RBD MIRRORING

RBD images can be asynchronously mirrored between two Ceph clusters. This capability uses the RBD journaling image feature to ensure crash-consistent replication between clusters. Mirroring is configured on a per-pool basis within peer clusters and can be configured to automatically mirror all images within a pool or only a specific subset of images. Mirroring is configured using the `rbd` command. The `rbd-mirror` daemon is responsible for pulling image updates from the remote, peer cluster and applying them to the image within the local cluster.

Note: RBD mirroring requires the Ceph Jewel release or later.

Important: To use RBD mirroring, you must have two Ceph clusters, each running the `rbd-mirror` daemon.

POOL CONFIGURATION

The following procedures demonstrate how to perform the basic administrative tasks to configure mirroring using the `rbd` command. Mirroring is configured on a per-pool basis within the Ceph clusters.

The pool configuration steps should be performed on both peer clusters. These procedures assume two clusters, named “local” and “remote”, are accessible from a single host for clarity.

See the rbd manpage for additional details of how to connect to different Ceph clusters.

Note: The cluster name in the following examples corresponds to a Ceph configuration file of the same name (e.g. /etc/ceph/remote.conf). See the ceph-conf documentation for how to configure multiple clusters.

ENABLE MIRRORING

To enable mirroring on a pool with `rbd`, specify the `mirror pool enable` command, the pool name, and the mirroring mode:

```
rbd mirror pool enable {pool-name} {mode}
```

The mirroring mode can either be `pool` or `image`:

- **pool**: When configured in `pool` mode, all images in the pool with the journaling feature enabled are mirrored.
- **image**: When configured in `image` mode, mirroring needs to be explicitly enabled on each image.

For example:

```
rbd --cluster local mirror pool enable image-pool pool
rbd --cluster remote mirror pool enable image-pool pool
```

DISABLE MIRRORING

To disable mirroring on a pool with `rbd`, specify the `mirror pool disable` command and the pool name:

```
rbd mirror pool disable {pool-name}
```

When mirroring is disabled on a pool in this way, mirroring will also be disabled on any images (within the pool) for which mirroring was enabled explicitly.

For example:

```
rbd --cluster local mirror pool disable image-pool
rbd --cluster remote mirror pool disable image-pool
```

ADD CLUSTER PEER

In order for the `rbd-mirror` daemon to discover its peer cluster, the peer needs to be registered to the pool. To add a mirroring peer Ceph cluster with `rbd`, specify the `mirror pool peer add` command, the pool name, and a cluster specification:

```
rbd mirror pool peer add {pool-name} {client-name}@{cluster-name}
```

For example:

```
rbd --cluster local mirror pool peer add image-pool client.remote@remote
rbd --cluster remote mirror pool peer add image-pool client.local@local
```

REMOVE CLUSTER PEER

To remove a mirroring peer Ceph cluster with `rbd`, specify the `mirror pool peer remove` command, the pool name, and the peer UUID (available from the `rbd mirror pool info` command):

```
rbd mirror pool peer remove {pool-name} {peer-uuid}
```

For example:

```
rbd --cluster local mirror pool peer remove image-pool 55672766-c02b-4729-8567-f13a66893445
rbd --cluster remote mirror pool peer remove image-pool 60c0e299-b38f-4234-91f6-eed0a367be08
```

DATA POOLS

When creating images in the destination cluster, `rbd-mirror` selects a data pool as follows:

1. If the destination cluster has a default data pool configured (with the `rbd_default_data_pool` configuration option), it will be used.
2. Otherwise, if the source image uses a separate data pool, and a pool with the same name exists on the destination cluster, that pool will be used.
3. If neither of the above is true, no data pool will be set.

IMAGE CONFIGURATION

Unlike pool configuration, image configuration only needs to be performed against a single mirroring peer Ceph cluster.

Mirrored RBD images are designated as either primary or non-primary. This is a property of the image and not the pool. Images that are designated as non-primary cannot be modified.

Images are automatically promoted to primary when mirroring is first enabled on an image (either implicitly if the pool mirror mode was **pool** and the image has the journaling image feature enabled, or explicitly enabled by the `rbd` command).

ENABLE IMAGE JOURNALING SUPPORT

RBD mirroring uses the RBD journaling feature to ensure that the replicated image always remains crash-consistent. Before an image can be mirrored to a peer cluster, the journaling feature must be enabled. The feature can be enabled at image creation time by providing the `--image-feature exclusive-lock,journaling` option to the `rbd` command.

Alternatively, the journaling feature can be dynamically enabled on pre-existing RBD images. To enable journaling with `rbd`, specify the `feature enable` command, the pool and image name, and the feature name:

```
rbd feature enable {pool-name}/{image-name} {feature-name}
```

For example:

```
rbd --cluster local feature enable image-pool/image-1 journaling
```

Note: The journaling feature is dependent on the exclusive-lock feature. If the exclusive-lock feature is not already enabled, it should be enabled prior to enabling the journaling feature.

Tip: You can enable journaling on all new images by default by adding `rbd default features = 125` to your Ceph configuration file.

ENABLE IMAGE MIRRORING

If the mirroring is configured in `image` mode for the image's pool, then it is necessary to explicitly enable mirroring for each image within the pool. To enable mirroring for a specific image with `rbd`, specify the `mirror image enable` command along with the pool and image name:

```
rbd mirror image enable {pool-name}/{image-name}
```

For example:

```
rbd --cluster local mirror image enable image-pool/image-1
```

DISABLE IMAGE MIRRORING

To disable mirroring for a specific image with `rbd`, specify the `mirror image disable` command along with the pool and image name:

```
rbd mirror image disable {pool-name}/{image-name}
```

For example:

```
rbd --cluster local mirror image disable image-pool/image-1
```

IMAGE PROMOTION AND DEMOTION

In a failover scenario where the primary designation needs to be moved to the image in the peer Ceph cluster, access to the primary image should be stopped (e.g. power down the VM or remove the associated drive from a VM), demote the current primary image, promote the new primary image, and resume access to the image on the alternate cluster.

Note: RBD only provides the necessary tools to facilitate an orderly failover of an image. An external mechanism is required to coordinate the full failover process (e.g. closing the image before demotion).

To demote a specific image to non-primary with `rbd`, specify the `mirror image demote` command along with the pool and image name:

```
rbd mirror image demote {pool-name}/{image-name}
```

For example:

```
rbd --cluster local mirror image demote image-pool/image-1
```

To demote all primary images within a pool to non-primary with `rbd`, specify the `mirror pool demote` command along with the pool name:

```
rbd mirror pool demote {pool-name}
```

For example:

```
rbd --cluster local mirror pool demote image-pool
```

To promote a specific image to primary with `rbd`, specify the `mirror image promote` command along with the pool and image name:

```
rbd mirror image promote [--force] {pool-name}/{image-name}
```

For example:

```
rbd --cluster remote mirror image promote image-pool/image-1
```

To promote all non-primary images within a pool to primary with `rbd`, specify the `mirror pool promote` command along with the pool name:

```
rbd mirror pool promote [--force] {pool-name}
```

For example:

```
rbd --cluster local mirror pool promote image-pool
```

Tip: Since the primary / non-primary status is per-image, it is possible to have two clusters split the IO load and stage failover / failback.

Note: Promotion can be forced using the `--force` option. Forced promotion is needed when the demotion cannot be propagated to the peer Ceph cluster (e.g. Ceph cluster failure, communication outage). This will result in a split-brain scenario between the two peers and the image will no longer be in-sync until a force resync command is issued.

FORCE IMAGE RESYNC

If a split-brain event is detected by the `rbd-mirror` daemon, it will not attempt to mirror the affected image until corrected. To resume mirroring for an image, first demote the image determined to be out-of-date and then request a resync to the primary image. To request an image resync with `rbd`, specify the `mirror image resync` command along with the pool and image name:

```
rbd mirror image resync {pool-name}/{image-name}
```

For example:

```
rbd mirror image resync image-pool/image-1
```

Note: The `rbd` command only flags the image as requiring a resync. The local cluster's `rbd-mirror` daemon process is responsible for performing the resync asynchronously.

MIRROR STATUS

The peer cluster replication status is stored for every primary mirrored image. This status can be retrieved using the `mirror image status` and `mirror pool status` commands.

To request the mirror image status with `rbd`, specify the `mirror image status` command along with the pool and image name:

```
rbd mirror image status {pool-name}/{image-name}
```

For example:

```
rbd mirror image status image-pool/image-1
```

To request the mirror pool summary status with `rbd`, specify the `mirror pool status` command along with the pool name:

```
rbd mirror pool status {pool-name}
```

For example:

```
rbd mirror pool status image-pool
```

Note: Adding `--verbose` option to the `mirror pool status` command will additionally output status details for every mirroring image in the pool.

RBD-MIRROR DAEMON

The two `rbd-mirror` daemons are responsible for watching image journals on the remote, peer cluster and replaying the journal events against the local cluster. The RBD image journaling feature records all modifications to the image in the order they occur. This ensures that a crash-consistent mirror of the remote image is available locally.

The `rbd-mirror` daemon is available within the optional `rbd-mirror` distribution package.

Important: Each `rbd-mirror` daemon requires the ability to connect to both clusters simultaneously.

Warning: Pre-Luminous releases: only run a single `rbd-mirror` daemon per Ceph cluster.

Each `rbd-mirror` daemon should use a unique Ceph user ID. To create a Ceph user, with ceph specify the `auth get-or-create` command, user name, monitor caps, and OSD caps:

```
ceph auth get-or-create client.rbd-mirror.{unique id} mon 'profile rbd' osd 'profile rbd'
```

The `rbd-mirror` daemon can be managed by `systemd` by specifying the user ID as the daemon instance:

```
systemctl enable ceph-rbd-mirror@rbd-mirror.{unique id}
```

The `rbd-mirror` can also be run in foreground by `rbd-mirror` command:

```
rbd-mirror -f --log-file={log_path}
```