケットごとの Ceph Object Gateway パフォーマンスカウンターを表示することもできます。

ラベル付き同期パフォーマンスカウンターの Prometheus への導入

この機能拡張により、ユーザーはゾーン間のレプリケーションの健全性に関する Prometheus からのリアルタイムの情報を収集し、Ceph Object Gateway のマルチサイト同期操作の監視性を高めることができます。

Ceph ダッシュボードでのバケットの追加および編集

この機能拡張により、Ceph ダッシュボードに対する Ceph Object Gateway の改善の一環として、Ceph ダッシュボードからバケットを適用、リスト表示、編集する機能が追加されました。

- ACL (パブリック、プライベート)
- タグ (追加/削除)

Ceph ダッシュボードでのバケットポリシーの追加、リスト表示、削除、および適用

この機能拡張により、Ceph ダッシュボードに対する Ceph Object Gateway の改善の一環として、Ceph ダッシュボードからバケットポリシーを追加、リスト表示、削除、適用する機能が追加されました。

3.3. CEPH ファイルシステム

MDS 動的メタデータバランサーはデフォルトでオフになっています。

以前は、バランサーの動作が不十分なために、`max_mds` ファイルシステム設定を増やすと、ツリーが望ましくない方法で断片化されていました。

この機能拡張により、MDS 動的メタデータバランサーはデフォルトでオフになります。Operator はバランサーを使用するために明示的にオンにする必要があります。

[Bugzilla:2227309](#)

CephFS はサブボリュームまたはディレクトリーツリーの静止をサポートします。

以前は、クライアント間で帯域外通信が存在する一貫したスナップショットバリアを越えて、複数のクライアントが読み取りと書き込みをインターリーブしていました。この通信が原因で、クライアントはスナップショットを介して相互に回復可能なチェックポイントに到達したと誤って認識するようになりました。

この機能拡張により、CephFS はサブボリュームまたはディレクトリーツリーの静止をサポートし、クラッシュ整合性のあるスナップショットの実行が可能になります。MDS がスナップショットを実行する前に、クライアントはすべての I/O を強制的に停止するようになりました。これにより、サブツリーのすべてのクライアントにチェックポイントが適用されます。

[Bugzilla:2235753](#)

MDS 常駐セグメントサイズ (RSS) パフォーマンスカウンターは、より高い優先度で追跡されます。

この機能拡張により、MDS 常駐セグメントサイズパフォーマンスカウンターがより高い優先度で追跡され、呼び出し元がその値を使用して有用な警告を生成できるようになります。これにより、Rook は MDS RSS サイズを識別し、それに応じて動作できるようになります。

[Bugzilla:2256560](#)

遅延したクライアントは、遅延した OSD がない場合にのみエビクトされるようになりました

以前は、MDS からのパフォーマンスダンプを監視すると、OSD が遅れていること (**objecter.op_laggy** および **objecter.osd_laggy**) が示されることがあり、クライアントの遅延の原因となっていました (キャップ取り消しのためにダーティーデータをフラッシュできませんでした)。

この機能拡張により、**defer_client_eviction_on_laggy_osds** オプションが true に設定され、OSD の遅延が原因でクライアントの遅延が発生した場合、OSD の遅延がなくなるまでクライアントエビクションは行われません。

[Bugzilla:2260003](#)

cephfs-mirror デーモンは、**perf dump** コマンドを介してスナップショット同期パフォーマンスカウンターをエクスポートします。

ceph-mds デーモンは、既存の **perf dump** コマンドに含まれるクライアントごとのパフォーマンスカウンターをエクスポートします。

[Bugzilla:2264177](#)

ディレクトリー情報をダンプするための新しい **dump dir** コマンドが導入されました。

この機能拡張により、ディレクトリー情報をダンプして出力する **dump dir** コマンドが導入されました。

[Bugzilla:2269687](#)

サブボリュームのスナップショットスケジューリングサポート

この機能拡張により、サブボリュームのスナップショットスケジューリングサポートが提供されます。すべてのスナップショットスケジューリングコマンドは、適切なサブボリュームとサブボリュームグループを参照するための **--subvol** および **--group** 引数を受け入れます。サブボリュームグループ引数なしでサブボリュームが指定された場合は、デフォルトのサブボリュームグループが考慮されます。また、サブボリュームを参照するときには有効なパスを指定する必要はなく、使用される引数解析の性質上、プレースホルダー文字列だけで十分となります。

例

```
# ceph fs snap-schedule add - 15m --subvol sv1 --group g1
# ceph fs snap-schedule status - --subvol sv1 --group g1
```

[Bugzilla:2238537](#)

MDS キャップを追加または変更する Ceph コマンドは、ユーザーが渡した MDS キャップが拒否された理由を説明する

以前は、MDS キャップを追加または変更する Ceph コマンドは、"Error EINVAL: mds capability parse failed, stopped at 'allow w' of 'allow w'" と出力していました。

この機能拡張により、コマンドは、ユーザーによって渡された MDS キャップが拒否された理由を説明し、"Error EINVAL: Permission flags in MDS caps must start with 'r' or 'rw' or be '*' or 'all'" と出力します。

[Bugzilla:2247586](#)

3.4. CEPH OBJECT GATEWAY

バケット通知を管理するための管理インターフェイスが追加されました。

以前は、バケット通知の管理に S3 REST API が使用されていました。ただし、管理者がその設定をオーバーライドする場合に、radosgw-admin ツールを使用して簡単にオーバーライドする方法はありませんでした。

この機能拡張により、バケット通知を管理するための次のコマンドを備えた管理インターフェイスが追加されました。

```
radosgw-admin notification get --bucket <bucket name> --notification-id <notification id>

radosgw-admin notification list --bucket <bucket name>

radosgw-admin notification rm --bucket <bucket name> [--notification-id <notification id>]
```

[Bugzilla:2130292](#)

RGW ラベルの付いたユーザーおよびバケット操作カウンターは、ceph counter dump の実行時に別のセクションに表示されるようになりました。

以前は、RGW ラベルの付いたすべての操作カウンターは **ceph counter dump** コマンドの出力の **rgw_op** セクションにありました。ただし、ユーザーラベルまたはバケットラベルのいずれかが付いていました。

この機能拡張により、**ceph counter dump** コマンドが実行されると、RGW ラベルの付いたユーザーおよびバケット操作カウンターがそれぞれ **rgw_op_per_user** セクションまたは **rgw_op_per_bucket** セクションに表示されるようになります。

[Bugzilla:2265574](#)

ユーザーは、-t コマンドラインオプションを使用して一時ファイルをディレクトリーに配置できるようになりました。

以前は、**/usr/bin/rgw-restore-bucket-index** ツールは **/tmp** のみを使用していたため、そのディレクトリーにはすべての一時ファイルを保持するのに十分な空き領域がない場合があります。

この機能拡張により、ユーザーは **-t** コマンドラインオプションを使用して一時ファイルを配置できるディレクトリーを指定でき、スペースが不足した場合に通知されるため、ツールを再実行するためのどのような調整を行う必要があるかがわかります。また、ユーザーは、ツールの一時ファイルによって、一時ファイルが存在するファイルシステム上の使用可能な領域が使い果たされているかどうかを定期的に確認することもできます。

[Bugzilla:2267715](#)

copy-object API を使用した暗号化オブジェクトのコピーがサポートされるようになる

以前は、Ceph Object Gateway では、サーバー側の暗号化サポートが開始されて以来、copy-object API を使用した暗号化オブジェクトのコピーはサポートされていませんでした。

この機能拡張により、copy-object API を使用した暗号化されたオブジェクトのコピーがサポートされ、copy-object 操作に依存するワークロードでもサーバー側の暗号化を使用できるようになります。

[Bugzilla:2149450](#)

新しい Ceph Object Gateway の admin-ops 機能が追加され、ユーザーメタデータの読み取りが許可され、関連付けられた承認キーの読み取りは許可されない

この機能拡張により、新しい Ceph Object Gateway admin-ops 機能が追加され、Ceph Object Gateway のユーザーメタデータの読み取りは許可されますが、関連付けられている承認キーの読み取りは許可されなくなります。これは、自動化およびレポートツールの権限を減らし、ユーザーになります。

したり、ユーザーのキーを表示したりすることを回避するためです。

[Bugzilla:2112325](#)

クラウド移行: 新たにサポートされる S3 互換プラットフォームを追加

このリリースでは、オブジェクトストレージをクラウドまたは他のオンプレミスの S3 エンドポイントに移動できるように、現在のライフサイクルの移行とストレージクラスモデルが拡張されています。クラウドアーカイブ機能では、IBM Cloud Object Store (COS) や IBM Storage Ceph などの S3 互換プラットフォームがサポートされるようになりました。

RGW バックエンドを使用した NFS

このリリースでは、Ceph Object Gateway バックエンドを備えた NFS が既存の機能とともに再 GA 化されます。

3.5. マルチサイトの CEPH OBJECT GATEWAY

radosgw-admin sync status コマンドに再試行メカニズムが導入されました。

以前は、マルチサイト同期がリモートゾーンに要求を送信するときに、ラウンドロビンストラテジーを使用してゾーンエンドポイントの1つを選択しました。そのエンドポイントが利用できない場合、**radosgw-admin sync status** コマンドで使用される http クライアントロジックは再試行メカニズムを提供しないため、入出力エラーが報告されます。

この機能拡張により、同期ステータスコマンドに再試行メカニズムが導入され、選択したエンドポイントが利用できない場合に、要求に対応するために別のエンドポイントが選択されます。

[Bugzilla:1995152](#)

NewerNoncurrentVersions、**ObjectSizeGreaterThan**、**ObjectSizeLessThan** フィルターがライフサイクルに追加されました。

この機能拡張により、ライフサイクルに **NewerNoncurrentVersions**、**ObjectSizeGreaterThan**、および **ObjectSizeLessThan** フィルターのサポートが追加されます。

[Bugzilla:2172162](#)

ユーザー S3 レプリケーション API がサポート対象になる

この機能拡張により、ユーザーの S3 レプリケーション API がサポートされるようになりました。これらの API を使用すると、ユーザーはバケットレベルでレプリケーションポリシーを設定できます。API は拡張され、送信元ゾーン名と宛先ゾーン名を指定するための追加パラメーターが含まれるようになりました。

[Bugzilla:2279461](#)

バケットの詳細な同期レプリケーション GA (パート 3)

このリリースでは、バケットの詳細なサポートとともに、バケットまたはバケットのグループを別の Red Hat Ceph Storage クラスターにレプリケートする機能が追加されました。ユーザービリティ要件は、Ceph Object Gateway マルチサイトと同じです。

3.6. RADOS

noautoscale フラグをオン/オフに設定すると、各プールの元の自動スケールモード設定が保持される

以前は、**noautoscale** フラグが設定されている場合、**pg_autoscaler** は各プールの **autoscale mode** 設定に保持されませんでした。このため、**noautoscale** フラグが設定されるたびに、各プールに対して **autoscale** モードを繰り返し設定する必要がありました。

この機能拡張により、**pg_autoscaler** モジュールは、**noautoscale flag** が設定された後も、**autoscaler** モードの個別のプール設定を保持します。**noautoscale** フラグをオン/オフに設定しても、各プールの元の自動スケールモード設定は保持されます。

[Bugzilla:2136766](#)

reset_purged_snaps_last OSD コマンドが導入される

この機能拡張により、OSD に **purged_snaps** キー (PSN) が存在しないがモニターに存在するケースを解決するために、**reset_purged_snaps_last** OSD コマンドが導入されました。**purged_snaps_last** コマンドはゼロに設定され、その結果、モニターは次の起動時にすべての **purged_snaps** 情報を OSD と共有することになります。

[Bugzilla:2251188](#)

BlueStore の RocksDB 圧縮の有効化

この機能拡張により、メタデータ (特に OMAP) が占めるスペースが少なくなるように、RocksDB 設定が変更され、データの内部圧縮が有効になります。

その結果、* データベースのサイズが小さくなる * コンパクション中の書き込みの増幅が小さくなる * 平均 I/O が増加する * CPU 使用率が増加する

[Bugzilla:2253313](#)

致命的な破損に対して OSD がより耐性を持つようになる

特別な OSD レイヤーオブジェクト "superblock" は以前、ディスクの先頭に配置されていたために上書きされ、致命的な破損が発生していました。

この機能拡張により、OSD "superblock" は冗長化され、ディスク上で移行されます。そのコピーはデータベースに保存されます。OSD は致命的な破損に対してより耐性を持つようになりました。

[Bugzilla:2079897](#)

3.7. RADOS ブロックデバイス (RBD)

rbd_diff_iterate2() API のパフォーマンスの向上

以前のバージョンでは、fast-diff モード (**whole_object == true** で **fast-diff** イメージ機能が有効な場合) で開始時点 (**fromsnapname == NULL**) との差分を取る際に、排他ロックが利用可能な場合、RBD **diff-iterate** のローカルでの実行が保証されませんでした。

この機能拡張により、**rbd_diff_iterate2()** API のパフォーマンスが向上し、**fast-diff** イメージ機能が有効になっている QEMU ライブディスクの同期およびバックアップのユースケースでパフォーマンスが向上します。

[Bugzilla:2258997](#)

第4章 バグ修正

本セクションでは、今回リリースされた Red Hat Ceph Storage で修正されたユーザーに大きな影響を及ぼすバグを説明します。また、セクションでは、以前のバージョンで見つかり修正された既知の問題を説明します。

4.1. CEPHADM ユーティリティー

停止した OSD を起動するために `cephadm` シェルで `--name NODE` フラグを使用しても、間違ったイメージコンテナが返されなくなる

以前は、`cephadm shell --name NODE` コマンドを使用すると、コマンドが間違ったバージョンのツールでコンテナを起動する場合があります。これは、ユーザーがホスト上に、OSD が使用しているものよりも新しい `ceph` コンテナイメージを持っている場合に発生していました。

この修正により、`Cephadm` は、`--name` フラグを指定した `cephadm` シェルコマンドを使用するとき、停止されたデーモンのコンテナイメージを決定します。ユーザーにとって `--name` フラグに関する問題がなくなり、コマンドは期待どおりに動作するようになりました。

[Bugzilla:2258542](#)

4.2. CEPH ANSIBLE ユーティリティー

Playbook は、実行中の RHEL バージョンに一致する RHCS バージョンリポジトリを削除するようになる

以前は、Playbook は、RHEL 9 に存在しないにもかかわらず、Red Hat Ceph Storage 4 リポジトリを RHEL 9 から削除しようとしていました。これにより、Playbook が失敗していました。

この修正により、Playbook は実行中の RHEL バージョンに一致する既存の Red Hat Ceph Storage バージョンリポジトリを削除し、正しいリポジトリが削除されます。

[Bugzilla:2258940](#)

4.3. NFS GANESHA

設定リロードプロセスで消費されたすべてのメモリーが解放される

以前は、リロードエクスポートでは、設定リロードプロセスによって消費されたメモリーがすべて解放されず、メモリーフットプリントが増加していました。

この修正により、設定リロードプロセスによって消費されるすべてのメモリーが解放され、メモリーフットプリントが削減されます。

[Bugzilla:2265322](#)

4.4. CEPH DASHBOARD

Ceph Dashboard でホストを複数指定してボリュームを作成できます。

この修正により、ユーザーは Ceph Dashboard でホストが複数あるボリュームを作成できるようになりました。

[Bugzilla:2241056](#)

未設定のサブボリュームサイズが無限に設定されなくなりました。

以前は、未設定のサブボリュームサイズが無限に設定されていたため、更新が失敗していました。

この修正により、サイズを無限に設定するコードが削除され、更新は想定どおりに機能します。

[Bugzilla:2251192](#)

カーネルマウントコマンドに不足しているオプションが追加されました。

以前は、ファイルシステムを接続するためのカーネルマウントコマンドにいくつかのオプションが欠落していたため、コマンドが意図したとおりに動作しませんでした。

この修正により、不足しているオプションが追加され、カーネルマウントコマンドが期待どおりに動作するようになります。

[Bugzilla:2266256](#)

Ceph Dashboard は、NFS v3 と v4 の両方に対応したエクスポート管理をサポートするようになりました。

以前は、Ceph Dashboard は NFSv4 対応のエクスポート管理のみをサポートしており、NFSv3 対応のエクスポートはサポートしていませんでした。これが原因で、NFSv3 のエクスポートを CLI で管理する場合に破損しました。

この修正により、チェックボックスが追加され NFSv3 ベースのエクスポート管理のサポートが有効になります。Ceph Dashboard は、v3 と v4 の両方に対応したエクスポート管理をサポートするようになりました。

[Bugzilla:2267814](#)

ゾーン作成時にアクセス/シークレットキーが必須ではなくなりました。

以前は、Ceph Object Gateway マルチサイトでゾーンを作成する場合、アクセスキーとシークレットキーが必須でした。このため、ユーザーは最初にゾーンに非システムユーザーのキーを設定し、その後システムユーザーのキーで更新する必要がありました。

この修正により、ゾーンの作成時にアクセスキーとシークレットキーは必須ではなくなります。

[Bugzilla:2275463](#)

マルチサイト設定をインポートしても、フォームの送信時にエラーが発生しなくなりました。

以前は、マルチサイト期間情報にはレルム名が含まれていませんでした。このため、マルチサイト設定をインポートすると、フォームの送信時にエラーが発生しました。

この修正により、期間情報から 'realm' の名前を取得するためのチェックが削除され、トークンのインポートが期待どおりに動作するようになりました。

[Bugzilla:2275861](#)

Ceph Object Gateway のメトリクスラベル名を Prometheus のラベル命名形式に合わせ、Prometheus で表示できるようになる

以前は、メトリクスラベル名が Prometheus のラベル命名形式と一致していなかったため、Ceph Object Gateway メトリクスが Prometheus で表示されませんでした。

この修正により、Ceph Object Gateway メトリクスラベル名のハイフン (-) がすべてアンダースコア (_) に置き換えられ、すべての Ceph Object Gateway メトリクスが Prometheus に表示されるようになります。

[Bugzilla:2276340](#)

Ceph ダッシュボードでフルネームにドットを含められるようになる

以前は、Ceph ダッシュボードでは、検証が誤っていたため、ドットを含むフルネームを作成または変更することができませんでした。

この修正により、Ceph ダッシュボードのフルネームにドットが含まれるように検証が適切に調整されています。

[Bugzilla:2249812](#)

4.5. CEPH ファイルシステム

一貫性を保つために、FSMap の変更を含む MDS メタデータがバッチで追加されるようになりました。

以前は、アップグレード中にモニターが MDS メタデータを追跡しなくなり、PAXOS トランザクションがキャンセルされ、MDS メタデータが利用できなくなることがありました。

この修正により、一貫性を確保するために、FSMap の変更を含む MDS メタデータがバッチで追加されます。`ceph mds metadata` コマンドは、アップグレード後も意図したとおりに機能するようになりました。

[Bugzilla:2144472](#)

ENOTEMPTY 出力が検出され、メッセージが正しく表示されるようになる

以前は、`subvolume group rm` コマンドを実行すると、ボリュームのプラグインで **ENOTEMPTY** 出力が検出されず、具体的なメッセージではなく一般的なエラーメッセージが表示されていました。

この修正により、サブボリュームグループ内にサブボリュームが存在する場合に、`subvolume group rm` コマンドで **ENOTEMPTY** 出力が検出され、メッセージが正しく表示されるようになります。

[Bugzilla:2240138](#)

MDS は、リクエストのクリーンアップの一環として、次のクライアント再生要求を自動的にキューに入れるようになる

以前は、MDS が `up:client-replay` 状態で次のクライアント要求を再生用にキューに入れず、MDS がハングすることがありました。

この修正により、要求のクリーンアップの一環として、次のクライアント再生要求が自動的にキューに入れられ、MDS はフェイルオーバー回復を正常に続行します。

[Bugzilla:2243105](#)

cephfs ミラーリングの全体的なパフォーマンスが向上しました。

この修正により、増分スナップショットの同期が修正され、`cephfs-mirroring` の全体的なパフォーマンスが向上します。

[Bugzilla:2248639](#)

loner member が true に設定されます。

以前は、ファイルロックが LOCK_EXCL_XSYN 状態の場合、非孤立クライアントには空のキャップが発行されていました。ただし、この状態の loner が **false** に設定されているため、ロッカーが Fcb キャップを発行する可能性があります。これは正しくありません。これが原因で、一部のクライアント要求で誤って上限が取り消され、待機時間が無限になり、要求が遅くなる可能性があります。

この修正により、loner メンバーが **true** に設定され、結果として対応する要求がブロックされなくなります。

[Bugzilla:2251258](#)

月次スナップショットのスナップスケジュール繰り返しと保持指定が m から M に変更される

以前は、月次スナップショットのスナップスケジュールの繰り返しと保持の仕様は、他の Ceph コンポーネントと一致していませんでした。

この修正により、仕様が **m** から **M** に変更され、他の Ceph コンポーネントと一致するようになりました。たとえば、毎月 5 つのスナップショットを保持するには、次のコマンドを発行する必要があります。

```
# ceph fs snap-schedule retention add /some/path M 5 --fs cephfs
```

[Bugzilla:2264348](#)

multi-mds クラスタで一部の inode が複製されたときに ceph-mds がクラッシュしなくなる

以前は、ceph-mds のロックアサーションが不正確だったため、multi-mds クラスタで一部の inode が複製されると ceph-mds がクラッシュしていました。

この修正により、アサーション内のロック状態が検証され、クラッシュが発生しなくなります。

[Bugzilla:2265415](#)

date、client_count、filters など、欠落していたフィールドが --dump の出力に追加される

この修正により、**date**、**client_count**、**filter** などの不足しているフィールドが **--dump** 出力に追加されます。

[Bugzilla:2272468](#)

MDS がリカバリー中にアサート関数で失敗しなくなる

以前は、MDS は障害が発生したランクを回復するときにメタデータの破損を誤って報告することがあり、そのためアサート関数が失敗していました。

この修正により、起動手順が修正され、リカバリー中にアサート関数で MDS が失敗しなくなります。

[Bugzilla:2272979](#)

ターゲット mon_host の詳細がピアリストとミラーデーモンステータスから削除される

以前は、スナップショットのミラーピアリストには、ピアリスト以外の情報も表示されていました。この出力により、MON IP が 1 つだけであるべきか、またはすべての MON ホスト IP が表示されるべきかについて混乱が生まれました。

この修正により、mon_host は fs snapshot mirror peer_list コマンドから削除され、ターゲット mon_host の詳細がピアリストとミラーデーモンステータスから削除されました。

[Bugzilla:2277143](#)

ターゲット `mon_host` の詳細がピアリストとミラーデーモンステータスから削除される

以前は、静止プロトコルコードによってリグレーションが発生していました。クライアント要求を強制終了すると、バッチ操作の新しいバッチヘッ드의選択がスキップされます。これにより、古いバッチヘッドリクエストがMDS キャッシュに永久に残り、遅いリクエストとして扱われていました。

この修正により、リクエストを強制終了するときに新しいバッチヘッドが選択され、バッチ操作によってリクエストの遅延が発生しなくなります。

[Bugzilla:2277944](#)

MDS が起動していない場合でも、ファイルシステムのアップグレードが実行される

以前は、すべてのMDSがダウンしている場合、モニターはMDSによるファイルシステムのアップグレードを許可しませんでした。このため、`fail_fs` 設定が `'true'` に設定されていると、アップグレードが失敗します。

この修正により、モニターはMDSが起動していない場合でもアップグレードを実行できるようになります。

[Bugzilla:2244417](#)

4.6. CEPH OBJECT GATEWAY

自動生成された内部トピックは、管理者のトピックリストコマンドに表示されなくなりました。

以前は、自動生成された内部トピックはトピックリストコマンドを介してユーザーに公開されていたため、ユーザーは自分が作成したトピックよりもはるかに多くのトピックを表示できました。

この修正により、内部の自動生成されたトピックは管理者のトピックリストコマンドに表示されなくなり、ユーザーには予想されるトピックのリストのみが表示されるようになりました。

[Bugzilla:1954461](#)

廃止されたバケット名フィールドはトピックリストコマンドに表示されなくなりました。

以前は、プルモード通知 (`pubsub`) の場合、通知はバケットに保存されていました。ただし、このモードは非推奨であるにもかかわらず、トピックリストコマンドには空のバケット名フィールドが引き続き表示されます。

この修正により、空のバケット名フィールドが削除されます。

[Bugzilla:1954463](#)

ライフサイクルの移行時に通知が送信されるようになりました。

以前は、移行時にディスパッチするロジック (有効期限とは別) が欠落していました。このため、移行時に通知が表示されませんでした。

この修正により、新しいロジックが追加され、ライフサイクルの移行時に通知が送信されるようになりました。

[Bugzilla:2166576](#)

`RGWCopObjRequest` が修正され、名前変更操作が期待どおりに動作するようになりました。

以前は、ジッパー変換後の **RGWCopyObjRequest** の初期化が正しくなかったため、名前変更操作が中断されていました。このため、多くの **rgw_rename()** シナリオではソースオブジェクトのコピーに失敗し、二次的な問題が原因でコピーが失敗したにもかかわらずソースも削除されました。

この修正により、**RGWCopyObjRequest** が修正され、さまざまな名前変更操作に対していくつかの単体テストケースが追加されます。

[Bugzilla:2217499](#)

Ceph Object Gateway への不正アクセスができなくなりました。

以前は、Ceph Object Gateway ロールを表す変数が初期化される前に利用されていたため、セグメント違反が発生していました。

この修正により、操作の順序が変更され、不正アクセスが発生しなくなります。必要に応じてロールが適用されます。

[Bugzilla:2252048](#)

間違った CSV オブジェクト構造ごとにエラーメッセージが表示されるようになりました。

以前は、CSV ファイルの中に、二重引用符が閉じられていない場合にアサートが発生し、その後クラッシュが発生していました。

この修正により、間違った CSV オブジェクト構造ごとにポップアップ表示されるエラーメッセージが導入されました。

[Bugzilla:2252396](#)

Ceph Dashboard でユーザー関連情報を照会する際に、'user not found' というエラーが発生しなくなりました。

以前は、Ceph Dashboard では、完全な **user_id** に名前空間が存在し、ダッシュボードで識別できないため、エンドユーザーは Ceph Object Gateway からユーザー関連情報を取得できず、“user not found” というエラーが発生しました。

この修正により、ユーザー情報を取得するために GET リクエストが管理者オペレーションに送信されたときに、テナント、名前空間、および **user_id** を含む完全なユーザー ID が返されるだけでなく、各フィールドが個別に返されるようになります。エンドユーザーは正しい **user_id** を取得できるようになり、これを使用して、Ceph Object Gateway から他のユーザー関連情報をさらに取得できます。

[Bugzilla:2255255](#)

Ceph Object Gateway は、新しいストリームエンコーディング形式で整形形式のペイロードを含むリクエストを渡すようになりました。

以前は、Ceph Object Gateway は **STREAMING-AWS4-HMAC-SHA256-PAYLOAD** および **STREAMING-UNSIGNED-PAYLOAD-TRAILER** エンコード形式を認識せず、リクエストが失敗していました。

この修正により、新しいエンコード形式に提供される新しい末尾の要求署名を認識、解析し、該当する場合は検証するロジックが実装されます。Ceph Object Gateway は、新しいストリームエンコーディング形式で整形形式のペイロードを含むリクエストを渡すようになりました。

[Bugzilla:2256967](#)

radosgw 管理バケットのチェック統計とバケット再シャード統計が正しく計算されるようになりました。

以前は、コードの変更により、バージョン管理されていないオブジェクトからバージョン管理されたオブジェクトに移行したオブジェクトがある場合、radosgw-admin バケットチェック統計の計算とバケット再シャード統計の計算が正しくありませんでした。

この修正により、計算が修正され、不正確なバケット統計出力は生成されなくなりました。

[Bugzilla:2257978](#)

マルチパートアップロードの失敗時に tail オブジェクトが失われなくなりました。

以前は、マルチパートのアップロード中に、タイムアウトなどのシナリオが原因で一部のアップロードに失敗した場合に、アップロードが再起動され、最初のクリーンアップ試行時に、その後に行われるクリーンアップ操作から tail オブジェクトが削除されていました。一部の tail オブジェクトが失われるため、結果として得られる Ceph Object Gateway マルチパートオブジェクトが破損していました。HEAD リクエストには応答しますが、GET リクエスト中は失敗します。

この修正により、コードで、最初の試行が正しくクリーンアップされるようになりました。結果として得られる Ceph Object Gateway マルチパートオブジェクトは破損しなくなり、クライアントが読み取ることができるようになります。

[Bugzilla:2262650](#)

CompleteMultipartUpload とその通知に ETag 値が表示されるようになりました。

以前は、通知に関連する変更が原因で、完了したマルチパートアップロードに対応するオブジェクトハンドルに結果の ETag が含まれませんでした。このため、**CompleteMultipartUpload** とその通知の結果として、マルチパートアップロードを完了するための ETag が存在しませんでした。(正しい ETag が計算され保存されたため、後続の操作には正しい ETag 結果が含まれます。)

この修正により、**CompleteMultipartUpload** はオブジェクトを更新し、期待どおりに出力するようになりました。**CompleteMultipartUpload** とその通知に ETag 値が存在するようになりました。

[Bugzilla:2266579](#)

Swift 経由でコンテナ (バケット) をリストしても、Ceph Object Gateway がクラッシュしなくなりました。

以前は、**swift-object-storage** 呼び出しパスに、対応するバケットでオブジェクトハンドルを更新するための呼び出しがありませんでした (ジッパーバックポートの問題)。このため、S3 Web サイトが同じバケットに対して設定されている場合に、Swift 経由でコンテナ (バケット) をリストすると、Ceph Object Gateway がクラッシュする原因になります。

この修正により、必要なジッパーロジックが追加され、クラッシュは発生しなくなりました。

[Bugzilla:2269038](#)

ライフサイクルポリシーのないバケットでライフサイクルを処理してもクラッシュしなくなる

以前は、ライフサイクルポリシーのないバケットでライフサイクルを手動で処理しようとする、null ポインター参照が発生し、radosgw-admin プログラムがクラッシュしていました。

この修正により、クラッシュを回避するために、ハンドルを操作する前に null バケットハンドルのチェックが行われます。

[Bugzilla:2270402](#)

データプールのゾーン詳細を変更できるようになる

`rgw::zone_create()` 関数は、ゾーン作成時にデフォルトの配置ターゲットとプール名を初期化します。この機能は、以前は `exclusive=false` に設定された `radosgw-admin` ゾーンにも使用されていました。ただし、`zone set` では、STANDARD ストレージクラスの `data_pool` を変更することはできません。

この修正により、`default-placement` ターゲットがすでに存在する場合は上書きされず、データプールのゾーンの詳細を期待どおりに変更できるようになります。

[Bugzilla:2254480](#)

浮動小数点数のモジュロ演算が正しい結果を返すようになる

以前は、浮動小数点数のモジュロ演算で間違った結果が返されていました。

この修正により、SQL エンジンが強化され、浮動小数点のモジュロ演算を処理し、正しい結果を返すようになりました。

[Bugzilla:2254125](#)

SQL ステートメントは、大文字と小文字を区別しないブール式の結果を正しく返す

以前は、SQL ステートメントの一部に大文字のブール式が含まれていたため、誤った解釈と誤った結果が生じていました。

この修正により、ステートメントの解釈は大文字と小文字を区別しなくなり、どの場合でも正しい結果が返されるようになります。

[Bugzilla:2254122](#)

SQL エンジンが正しい NULL 値を返す

以前は、NULL を型へキャストすることが SQL ステートメントに含まれていたため、NULL を返す代わりに間違った結果が返されていました。

この修正により、SQL エンジンが NULL からのキャストを識別し、NULL を返します。

[Bugzilla:2254121](#)

ETags 値が CompleteMultipartUpload とその通知に表示されるようになる

以前は、通知に関連する変更が原因で、完了したマルチパートアップロードに対応するオブジェクトハンドルの結果の ETag が含まれませんでした。その結果、**CompleteMultipartUpload** とその通知には ETag が存在しませんでした。(正しい ETag が計算され保存されたため、後続の操作には正しい ETag 結果が含まれます。)

この修正により、**CompleteMultipartUpload** はオブジェクトを更新し、期待どおりに出力するようになりました。**CompleteMultipartUpload** とその通知に ETag 値が表示されるようになりました。

[Bugzilla:2249744](#)

オブジェクト名にバックスラッシュ (/) が埋め込まれたワークロードをクラウド同期に送信しても、同期エラーが発生しなくなる

以前は、クラウド同期中にオブジェクトパスの URL エスケープが不適切だったため、名前にバックスラッシュ (/) が埋め込まれたオブジェクトがワークロードに含まれていた場合、つまり仮想ディレクトリーパスが使用されていた場合に同期が失敗していました。

この修正により、不適切なエスケープが修正され、オブジェクト名に埋め込まれたバックスラッシュ (/) を含むワークロードが期待どおりにクラウド同期に送信できるようになります。

[Bugzilla:2249068](#)

ブール式を含む SQL ステートメントがブール型を返す

以前は、ブール式 (射影) を含む SQL ステートメントは、ブール型ではなく文字列型を返していました。

この修正により、エンジンは、ステートメント構文に従って文字列をブール式として識別し、ブール型 (true/false) を正常に返すようになりました。

[Bugzilla:2254582](#)

作業スケジューラーが、`should_work` 関数で次の日付を考慮するようになる

以前は、ライフサイクルを現在の時刻で実行を開始するかどうかを決定する `should_work` 関数で使われたロジックでは、次の日付が考慮されていませんでした。その結果、`AB<XY` の場合、カスタム作業時間 "`XY:TW-AB:CD`" によってライフサイクル処理が中断されていました。

この修正により、作業スケジューラーは次の日付を考慮するようになり、さまざまなカスタムライフサイクル作業スケジュールが期待どおりに機能するようになりました。

[Bugzilla:2255938](#)

`merge_and_store_attrs()` メソッドによって属性更新操作が失敗しなくなる

以前は、変更されたバケットインスタンス属性と変更されていないバケットインスタンス属性の調整を処理する `merge_and_store_attrs()` メソッドのバグにより、一部の属性更新操作が暗黙的に失敗していました。このため、バケットのサブセットに対する一部のメタデータ操作は失敗していました。たとえば、レート制限が設定されているバケットでは、バケット所有者の変更ができませんでした。

この修正により、`merge_and_store_attrs()` メソッドが修正され、影響を受けていたすべてのシナリオが正しく動作するようになりました。

[Bugzilla:2262919](#)

チェックサムと不正な形式のトレーラーがクラッシュを引き起こさなくなる

以前は、`java AWS4Test.testMultipartUploadWithPauseAWS4` の実行中に `AWSv4ComplMulti` から例外が発生し、一部のクライアント入力、具体的にはチェックサムトレーラーを使用するクライアント入力によってクラッシュが発生していました。

この修正により、`do_aws4_auth_completion()` に例外ハンドラーが実装されます。チェックサムと不正な形式のトレーラーによってクラッシュが発生することはなくなりました。

[Bugzilla:2266092](#)

改良された末尾のチャンク境界検出の実装

以前は、長さ 0 の末尾のチャンク境界フォーマットの有効な形式が 1 つ処理されませんでした。このため、Ceph Object Gateway は末尾のチャンクの開始を正しく認識できず、403 エラーが発生していました。

この修正により、改良された末尾のチャンク境界検出が実装され、匿名アクセスの場合に予期しない 403 エラーが発生しなくなりました。

[Bugzilla:2266411](#)

Kafka メッセージとアイドルタイムアウトのデフォルト値によるハングが発生しなくなる

以前は、Kafka メッセージとアイドルタイムアウトのデフォルト値により、Kafka ブローカーの待機中にまれにハングが発生していました。

この修正により、タイムアウトが調整され、ハングしなくなりました。

[Bugzilla:2269381](#)

Delete bucket tagging が失敗しなくなる

以前は、RADOS SAL `merge_and_store_attrs()` のロジックが間違っていたため、属性が削除されませんでした。これは `DeleteLifecycle` にも影響を及ぼしました。その結果、一部のコードパスでは純粋な属性削除が有効になりませんでした。

この修正により、バケットタグを保存するロジックでは、`merge_and_store_attrs()` ではなく RADOS SAL `put_info()` が使用されます。Delete bucket tagging が、想定どおりに正常に行われるようになりました。

[Bugzilla:2271806](#)

mtime オブジェクトが S3 PutACL で進み、ACL の変更が適切にレプリケートされるようになりました。

以前は、**S3 PutACL** 操作では **mtime** オブジェクトが更新されませんでした。これにより、タイムスタンプベースのオブジェクト変更チェックが誤って `false` を返したため、一度適用された ACL の変更はレプリケートされませんでした。

この修正により、**mtime** オブジェクトは **S3 PutACL** で常に進み、ACL の変更が適切にレプリケートされます。

[Bugzilla:2271938](#)

すべての移行ケースで通知をディスパッチできるようになる

以前は、移行時に通知をディスパッチするロジックが誤ってクラウド移行ケースにスコープされていたため、プール移行時の通知が送信されませんでした。

この修正により、通知ディスパッチがプール移行スコープに追加され、すべての移行ケースで通知をディスパッチできるようになります。

[Bugzilla:2279607](#)

2106 年以降の RetainUntilDate は切り捨てられなくなり、新しい PutObjectRetention リクエストに対して期待どおりに動作します。

以前は、2106 年以降の **RetainUntilDate** を指定する **PutObjectRetention** リクエストは切り捨てられ、オブジェクトロックの強制にはそれより前の日付が使用されていました。これは、期間が日数で指定される `PutBucketObjectLockConfiguration` リクエストに影響を与えることはありませんでした。

この修正により、**RetainUntilDate** は新しい **PutObjectRetention** リクエストに対して期待どおりに保存され、動作するようになりました。以前存在していたリクエストは自動的に修復されません。既存のリクエストを修正するには、`x-amz-object-lock-retain-until-date` に基づいて **HeadObject** リクエストを使用してリクエストを識別し、**RetainUntilDate** を使用して再度保存します。

詳細は、[S3 の put オブジェクト保持](#) を参照してください。

[Bugzilla:2265890](#)

バケットライフサイクル処理ルールが停止しなくなる

以前は、シャードごとのバケットライフサイクルルールの列挙に、バケットのライフサイクルルールの同時削除に関連する論理エラーが含まれていました。このため、シャードはそのシャードの処理を停止する状態になり、一部のバケットライフサイクルルールが処理されなくなる可能性があります。

この修正により、列挙は削除されたエントリーをスキップできるようになり、この問題に関連するライフサイクル処理の停止が解決されました。

[Bugzilla:2270334](#)

バージョン管理されたバケット内のオブジェクトを削除すると、統計の不一致が発生する

バージョン管理されたバケットには最新のオブジェクトと最新でないオブジェクトが混在しているため、オブジェクトを削除すると、ローカルサイトとリモートサイトでバケットとユーザーの統計に不一致が生じる可能性があります。これにより、どちらのサイトでもオブジェクトリークが発生することはなく、統計の不一致が発生するだけです。

[Bugzilla:1871333](#)

4.7. マルチサイトの CEPH OBJECT GATEWAY

Ceph Object Gateway はオブジェクトの削除中にデッドロックに陥らなくなりました。

以前は、オブジェクトの削除中に、Ceph Object Gateway S3 の **DeleteObjects** がマルチサイトデプロイメントと一緒に実行され、Ceph Object Gateway がデッドロックに陥り、新しいリクエストの受け入れを停止していました。これは、**DeleteObjects** 要求が一度に複数のオブジェクトの削除を処理するために発生しました。

この修正により、レプリケーションログがシリアル化され、デッドロックが防止されます。

[Bugzilla:2249651](#)

起動時での CURL パスの正規化が無効化される

以前は、CURL によってデフォルトで実行されるパス正規化 (RGW レプリケーションスタックの一部) により、レプリケーション中にオブジェクト名が不正に再フォーマットされていました。このため、名前に埋め込まれた `.` および `..` を含むオブジェクトは複製されませんでした。

この修正により、起動時に CURL パスの正規化が無効になり、影響を受けるオブジェクトが期待どおりに複製されるようになります。

[Bugzilla:2265148](#)

プライマリーサイトで転送されたリクエストの認証が失敗しなくなりました。

以前は、STS によって返された一時的な認証情報を使用してリクエストに署名した場合、セカンダリーに発行された S3 リクエストは失敗していました。リクエストがプライマリーに転送され、転送されたリクエストのセッショントークン内の一時的な認証情報と一致しないシステムユーザーの認証情報を使用して署名されたために、失敗が発生しました。認証情報が一致しないため、プライマリーサイトで転送されたリクエストの認証に失敗して、S3 操作が失敗します。

この修正により、リクエストがセカンダリーからプライマリーに転送される場合に、セッショントークン内の一時的な認証情報を使用して認証がバイパスされます。システムユーザーの認証情報を使用して、認証が正常に完了します。

[Bugzilla:2271399](#)

4.8. RADOS

Ceph は、プールにオブジェクトが1つも保存されていない場合、**POOL_APP_NOT_ENABLED** 警告を報告します。

以前は、プールが空の場合、Ceph ステータスはプールアプリケーションの警告を報告できず、RGW プールに対してアプリケーションタグが有効になっていると RGW バケットの作成に失敗しました。

この修正により、プールにオブジェクトが1つも保存されていない場合でも、Ceph は **POOL_APP_NOT_ENABLED** 警告を報告します。

[Bugzilla:2029585](#)

ストレッチクラスター内の2つのサイト間で OSD の重みが不均等であるかどうかのチェックが追加されました。

以前は、ストレッチクラスターのデプロイメント後に OSD の重みが等しいかどうかのチェックは行われていませんでした。これが原因で OSD の重みを不均等にすることができていました。

この修正により、ストレッチクラスター内の2つのサイト間の OSD 重みの不均一性をチェックする機能が追加されました。クラスターは、2つのサイト間の OSD の重みが不均一であるという警告を発するようになりました。

[Bugzilla:2125107](#)

norecover フラグが設定されている間は、Autoscaler は実行されなくなりました。

以前は、**norecover** フラグが設定されている間に autoscaler が実行され、新しい PG が作成され、これらの PG をバックフィルする必要がありました。オブジェクトが不足している場合や、機能が低下している場合にクライアント I/O が無期限に停止しないように、I/O がブロックされているときには、**norecover** フラグを設定した状態で autoscaler を実行できていました。

この修正により、**norecover** フラグが設定されている間は autoscaler は実行されなくなります。

[Bugzilla:2134786](#)

ceph config dump のコマンド出力の一貫性が保たれるようになりました。

以前は、pretty print 形式の出力のない **ceph config dump** コマンドは、ローカライズされたオプション名とその値を表示していました。正規化されたオプションとローカライズされたオプションの例を以下に示します。

```
Normalized: mgr/dashboard/ssl_server_port
```

```
Localized: mgr/dashboard/x/ssl_server_port
```

ただし、pretty-printed 形式のコマンド (たとえば、JSON) では、上記の例に示すように、正規化されたオプション名のみが表示されました。**ceph config dump** コマンドの結果が、pretty-print オプションの有無で矛盾していました。

この修正により、出力は一貫しており、**ceph config dump --format TYPE** コマンド (**TYPE** は pretty print タイプ) を使用するとき、常にローカライズされたオプション名が表示されます。

[Bugzilla:2213766](#)

MGR モジュールは常に1つの CPU コアを占有しなくなり、CPU 使用率は正常になりました。

以前は、配置グループ autoscaler モジュールから負荷の高い呼び出しで、Monitor から OSDMap を取得すると、MGR モジュールが常に CPU コアを1つ専有するようになりました。これが原因で、MGR デーモンの CPU 使用率が高くなりました。

この修正により、配置グループの autoscaler モジュールから実行される OSD マップ呼び出しの数が削減され、CPU 使用率が正常になりました。

[Bugzilla:2241030](#)

OSD の親 (ホスト) の正しい CRUSH の場所が決定されました。

以前は、`osd_memory_target_autotune` オプションが有効になっている場合、メモリーターゲットはホストレベルで適用されていました。これは、メモリーを自動調整するときにホストマスクを使用することによって実行されました。しかし、メモリーターゲットに適用されたコードでは、変更がホストの OSD に伝播されるために、親ホストの正しい CRUSH の場所を決定できません。その結果、マシンによってホストされている OSD のいずれも設定オブザーバーから通知されず、それらの OSD セットの `osd_memory_target` は変更されませんでした。

この修正により、OSD の親 (ホスト) の正しい CRUSH の場所がホストマスクに基づいて決定されます。これにより、変更がホスト上の OSD に伝播されます。自動チューナーが新しい `osd_memory_target` を適用し、変更が反映されるたびに、マシンによってホストされているすべての OSD に通知されます。

[Bugzilla:2244604](#)

クラッシュ/シャットダウンテスト中にモニターが選挙で停止しなくなりました。

以前は、MonitorMap の `disallowed_leaders` 属性が `stretch_mode` に入る場合にのみ条件付きで入力されていました。ただし、復活したモニターが `probing` 状態にあるため、すぐには `Stretch_Mode` にならないインスタンスがありました。これが原因で、クラスター全体のモニター間で `disallowed_leaders` の設定が一致しなくなりました。このため、モニターはリーダーを選出できず、選出が停止し、Ceph が応答しなくなります。

この修正により、`disallowed_leaders` 属性に対応するためにモニターを `Stretch_mode` にする必要がなくなりました。クラッシュ/シャットダウンテスト中にモニターが選挙で停止しなくなりました。

[Bugzilla:2248939](#)

属性の取得エラーメッセージが表示されなくなりました。

以前は、`ceph-objectstore-tool` は `--op list` を使用すると pgmeta オブジェクトをリストし、`Error getting attr on` というメッセージを表示していました。

この修正により、pgmeta オブジェクトはスキップされ、エラーメッセージが表示されなくなります。

[Bugzilla:2251004](#)

アロケーター内の LBA アライメントは使用されなくなり、OSD デーモンは割り当ての失敗が原因でアサートしなくなりました。

以前は、OSD デーモンがアサートして再起動に失敗し、データが使用できなくなったり、データが失われたりすることがありました。これは、アロケーターの要求が 4000 に達し、アロケーターが別の割り当て単位で設定されている場合に OSD デーモンがアサートしないことが原因で発生していました。

この修正により、アロケーター内の LBA アライメントは使用されず、OSD デーモンが割り当て失敗によりアサートされなくなりました。

[Bugzilla:2260306](#)

`libcephsqlite` ライブラリーを使用する `sqlite` データベースは、短い読み取りでメモリーページを正しくゼロにできないことが原因で破損することがなくなりました。

以前は、libcephsqlite は短い読み取りを正しく処理せず、sqlite データベースの破損を引き起こす可能性があります。

この修正により、libcephsqlite は、破損の可能性を回避するために、短い読み取りに対してページを正しくゼロにします。

[Bugzilla:2240139](#)

4.9. RBD ミラーリング

強制昇格中にピアサイトがダウンした場合、イメージステータスの説明に "orphan (force promoting)" と表示されるようになりました。

以前は、強制昇格時にピアサイトがダウンすると、イメージステータスの説明に "local image linked to unknown peer" と表示されていましたが、これは明確な説明ではありません。

この修正により、ミラーデーモンが改善され、イメージステータスの説明が "orphan (force promoting)" として表示されるようになります。

[Bugzilla:2190366](#)

rbd_support モジュールは、クライアントの繰り返しブロックリストからの回復に失敗しなくなりました。

以前は、rbd_support モジュールの再帰デッドロック、**rbd_support** モジュールの librbd クライアントの競合状態、および ceph-mgr をクラッシュさせることがある librbd cython バインディングのバグが原因で、**rbd_support** モジュールがクライアントの繰り返しのブロックリストから回復できないことがありました。

このリリースでは、これら 3 つの問題がすべて修正され、rbd_support`モジュールはクライアントの繰り返しのブロックリストからの回復に失敗しなくなりました。

[Bugzilla:2247531](#)

第5章 既知の問題

本セクションでは、Red Hat Ceph Storage の今回リリースで見つかった既知の問題を説明します。

5.1. CEPHADM ユーティリティー

Cephadm は OSD をドレーンするときに以前の OSD の重みを維持しません。

Cephadm は、OSD をドレーンするときに以前の OSD の重みを維持しません。このため、**ceph orch osd rm <osd-id>** コマンドを実行し、その後 OSD の削除を停止すると、Cephadm で OSD のクラッシュウェイトが元の値に戻らなくなりました。クラッシュの重みは 0 のままになります。

回避策として、ユーザーは OSD のクラッシュの重みを手動で元の値に調整するか、OSD を完全に削除して新しい OSD をデプロイする必要があります。削除プロセスが開始されるまで OSD のクラッシュの重みは元の値に戻らないため、**ceph orch osd rm** 操作をキャンセルする際に注意する必要があります。

[Bugzilla:2247211](#)

Ceph Object Gateway のレلم bootstrap コマンドを繰り返し使用すると、ゾングループのホスト名の設定が失敗する

Ceph Object Gateway のレلم bootstrap コマンドを使用して、ゾングループのホスト名を複数回設定すると失敗します。このため、Ceph Object Gateway のレلم bootstrap コマンドを繰り返し使用してレلم/ゾングループ/ゾーンを再作成しても、**zonegroup_hostnames** フィールドでは正しく機能せず、ホスト名がゾングループに設定されません。

回避策として、radosgw-admin ツールを使用してゾングループのホスト名を手動で設定します。[Bugzilla:2241321](#)

5.2. CEPH OBJECT GATEWAY

大きな Parquet オブジェクトに対するクエリーを処理すると、Ceph Object Gateway プロセスが停止します。

以前は、Parquet オブジェクトに対するクエリーを処理する際に、そのオブジェクトがチャンクごとに読み取られることがあり、これらのチャンクがかなり大きくなるがありました。これにより、Ceph Object Gateway は、ローエンドのマシンには大きすぎるメモリーに大容量のバッファをロードすることになります。特に、Ceph Object Gateway が大量のメモリーを消費する OSD プロセスと同じ場所に配置されている場合は、この傾向が顕著になります。この状況により、OS は Ceph Object Gateway プロセスを強制終了します。

回避策として、Ceph Object Gateway を別のノードに配置すると、Ceph Object Gateway により多くのメモリーが残り、処理を正常に完了できるようになります。

[Bugzilla:2275323](#)

現在の RGW STS 実装では、1024 バイトを超える暗号化鍵はサポートされていません。

現在の RGW STS 実装では、1024 バイトを超える暗号化鍵はサポートされていません。

回避策として、**Keycloak: realm settings - keys** で、'rsa-enc-generated' プロバイダーを編集して、優先度を 100 ではなく 90 に設定し、**keySize** を 2048 ではなく 1024 に設定します。

[Bugzilla:2276931](#)

オブジェクト圧縮と暗号化のための Intel QAT アクセラレーション

圧縮と暗号化を有効にするときにノードの CPU 使用率を削減し、Ceph Object Gateway のパフォーマンスを向上させることができるように、Intel QuickAssist テクノロジー (QAT) が実装されています。このリリースでは、QAT は新しい設定 (Greenfield) でのみ設定できます。この機能は制限されています。QAT Ceph Object Gateway デーモンは、非 QAT (通常の) Ceph Object Gateway デーモンと同じクラスター内に設定できません。

[Bugzilla:2284394](#)

第6章 ソース

更新された Red Hat Ceph Storage ソースコードパッケージは、以下の場所から入手できます。

- Red Hat Enterprise Linux 9 の場合:
<https://ftp.redhat.com/redhat/linux/enterprise/9Base/en/RHCEPH/SRPMS/>