Red Hat OpenShift Container Storage 4.8

4.8 Release Notes

Release notes for feature and enhancements, known issues, and other important release information

Last Updated: 2022-05-13

Red Hat OpenShift Container Storage 4.8 4.8 Release Notes

Release notes for feature and enhancements, known issues, and other important release information

Legal Notice

Abstract

The release notes for Red Hat OpenShift Container Storage 4.8 summarize all new features and enhancements, notable technical changes, and any known bugs upon general availability.

Table of Contents

CHAPTER 1. INTRODUCTION

Red Hat OpenShift Container Storage is software-defined storage that is optimized for container environments. It runs as an operator on OpenShift Container Platform to provide highly integrated and simplified persistent storage management for containers.

Red Hat OpenShift Container Storage is integrated into the latest Red Hat OpenShift Container Platform to address platform services, application portability, and persistence challenges. It provides a highly scalable backend for the next generation of cloud-native applications, built on a new technology stack that includes Red Hat Ceph Storage, the Rook.io Operator, and NooBaa's Multicloud Object Gateway technology.

Red Hat OpenShift Container Storage provides a trusted, enterprise-grade application development environment that simplifies and enhances the user experience across the application lifecycle in a number of ways:

- Provides block storage for databases.
- Shared file storage for continuous integration, messaging, and data aggregation.
- Object storage for cloud-first development, archival, backup, and media storage.
- Scale applications and data exponentially.
- Attach and detach persistent data volumes at an accelerated rate.
- Stretch clusters across multiple data-centers or availability zones.
- Establish a comprehensive application container registry.
- Support the next generation of OpenShift workloads such as Data Analytics, Artificial Intelligence, Machine Learning, Deep Learning, and Internet of Things (IoT).
- Dynamically provision not only application containers, but data service volumes and containers, as well as additional OpenShift Container Platform nodes, Elastic Block Store (EBS) volumes and other infrastructure services.

1.1. ABOUT THIS RELEASE

Red Hat OpenShift Container Storage 4.8 (RHBA-2021:3002 and RHBA-2021:3003) is now available. New enhancements, features, and known issues that pertain to OpenShift Container Storage 4.8 are included in this topic.

Red Hat OpenShift Container Storage 4.8 is supported on the Red Hat OpenShift Container Platform versions 4.8. For more information, see Red Hat OpenShift Container Storage Supportability and Interoperability Guide.

With the release of OpenShift Container Storage 4.8, version 4.5 is now end of life. For more information, see Red Hat OpenShift Container Platform Life Cycle Policy .

CHAPTER 2. NEW FEATURES

This section describes new features introduced in Red Hat OpenShift Container Storage 4.8.

Compact deployment general availability support

OpenShift Container Storage can now be installed on a three-node OpenShift compact bare metal cluster, where all the workloads run on three strong master nodes. There are no worker or storage nodes.

For information on how to configure OpenShift Container Platform on a compact bare metal cluster, see Configuring a three-node cluster and Delivering a Three-node Architecture for Edge Deployments.

Caching policy for object buckets

In Red Hat OpenShift Container Storage's Multicloud Object Gateway, you can now create a cache bucket. A cache bucket is a namespace bucket with a hub target and a cache target. For more information, see Caching policy for object buckets.

Persistent volume encryption through storage class general availability support

You can encrypt persistent volumes (block only) with storage class encryption using an external Key Management System (KMS) to store device encryption keys. Persistent volume encryption is only available for RBD persistent volumes. Storage class encryption is supported in OpenShift Container Storage 4.7 or higher. For more information, see how to create a storage class with persistent volume encryption.

Persistent Volume encryption now also supports snapshots and clones.

Thick provisioned storage for VMware platform

You can now use thick-provisioned storage, in addition to the thin-provisioned storage for VMware-hosted OpenShift Container Platform for better performance and security. When you need the flexibility to use the thick-provisioned storage in OpenShift Container storage, you must create a storage class with the zeroedthick or eagerzeroedthick disk format in OpenShift Container Platform. You get the option to select the storage class that is created in addition to the default thin storage class while you create the OpenShift Container Storage cluster service.

For more information, see Creating an OpenShift Container Storage Cluster Service

New pool management user interface

The new management capability provides you with a simple, easy to use interface to create a storage class or delete the pool that was automatically attached to it; or, if you want to update the characteristics of an existing pool (such as compression, replica). This feature is not a replacement for the existing storage class configuration. For more information, see Block pools chapter in *Managing and Allocating Storage Resources guide*.

Support for an air-gap disconnected environment on IBM Power Systems and IBM Z infrastructure

Deployment of OpenShift Container Storage 4.8 is now possible in air-gapped environment, which does not have internet connectivity.

Multicloud Object Gateway on IBM Power Systems and IBM Z infrastructure

Red Hat OpenShift Container Storage 4.8 on IBM Power Systems and IBM Z infrastructure adds support for Noobaa's multicloud object service which provides multicloud and hybrid capabilities for object workloads. By default, Multicloud Object Gateway uses a default backing store which is cloud

native or RGW.

Encrypted storage data on IBM Power Systems

Administrators can now choose to encrypt all data in the OpenShift Container Storage 4.8 cluster as part of the deployment process. See Data encryption options for more information.

Support for DASD on IBM Z infrastructure

DASD is now supported for storage nodes on IBM Z infrastructure.

CHAPTER 3. ENHANCEMENTS

This section describes major enhancements introduced in Red Hat OpenShift Container Storage 4.8.

Added a new alert to improve notification to the users in case one or more OSD requests are taking a long time to process

This alert is important to notify OpenShift Container Storage administrators about the slow operations which can be an indication of extreme load, a slow storage device, or a software bug. Users can check ceph status to find out the cause for slowness.

ClusterObjectStoreState alert message is generated when RADOS Object Gateway (RGW) is not available or is unhealthy.

Previously, the **ClusterObjectStoreState** alert message was not generated if the RADOS Object Gateway (RGW) was not available or was unhealthy. With a fix implemented in the OpenShift Container Storage operator, users can now see the ClusterObjectStoreState alert when RADOS Object Gateway (RGW) is not available or is unhealthy.

Ability to enable or disable compression in a pool

With OpenShift Container Storage 4.8 onwards, you can enable or disable the compression in a pool as a day two operation using the user interface.

Added ability to create namespace buckets using the OpenShift Container Platform user interface

Namespace buckets can be added using the OpenShift Container Platform user interface. Namespace buckets provide an aggregated view of existing object buckets in the cloud or S3 compatible storage on premise. For more information about adding namespace buckets using the user interface, see Adding namespace bucket using the OpenShift Container Platform user interface.

Utilizing all the available devices during initial deployment and scaling up for local storage devices

For all the local storage devices in attached mode deployments, the storage cluster now utilizes all the locally available storage devices. Similarly, during scaling up by adding capacity, all the available storage devices can be added.

Prevent adding **no-out** flag on the failure domain if an OSD is down due to reasons other than node drain

When an OSD is down due to disk failure, a **no-out** flag is added on the failure domain. This prevents the OSD from being marked out using standard ceph mon_osd_down_out_interval. With this update, when an OSD is down due to reasons other than node drain, say, disk failure, in such a situation, if the pgs are unhealthy then rook will create a blocking PodDisruptionBudget on other failure domains to prevent further node drains on them. **noout** flag won't be set on the node in this case. If the OSD is down but all the pgs are **active+clean**, the cluster will be treated as fully healthy. The default PodDisruptionBudget (with maxUnavailable=1) will be added back and the blocking ones will be deleted.

CHAPTER 4. TECHNOLOGY PREVIEWS

Technology Preview features are provided with a limited support scope, as detailed on the Customer Portal: Technology Preview Features Support Scope .

This section describes technology preview features introduced in Red Hat OpenShift Container Storage 4.8 under Technology Preview support limitations.

Disaster recovery using arbiter

With this release, Red Hat OpenShift Container Storage provides Metro-DR stretched cluster (arbiter) feature that allows you to enable a single cluster to be stretched across two zones with a third zone as the location for the arbiter during the storage cluster creation.

For more information, see Disaster Recovery in the Planning your deployment guide.

Multi Network plugin (Multus) support

Supports the ability to use multi-container network plugin (multus) to improve security and performance by isolating networks. This feature has been tested only on bare metal and VMWare deployments. For more information about multus, see Multi network plug-in (Multus) support .

NOTE

In the case of a deletion of a plugin pod, the data is not accessible until a node restart takes place. This is a known issue. For more information, see If a plugin pod is deleted, the data becomes inaccessible until a node restart takes place.

CHAPTER 5. DEVELOPER PREVIEWS

This section describes developer preview features introduced in Red Hat OpenShift Container Storage 4.8.

Developer preview feature is subject to Developer preview support limitations. Developer preview releases are not intended to be run in production environments. The clusters deployed with the developer preview features are considered to be development clusters and are not supported through the Red Hat Customer Portal case management system. If you need assistance with developer preview features, reach out to the ocs-devpreview@redhat.com mailing list and a member of the Red Hat Development Team will assist you as quickly as possible based on availability and work schedules.

Data segregation per hosts group

Workloads can use specific isolated IO paths and spread across specific storage nodes to limit the impact in case of partial cluster failure or a noisy neighbour and isolate tenants.

For more information, see knowledge base article.

Regional disaster recovery

Red Hat OpenShift Container Storage provides multi-cluster asynchronous replication of storage volumes across two OpenShift Container Storage clusters serving two Openshift Container Platform clusters. Any stateful application, including its stateless counterparts need some preparation prior to deploying the same on a peer cluster.

Object Storage Device weight and pod affinity configuration

You can now efficiently use partitioned disks with a non OpenShift Container Storage device to maximize capacity utilization without blocking that workload. You can assign an object storage device (OSD) to all the solid state drives (SSDs) on a host machine including a partitioned device that has a dedicated partition for the operating system. You can reduce the load on the OSD for better efficiency as it shares the same physical device. To reduce the load, you can configure the OSD weight and pod affinity parameters in the Storage Cluster CR.

For more information, see knowledge base article.

Flexibility in OpenShift Container Storage component deployment

Flexibility in component deployment is now possible with the capability to disable the components multicloud object gateway, RGW, and CephFS during deployment. It is also possible to disable or enable these components after the OpenShift Container Storage deployment. This flexibility helps to reduce resource costs when using Amazon S3.

For more information, see knowledge base article.

CHAPTER 6. BUG FIXES

This section describes notable bug fixes introduced in Red Hat OpenShift Container Storage 4.8.

Arbiter and flexible scaling can't be enabled at the same time.

When arbiter and flexible scaling both are enabled, the storage cluster was shown in **READY** state even though there were logs or messages with the error **arbiter and flexibleScaling both can't be enabled**. This was happening because of the incorrect specs of the storage cluster CR. With this update, the storage cluster is in "ERROR" state with the correct error message.

(BZ#1946595)

Buckets are always deleted when cleanup is required by the library

Previously, on OBC creation failure, lib-bucket-provisioner sent a delete request to the provisioner for cleanup purposes before retrying. Noobaa provisioners would look at the reclaim policy of the object bucket but in some cases did not delete the underlying bucket. With this update, on cleanup scenarios, the underlying bucket should have been deleted regardless of the reclaim policy.

(BZ#1947796)

Collect configuration of each OSD attached

Previously, there was no way to find out detailed configuration of each OSD. With this update, **must-gather** collects all the configuration of an OSD to further improve the debugging.

(BZ#1962755)

gRPC metrics are now disabled by default

Earlier, the **cephcsi** pods exposed the remote procedure call (gRPC) metrics for debugging purposes. The **cephcsi** node plugin pods used the host ports 9091 for CephFS and 9090 for RBD on the node where the **cephcsi** node plugin pods were running. This meant the **cephcsi** pods failed to come up. With this update, gRPC metrics are disabled by default and **cephcsi** pods do not use ports 9091 and 9090 on the node where the node plugin pods are running.

(BZ#1923819)

MDS report oversized cache

Rook has not previously applied mds_cache_memory_limit upon upgrades. This means OpenShift Container Storage 4.2 clusters that did not have that option applied were not updated with the correct value, which is typically half the size of the pod's memory limit. Therefore, MDSs in standby-replay may report oversized cache.

(BZ#1944148)

Newly restored PVC can now be mounted on nodes

Previously, a bug in Ceph-CSI driver caused a wrong 'RBD image not found' error while mounting newly restored PVC with deleted parent snapshot on nodes with Red Hat Enterprise Linux version of less than 8.2 (with no deep flattening feature). This issue was fixed by flattening the newly restored PVC before mounting on nodes with Red Hat Enterprise Linux version of less than 8.2 (with no deep flattening feature).

(BZ#1956232)

Reliable mon quorum

Earlier, if the operator was restarted during a mon failover, the operator could erroneously remove the new mon. Hence, the mon quorum was at risk when the operator removed the new mon. With this update, the operator will restore the state when the mon failover is in progress and properly complete the mon failover after the operator is restarted. Now, the mon quorum is more reliable in the node drains and mon failover scenarios.

(BZ#1955831)

CHAPTER 7. KNOWN ISSUES

This section describes known issues in Red Hat OpenShift Container Storage 4.8.

Arbiter nodes can not be labelled with the OpenShift Container Storage node label

Arbiter nodes are considered as valid non-arbiter nodes if they are labelled with the OpenShift Container Storage node label, **cluster.ocs.openshift.io/openshift-storage**. This means the placement for the non-arbiter resources becomes undetermined. To work around this issue, do not label the arbiter nodes with the OpenShift Container Storage node label so that only arbiter resources are placed on the arbiter nodes.

(BZ#1947110)

Ceph status is HEALTH_WARN after disk replacement

After disk replacement, a warning **1 daemons have recently crashed** is seen even if all OSD pods are up and running. This warning causes a change in Ceph's status. The Ceph status should be HEALTH_OK instead of HEALTH_WARN. To workaround this issue, rsh to the ceph-tools pod and silence the warning, the Ceph health will then be back to HEALTH_OK.

(BZ#1896810)

Monitoring spec is reset in CephCluster resource

Monitoring spec becomes empty whenever **ocs-operator** is restarted or during an upgrade. This has no impact on the functionality but If you are looking for the monitoring endpoint details, you will find it empty.

To resolve this issue, update the **rook-ceph-external-cluster-details** secret after upgrading from 4.7 to 4.8 so that the details of all endpoints (such as comma-separated IP addresses of active and standby MGRs) are updated into the "MonitoringEndpoint" data key. This also helps to avoid any problems in the future raised due to differences in the number of endpoints in fresh versus upgraded clusters.

(BZ#1984735)

Issue with noobaa-db-pg-0

noobaa-db-pg-0 pod does not migrate to other nodes when the hosting node goes down. NooBaa will not work when a node is down as migration of noobaa-db-pg-0 pod is blocked.

(BZ#1783961)

If a plugin pod is deleted, the data becomes inaccessible until a node restart takes place

The issue is caused because **netns** of the mount gets destroyed when the **csi-cephfsplugin** pod is restarted which results in **csi-cephfsplugin** locking up all mounted volumes. This issue is seen only in clusters enabled with multus.

The issue is resolved when you restart the node on which **csi-cephfsplugin** was restarted after the deletion.

(BZ#1979561)

Encryption passphrase is stored in the source KMS for restoring volumes from the snapshot

When the parent and the restored PVC have different StorageClasses with different backend paths in the KMS settings, the restored PVC goes into the **Bound** state and the encryption passphrase is created in the backend path of the KMS settings from the snapshot. The restored PVC can fails to get

attached to a Pod as the checks for the encryption passphrase use the settings linked from the 2nd StorageClass path, where the encryption passphrase can not be found in the backend path.

To prevent the issue, PVCs should always use the same KMS settings when creating snapshots and restoring them.

(BZ#1975730)

Keys are still listed in Vault after deleting encrypted PVCs while using the kv-v2 secret engine

Hashicorp Vault added a feature for the key-value store v2 where deletion of the stored keys makes it possible to recover the contents in case the metadata of the deleted key is not removed in a separate step. When using key-value v2 storage for secrets in Hashicorp Vault, deletion of volumes will not remove the metadata of the encryption passphrase from the KMS. Though it is possible to restore the encryption passphrase at a later time. These partially deleted keys are not automatically cleaned up by the KMS.

You can resolve this issue by manually deleting the metadata of the removed keys. Any key that has the **deletion_time** set in the metadata can be assumed to have been deleted when key-value storage v1 was used but kept available with v2.

(BZ#1979244)

Restore Snapshot/Clone operations with greater size than parent PVC results in endless loop

Ceph CSI does not support restoring a snapshot or creating clones with a size greater than the parent PVC. Therefore, Restore Snapshot/Clone operations with a greater size results in an endless loop. To workaround this issue, delete the pending PVC. In order to get a larger PVC, complete one of the following based on the operation you are using: If using Snapshots, restore the existing snapshot to create a volume of the same size as the parent PVC, then attach it to a pod and expand the PVC to the required size. For more information, see Volume snapshots. If using Clone, clone the parent PVC to create a volume of the same size as the parent PVC, then attach it to a pod and expand the PVC to the required size. For more information, see Volume cloning.

(BZ#1870334)

PVC restored from a snapshot or cloned from a thick provisioned PVC, is not thick provisioned

When the snapshot of a thick provisioned PVC is restored using **thick provisioning** enabled storage class, the restored volume is not thick provisioned. The restored PVC reaches the **Bound** state without thick provisioning. This can only be fixed when RHCS-5.x is used. Older Ceph versions do not support copying of zero-filled data blocks (used when thick-provisioning).

Currently, to resolve the issue with RHCS-4.x based deployments is to mark PVC-cloning and snapshot-restoring of thick-provisioned volumes as a limitation. The newly created volumes will become thin-provisioned.

(BZ#1959793)

Deleting the pending PVC and RBD provisioner leader pod while the thick provisioning is progressing, will leave a stale image and OMAP metadata

When an RBD PVC is being thick provisioned, the Persistent Volume Claim (PVC) is in a **Pending** state. If the RBD provisioner leader and the PVC itself are deleted, the RBD image and OMAP metadata will not be deleted.

To address this issue, do not delete the PVC while the thick provisioning is in progress.

(BZ#1962956)

Provisioning attempts did not stop once the storage cluster utilization reached 85% or even after deleting the PVC.

If the storage cluster utilization reaches 85% while an RBD thick PVC is being provisioned, the provisioning attempt will not stop automatically by deleting the pending PVC and the RBD image will not get deleted even after deleting the pending PVC.

The best approach is not to start provisioning if the requested size is beyond the available storage.

(BZ#1965016)

Keys for OSDs in the Vault are not deleted during uninstall when kv-v2 is used

Key encryption keys **data** are soft-deleted from Vault during cluster deletion when the Vault K/V Secret engine is version 2. This means any version of the Key can be retrieved and so the deletion is undone. The metadata is still visible so the key can be restored. If this is causing inconvenience, the key can still be deleted manually using the vault command with the "destroy" argument.

(BZ#1975323)

Deletion of CephBlockPool gets stuck and blocks the creation of new pools

In a Multus enabled cluster, the Rook Operator does not have the network annotations and thus does not have access to the OSD network. This means that when running "rbd" type commands during pool cleanup, the command hangs since it cannot contact the OSDs. The workaround is to delete the **CephBlockPool** manually using the toolbox.

(BZ#1983756)

Device replacement action cannot be performed through the user interface for an encrypted OpenShift Container Storage cluster

On an encrypted OpenShift Container Storage cluster, the discovery result CR discovers the device backed by a Ceph OSD (Object Storage Daemon) differently from the one reported in the Ceph alerts. When clicking the alert, the user is presented with Disk not found message. Due to the mismatch, console UI cannot enable the disk replacement option for an OpenShift Container Storage user. To workaround this issue, use the CLI procedure for failed device replacement in the Replacing Devices guide.

(BZ#1906002)

False Alert for PVCs with volumeMode as block

Due to a change in Kubernetes, there is a regression in a Prometheus alert of OpenShift Container Platform. This change has impacted the following:

Alert: KubePersistentVolumeFillingUp.

PVCs: PVCs in volumeMode: Block

Matching regular expression in the namespaces: "(openshift-.|kube-.|default|logging)"

Metric: **kubelet_volume_stats_available_bytes**

As a result, the alert, kubelet_volume_stats_available_bytes reports the available size as 0 from the time

of PVC creation, and a false alert is fired for all the PVCs in volumeMode: Block in namespaces that match the regular expression: "(openshift-.|kube-.|default|logging)". This impacts all the PVCs created for OSD device sets of OpenShift Container Storage deployed in internal and internal-attached modes and on different infrastructures like Amazon Web Services, VMware, Baremetal, and so on. This also impacts the customer workload PVCs.

Currently, there is no workaround available for this issue until it is fixed in one of the upcoming minor releases of OpenShift Container Platform 4.8.z. Hence, address any alert regarding storage capacity by OpenShift Container Storage very promptly and with urgency.

(BZ#1984817)

Critical alert notification is sent after installation of arbiter storage cluster, when ceph object user for **cephobjectstore** fails to be created during storage cluster reinstallation.

In a storage cluster containing a **CephCluster** and one or more **CephObjectStores**, if the **CephCluster** resource is deleted before all of the **CephObjectStore** resources are fully deleted, the Rook Operator can still keep connection details about the CephObjectStore(s) in memory. If the same **CephCluster** and CephObjectStore(s) are re-created, the CephObjectStore(s) may enter "Failed" state.

To avoid this issue, delete the CephObjectStore(s) completely before removing the CephCluster.

- If you do not wish to wait for the CephObjectStore(s) to be deleted, restarting the Rook Operator (by deleting the Operator Pod) will avoid the issue if done after uninstall.
- If you are actively experiencing this issue, restarting the Rook Operator will resolve it by clearing the Operator's memory of old **CephObjectStore** connection details.

(BZ#1974344)