Red Hat Ceph Storage 5

문제 해결 가이드

Red Hat Ceph Storage 문제 해결

Last Updated: 2022-08-14

Red Hat Ceph Storage 5 문제 해결 가이드

Red Hat Ceph Storage 문제 해결

Enter your first name here. Enter your surname here.
Enter your organisation's name here. Enter your organisational division here.
Enter your email address here.

법적 공지

초록

이 문서에서는 Red Hat Ceph Storage에서 일반적인 문제를 해결하는 방법을 설명합니다. Red Hat은 코드, 문서, 웹 속성에서 문제가 있는 용어를 교체하기 위해 최선을 다하고 있습니다. 먼저 마스터(master), 슬레이브(slave), 블랙리스트(blacklist), 화이트리스트(whitelist) 등 네 가지 용어를 교체하고 있습니다. 이러한 변경 작업은 작업 범위가 크므로 향후 여러 릴리스에 걸쳐 점차 구현할 예정입니다. 자세한 내용은 CTO Chris Wright의 메시지에서 참조하십시오.

차례

1장. 초기 문제 해결

스토리지 관리자는 Red Hat 지원에 문의하기 전에 Red Hat Ceph Storage 클러스터의 초기 문제 해결을 수행할 수 있습니다. 이 장에는 다음 정보가 포함됩니다.

- *문제 식별.*
- *Ceph 상태 이해.*
- *Ceph 클러스터에 대한 상태 변경 경고.*
- *Ceph 로그 이해.*
- *'sos 보고서 생성'을 생성합니다.*

1.1. 사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.

1.2. 문제 식별

Red Hat Ceph Storage 클러스터의 오류 원인을 확인하려면 Procedure 섹션에 있는 질문에 대답하십시오.

사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.

절차

1. 지원되지 않는 구성을 사용할 때 특정 문제가 발생할 수 있습니다. 구성이 지원되는지 확인합니다.
2. 이 문제의 원인이 되는 Ceph 구성 요소를 알고 있습니까?
   a. 아니요. *Red Hat Ceph Storage 문제 해결 가이드에서 Ceph 스토리지 클러스터 절차의 상태를 진단합니다.*
   b. Ceph 모니터. *Red Hat Ceph Storage 문제 해결 가이드의 Ceph 모니터 문제 해결 섹션을 참조하십시오.*
   c. Ceph OSD. *Red Hat Ceph Storage 문제 해결 가이드의 Ceph OSD 문제 해결 섹션을 참조하십시오.*
   d. Ceph 배치 그룹. *Red Hat Ceph Storage 문제 해결 가이드의 Ceph 배치 그룹 문제 해결 섹션을 참조하십시오.*
   e. 다중 사이트 Ceph 오브젝트 게이트웨이. *Red Hat Ceph Storage 문제 해결 가이드의 다중 사이트 Ceph 오브젝트 게이트웨이 문제 해결 섹션을 참조하십시오.*

추가 리소스

- 자세한 내용은 Red Hat Ceph Storage: 지원 구성 문서를 참조하십시오.

1.3. 스토리지 클러스터 상태 진단

다음 절차에서는 Red Hat Ceph Storage 클러스터의 상태를 진단하는 기본 단계를 나열합니다.

사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.

절차

1. Cephadm 쉘에 로그인합니다.

   예제

   ```
   [root@host01 ~]# cephadm shell
   ```

2. 스토리지 클러스터의 전체 상태를 확인합니다.

   예제

   ```
   [ceph: root@host01 /]# ceph health detail
   ```

   명령에서 **HEALTH_WARN** 또는 **HEALTH_ERR** 을 반환하는 경우 자세한 *내용은 Ceph 상태* 이해를 참조하십시오.

3. 스토리지 클러스터의 로그를 모니터링합니다.

   예제

   ```
   [ceph: root@host01 /]# ceph -W cephadm
   ```

4. 클러스터의 로그를 파일에 캡처하려면 다음 명령을 실행합니다.

   예제

   ```
   [ceph: root@host01 /]# ceph config set global log_to_file true
   [ceph: root@host01 /]# ceph config set global mon_cluster_log_to_file true
   ```

   로그는 기본적으로 **/var/log/ceph/** 디렉터리에 있습니다. Ceph 로그 이해에 나열된 오류 메시지의 *Ceph 로그를* 확인하십시오.

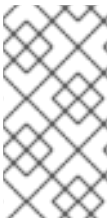
5. 로그에 충분한 정보가 포함되어 있지 않으면 디버깅 수준을 늘리고 실패한 작업을 재현하십시오. 자세한 *내용은 로깅* 구성을 참조하십시오.

1.4. CEPH 상태 이해

ceph health 명령은 Red Hat Ceph Storage 클러스터의 상태에 대한 정보를 반환합니다.

- **HEALTH_OK** 는 클러스터가 정상임을 나타냅니다.
- **HEALTH_WARN** 은 경고를 나타냅니다. 경우에 따라 Ceph 상태가 **HEALTH_OK** 로 자동 돌아갑니다. 예를 들어 Red Hat Ceph Storage 클러스터가 리밸런싱 프로세스를 완료하는 경우. 그러나 클러스터가 **HEALTH_WARN** 상태에 있는 경우 더 긴 문제 해결을 고려하십시오.
- **HEALTH_ERR** 은 즉각적인 주의가 필요한 더 심각한 문제를 나타냅니다.

ceph 상태 세부 정보 및 **ceph -s** 명령을 사용하여 자세한 출력을 가져옵니다.



참고

mgr 데몬이 실행되지 않으면 상태 경고가 표시됩니다. Red Hat Ceph Storage 클러스터의 마지막 **mgr** 데몬이 제거된 경우 Red Hat Storage 클러스터의 임의의 호스트에 **mgr** 데몬을 수동으로 배포할 수 있습니다. *Red Hat Ceph Storage 5 관리 가이드에서 mgr 데몬 수동 배포를 참조하십시오.*

추가 리소스

- *Red Hat Ceph Storage 문제 해결 가이드의 Ceph Monitor 오류 메시지 표를 참조하십시오.*
- *Red Hat Ceph Storage 문제 해결 가이드의 Ceph OSD 오류 메시지 표를 참조하십시오.*
- *Red Hat Ceph Storage 문제 해결 가이드의 배치 그룹 오류 메시지 표를 참조하십시오.*

1.5. CEPH 클러스터의 상태 변경 경고

특정 시나리오에서는 사용자가 경고를 이미 인식하고 즉시 조치를 취할 수 없기 때문에 일시적으로 일부 경고를 음소거할 수 있습니다. Ceph 클러스터의 전반적인 보고된 상태에 영향을 미치지 않도록 상태 점검을 음소거할 수 있습니다.

경고는 상태 점검 코드를 사용하여 지정됩니다. 한 가지 예로 OSD가 유지 관리를 위해 다운되면 **OSD_DOWN** 경고가 예상됩니다. 경고가 전체 유지 관리 기간 동안 **HEALTH_OK** 대신 cluster를 **HEALTH_WARN** 에 배치하기 때문에 유지보수가 끝날 때까지 경고를 음소거할 수 있습니다.

경고의 범위가 더 심해지면 대부분의 건강 마케터도 사라집니다. 예를 들어 OSD가 1개이고 경고가 음소거되면 하나 이상의 추가 OSD가 다운되면 음소거가 사라집니다. 이는 경고 또는 오류를 트리거하는 항목 수를 나타내는 개수와 관련된 모든 상태 경고에 적용됩니다.

사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.
- 노드에 대한 root 수준의 액세스입니다.
- 상태 경고 메시지입니다.

절차

1. Cephadm 쉘에 로그인합니다.

   예제

   ```
   [root@host01 ~]# cephadm shell
   ```

2. **ceph health detail** 명령을 실행하여 Red Hat Ceph Storage 클러스터의 상태를 확인합니다.

   예제

   ```
   [ceph: root@host01 /]# ceph health detail

   HEALTH_WARN 1 osds down; 1 OSDs or CRUSH {nodes, device-classes} have
   ```

```
{NOUP,NODOWN,NOIN,NOOUT} flags set
[WRN] OSD_DOWN: 1 osds down
    osd.1 (root=default,host=host01) is down
[WRN] OSD_FLAGS: 1 OSDs or CRUSH {nodes, device-classes} have
{NOUP,NODOWN,NOIN,NOOUT} flags set
    osd.1 has flags noup
```

OSD 중 하나가 down이므로 스토리지 클러스터가 **HEALTH_WARN** 상태에 있음을 알 수 있습니다.

3. 경고를 만듭니다.

구문

```
ceph health mute HEALTH_MESSAGE
```

예제

```
[ceph: root@host01 /]# ceph health mute OSD_DOWN
```

4. 선택 사항: 상태 점검 음소거에는 연결된 TTL(Time to Live)이 있을 수 있으므로 지정된 기간이 경과된 후 음소거가 자동으로 만료됩니다. 명령에서 TTL을 선택적 기간 인수로 지정합니다.

구문

```
ceph health mute HEALTH_MESSAGE DURATION
```

DURATION 은 **s**,**sec**,**m**,**min**,**h** 또는 **hour** 로 지정할 수 있습니다.

예제

```
[ceph: root@host01 /]# ceph health mute OSD_DOWN 10m
```

이 예에서 경고 **OSD_DOWN** 은 10분 동안 변경되지 않습니다.

5. Red Hat Ceph Storage 클러스터 상태가 **HEALTH_OK** 로 변경되었는지 확인합니다.

예제

```
[ceph: root@host01 /]# ceph -s
  cluster:
    id:     81a4597a-b711-11eb-8cb8-001a4a000740
    health: HEALTH_OK
            (muted: OSD_DOWN(9m) OSD_FLAGS(9m))

  services:
    mon: 3 daemons, quorum host01,host02,host03 (age 33h)
    mgr: host01.pzhfuh(active, since 33h), standbys: host02.wsnngf, host03.xwzphg
    osd: 11 osds: 10 up (since 4m), 11 in (since 5d)

  data:
    pools:   1 pools, 1 pgs
```

```
    objects: 13 objects, 0 B
    usage:   85 MiB used, 165 GiB / 165 GiB avail
    pgs:     1 active+clean
```

이 예에서는 *OSD_DOWN* 및 *OSD_FLAG* 가 음소거되고 9분 동안 mute가 활성화되어 있음을 확인할 수 있습니다.

6. 선택 사항: 경고가 지워지는 후에도 음소거를 유지할 수 **있습니다**.

구문

```
ceph health mute HEALTH_MESSAGE DURATION --sticky
```

예제

```
[ceph: root@host01 /]# ceph health mute OSD_DOWN 1h --sticky
```

7. 다음 명령을 실행하여 mute를 제거할 수 있습니다.

구문

```
ceph health unmute HEALTH_MESSAGE
```

예제

```
[ceph: root@host01 /]# ceph health unmute OSD_DOWN
```

추가 리소스

- 자세한 내용은 *Red Hat Ceph Storage 문제 해결 가이드에서 Ceph 클러스터의 상태 메시지를* 참조하십시오.

1.6. CEPH 로그 이해

Ceph는 로깅이 활성화된 후 **/var/log/ceph/** 디렉터리에 로그를 저장합니다.

***CLUSTER_NAME*.log** 는 글로벌 이벤트가 포함된 기본 스토리지 클러스터 로그 파일입니다. 기본적으로 로그 파일 이름은 **ceph.log** 입니다. Ceph Monitor 노드만 기본 스토리지 클러스터 로그가 포함됩니다.

각 Ceph OSD 및 모니터에는 ***CLUSTER_NAME*-osd.*NUMBER*.log 및 *CLUSTER_NAME*-mon.*HOSTNAME*.log** 라는 자체 로그 파일이 있습니다.

Ceph 하위 시스템의 디버깅 수준을 늘리면 Ceph도 해당 하위 시스템에 대한 새 로그 파일을 생성합니다.

추가 리소스

- 로깅에 대한 자세한 내용은 *Red Hat Ceph Storage 문제 해결 가이드 구성을 참조하십시오* . *https://access.redhat.com/documentation/en-us/red_hat_ceph_storage/5/html-single/troubleshooting_guide/#configuring-logging*
- *Red Hat Ceph Storage 문제 해결 가이드의 Ceph 로그 테이블의 Common Ceph Monitor 오류 메시지를 참조하십시오.*

- Red Hat Ceph Storage 문제 해결 가이드의 Ceph 로그 테이블에 있는 Common Ceph OSD 오류 메시지를 참조하십시오.

1.7. SOS 보고서 생성

sos report 명령을 실행하여 Red Hat Enterprise Linux에서 Red Hat Ceph Storage 클러스터의 구성 세부 정보, 시스템 정보 및 진단 정보를 수집할 수 있습니다. Red Hat 지원 팀은 이 정보를 사용하여 스토리지 클러스터의 추가 문제 해결을 위해 사용합니다.

사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.
- 노드에 대한 루트 수준 액세스입니다.

절차

1. **sos** 패키지를 설치합니다.

   예제

   ```
   [root@host01 ~]# dnf install sos
   ```

   참고

   sos-4.0.11.el8 패키지 이상 버전을 설치하여 Ceph 명령 출력을 올바르게 캡처합니다.

2. **sos 보고서를** 실행하여 스토리지 클러스터의 시스템 정보를 가져옵니다.

   예제

   ```
   [root@host01 ~]# sosreport -a --all-logs -e ceph
   ```

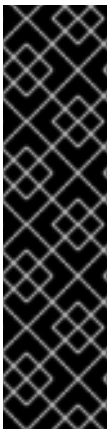
   보고서는 **/var/tmp** 파일에 저장됩니다.

추가 리소스

- What is an sosreport and how to create one in Red Hat Enterprise Linux? 자세한 내용은 Knowledgebase 문서입니다.

2장. 로깅 구성

이 장에서는 다양한 Ceph 하위 시스템에 대한 로깅을 구성하는 방법을 설명합니다.



중요

로깅은 리소스 집약적입니다. 또한 상세 로깅은 상대적으로 짧은 시간에 많은 양의 데이터를 생성할 수 있습니다. 클러스터의 특정 하위 시스템에서 문제가 발생하면 해당 하위 시스템의 로깅만 활성화합니다. 자세한 내용은 2.2절. "Ceph 하위 시스템" 를 참조하십시오.

또한 로그 파일의 회전을 설정하는 것이 좋습니다. 자세한 내용은 2.5절. "로그 교체 가속화" 를 참조하십시오.

문제가 발생하면 하위 시스템 로그 및 메모리 수준을 기본값으로 변경합니다. 모든 Ceph 하위 시스템 목록과 해당 기본값을 보려면 부록 A. Ceph 하위 시스템의 기본 로깅 수준 값 을 참조하십시오.

다음을 통해 Ceph 로깅을 구성할 수 있습니다.

- 런타임 시 **ceph** 명령 사용. 이것이 가장 일반적인 접근법입니다. 자세한 내용은 2.3절. "런타임 시 로깅 구성" 를 참조하십시오.
- Ceph 구성 파일 업데이트. 클러스터를 시작할 때 문제가 발생하는 경우 이 방법을 사용합니다. 자세한 내용은 2.4절. "구성 파일에서 로깅 구성" 를 참조하십시오.

2.1. 사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.

2.2. CEPH 하위 시스템

이 섹션에는 Ceph 하위 시스템 및 로깅 수준에 대한 정보가 포함되어 있습니다.

Ceph 하위 시스템 및 로깅 수준 이해

Ceph는 여러 하위 시스템으로 구성됩니다.

각 하위 시스템에는 다음과 같은 로깅 수준이 있습니다.

- **/var/log/ceph/** 디렉토리(log 수준)에 기본적으로 저장된 출력 로그
- 메모리 캐시에 저장된 로그(메모리 수준)

일반적으로 Ceph는 다음을 제외하고 메모리에 저장된 로그를 출력 로그에 보내지 않습니다.

- 치명적인 신호가 발생합니다.
- 소스 코드에서 어설션이 트리거됩니다.
- 요청하신 경우

이러한 각 하위 시스템에 대해 서로 다른 값을 설정할 수 있습니다. Ceph 로깅 수준은 1에서 **20** 까지이며 **1** 은 terse이고 **20** 은 자세한 정보를 제공합니다.

로그 수준과 메모리 수준에 단일 값을 사용하여 둘 다 동일한 값으로 설정합니다. 예를 들어 ***debug_osd = 5*** *는* ***ceph-osd*** *데몬의 디버그 수준을* ***5*** *로 설정합니다.*

출력 로그 수준과 메모리 수준에 다른 값을 사용하려면 슬래시(/)로 값을 분리합니다. 예를 들어 ***debug_mon = 1/5*** *는* ***ceph-mon*** *데몬의 디버그 로그 수준을* ***1*** *로 설정하고 메모리 로그 수준을* ***5*** *로 설정합니다.*

표 2.1. Ceph subsystems 및 Logging Defaults

하위 시스템	로그 수준	메모리 수준	설명
asok	1	5	관리 소켓
auth	1	5	인증
클라이언트	0	5	**librados** 를 사용하여 클러스터에 연결하는 모든 애플리케이션 또는 라이브러리
bluestore	1	5	BlueStore OSD 백엔드
journal	1	5	OSD 저널
mds	1	5	메타데이터 서버
monc	0	5	모니터 클라이언트는 대부분의 Ceph 데몬과 모니터 간의 통신을 처리합니다.
월요일	1	5	모니터
ms	0	5	Ceph 구성 요소 간 메시징 시스템
osd	0	5	OSD 데몬
Paxos	0	5	합의를 설정하는 데 사용하는 알고리즘
rados	0	5	신뢰할 수 있는 자동 배포 오브젝트 스토어, Ceph의 핵심 구성 요소
rbd	0	5	Ceph 블록 장치
rgw	1	5	Ceph Object Gateway

로그 출력 예

다음 예제에서는 모니터와 OSD의 상세 수준을 높일 때 로그에서 메시지 유형을 보여줍니다.

디버그 설정 모니터링

```
debug_ms = 5
debug_mon = 20
```

```
debug_paxos = 20
debug_auth = 20
```

모니터 디버그 설정의 로그 출력 예

```
2022-05-12 12:37:04.278761 7f45a9afc700 10 mon.cephn2@0(leader).osd e322 e322: 2 osds: 2 up,
2 in
2022-05-12 12:37:04.278792 7f45a9afc700 10 mon.cephn2@0(leader).osd e322
min_last_epoch_clean 322
2022-05-12 12:37:04.278795 7f45a9afc700 10 mon.cephn2@0(leader).log v1010106 log
2022-05-12 12:37:04.278799 7f45a9afc700 10 mon.cephn2@0(leader).auth v2877 auth
2022-05-12 12:37:04.278811 7f45a9afc700 20 mon.cephn2@0(leader) e1 sync_trim_providers
2022-05-12 12:37:09.278914 7f45a9afc700 11 mon.cephn2@0(leader) e1 tick
2022-05-12 12:37:09.278949 7f45a9afc700 10 mon.cephn2@0(leader).pg v8126 v8126: 64 pgs: 64
active+clean; 60168 kB data, 172 MB used, 20285 MB / 20457 MB avail
2022-05-12 12:37:09.278975 7f45a9afc700 10 mon.cephn2@0(leader).paxosservice(pgmap
7511..8126) maybe_trim trim_to 7626 would only trim 115 < paxos_service_trim_min 250
2022-05-12 12:37:09.278982 7f45a9afc700 10 mon.cephn2@0(leader).osd e322 e322: 2 osds: 2 up,
2 in
2022-05-12 12:37:09.278989 7f45a9afc700  5 mon.cephn2@0(leader).paxos(paxos active c
1028850..1029466) is_readable = 1 - now=2021-08-12 12:37:09.278990 lease_expire=0.000000 has
v0 lc 1029466
....
2022-05-12 12:59:18.769963 7f45a92fb700  1 -- 192.168.0.112:6789/0 <== osd.1
192.168.0.114:6800/2801 5724 ==== pg_stats(0 pgs tid 3045 v 0) v1 ==== 124+0+0 (2380105412 0
0) 0x5d96300 con 0x4d5bf40
2022-05-12 12:59:18.770053 7f45a92fb700  1 -- 192.168.0.112:6789/0 --> 192.168.0.114:6800/2801
-- pg_stats_ack(0 pgs tid 3045) v1 -- ?+0 0x550ae00 con 0x4d5bf40
2022-05-12 12:59:32.916397 7f45a9afc700  0 mon.cephn2@0(leader).data_health(1) update_stats
avail 53% total 1951 MB, used 780 MB, avail 1053 MB
....
2022-05-12 13:01:05.256263 7f45a92fb700  1 -- 192.168.0.112:6789/0 --> 192.168.0.113:6800/2410
-- mon_subscribe_ack(300s) v1 -- ?+0 0x4f283c0 con 0x4d5b440
```

OSD 디버그 설정

```
debug_ms = 5
debug_osd = 20
```

OSD 디버그 설정의 로그 출력 예

```
2022-05-12 11:27:53.869151 7f5d55d84700  1 -- 192.168.17.3:0/2410 --> 192.168.17.4:6801/2801 --
osd_ping(ping e322 stamp 2021-08-12 11:27:53.869147) v2 -- ?+0 0x63baa00 con 0x578dee0
2022-05-12 11:27:53.869214 7f5d55d84700  1 -- 192.168.17.3:0/2410 --> 192.168.0.114:6801/2801
-- osd_ping(ping e322 stamp 2021-08-12 11:27:53.869147) v2 -- ?+0 0x638f200 con 0x578e040
2022-05-12 11:27:53.870215 7f5d6359f700  1 -- 192.168.17.3:0/2410 <== osd.1
192.168.0.114:6801/2801 109210 ==== osd_ping(ping_reply e322 stamp 2021-08-12
11:27:53.869147) v2 ==== 47+0+0 (261193640 0 0) 0x63c1a00 con 0x578e040
2022-05-12 11:27:53.870698 7f5d6359f700  1 -- 192.168.17.3:0/2410 <== osd.1
192.168.17.4:6801/2801 109210 ==== osd_ping(ping_reply e322 stamp 2021-08-12
11:27:53.869147) v2 ==== 47+0+0 (261193640 0 0) 0x6313200 con 0x578dee0
....
2022-05-12 11:28:10.432313 7f5d6e71f700  5 osd.0 322 tick
2022-05-12 11:28:10.432375 7f5d6e71f700 20 osd.0 322 scrub_random_backoff lost coin flip,
```

```
randomly backing off
2022-05-12 11:28:10.432381 7f5d6e71f700 10 osd.0 322 do_waiters -- start
2022-05-12 11:28:10.432383 7f5d6e71f700 10 osd.0 322 do_waiters -- finish
```

추가 리소스

- 런타임 시 로깅 구성
- 구성 파일에서 로깅 구성

2.3. 런타임 시 로깅 구성

시스템 런타임에서 Ceph 하위 시스템의 로깅을 구성하여 발생할 수 있는 모든 문제를 해결할 수 있습니다.

사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.
- Ceph 디버거에 액세스합니다.

절차

1. Ceph 디버깅 출력을 활성화하려면 **dout()** 를 런타임 시 다음을 수행합니다.

```
ceph tell TYPE.ID injectargs --debug-SUBSYSTEM VALUE [--NAME VALUE]
```

2. 교체:
   - **TYPE** with the type of Ceph daemon (**osd**,**mon**, or **mds**)
   - Ceph 데몬의 특정 **ID** 가 있는 ID입니다. 또는 ***** 를 사용하여 특정 유형의 모든 데몬에 런타임 설정을 적용합니다.
   - 특정 하위 시스템을 사용하여 **제출**.
   - **VALUE 1** 에서 **20** 의 숫자가 있으며, 여기서 **1** 은 정점이고 **20** 은 상세 정보입니다.
     예를 들어, **osd.0** 이라는 OSD 하위 시스템에서 OSD 하위 시스템의 로그 수준을 0으로 설정하고 메모리 수준을 5로 설정하려면 다음을 수행합니다.

```
# ceph tell osd.0 injectargs --debug-osd 0/5
```

런타임 시 구성 설정을 보려면 다음을 수행합니다.

1. 실행 중인 Ceph 데몬(예: **ceph-osd** 또는 **ceph-mon**)을 사용하여 호스트에 로그인합니다.
2. 구성을 표시합니다.

   구문

```
ceph daemon NAME config show | less
```

   예제

-

```
[ceph: root@host01 /]# ceph daemon osd.0 config show | less
```

추가 리소스

- 자세한 내용은 Ceph 하위 시스템을 참조하십시오.
- 자세한 내용은 구성 파일의 구성 로깅 을 참조하십시오.
- Red Hat Ceph Storage 5 구성 가이드 의 Ceph 디버깅 및 로깅 구성 참조 장.

2.4. 구성 파일에서 로깅 구성

정보, 경고 및 오류 메시지를 로그 파일에 기록하도록 Ceph 하위 시스템을 구성합니다. Ceph 구성 파일의 디버깅 수준을 기본 **/etc/ceph/ceph.conf** 에서 지정할 수 있습니다.

사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.

절차

1. Ceph 디버깅 출력을 활성화하려면 부팅 시 **dout()** 를 Ceph 구성 파일에 디버깅 설정을 추가합니다.
   a. 각 데몬에 공통된 하위 시스템의 경우 **[global]** 섹션 아래에 설정을 추가합니다.
   b. 특정 데몬에 대한 하위 시스템의 경우 **[mon]**, **[osd]** 또는 **[mds]** 와 같은 데몬 섹션의 설정을 추가합니다.

   예제

```
[global]
        debug_ms = 1/5

[mon]
        debug_mon = 20
        debug_paxos = 1/5
        debug_auth = 2

[osd]
        debug_osd = 1/5
        debug_monc = 5/20

[mds]
        debug_mds = 1
```

추가 리소스

- Ceph 하위 시스템
- 런타임 시 로깅 구성
- Red Hat Ceph Storage 5 구성 가이드 의 Ceph 디버깅 및 로깅 구성 참조 장

2.5. 로그 교체 가속화

Ceph 구성 요소에 대한 디버깅 수준을 늘리면 대량의 데이터가 생성될 수 있습니다. 거의 전체 디스크가 있는 경우 **/etc/logrotate.d/ceph** 에서 Ceph 로그 회전 파일을 수정하여 로그 회전을 가속화할 수 있습니다. Cron 작업 스케줄러는 이 파일을 사용하여 로그 회전을 예약합니다.

사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.
- 노드에 대한 루트 수준 액세스입니다.

절차

1. 회전 빈도 다음에 로그 회전 파일에 크기 설정을 추가합니다.

```
rotate 7
weekly
size SIZE
compress
sharedscripts
```

예를 들어 로그 파일을 500MB에 도달하면 회전하는 방법은 다음과 같습니다.

```
rotate 7
weekly
size 500 MB
compress
sharedscripts
size 500M
```

2. **crontab** 편집기를 엽니다.

```
[root@mon ~]# crontab -e
```

3. 항목을 추가하여 **/etc/logrotate.d/ceph** 파일을 확인합니다. 예를 들어, Cron에 30분마다 **/etc/logrotate.d/ceph** 를 확인하도록 지시하려면 다음을 수행합니다.

```
30 * * * * /usr/sbin/logrotate /etc/logrotate.d/ceph >/dev/null 2>&1
```

3장. 네트워킹 문제 해결

이 장에서는 네트워킹 및 chrony for Network Time Protocol(NTP)과 연결된 기본 문제 해결 절차를 설명합니다.

3.1. 사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.

3.2. 기본 네트워킹 문제 해결

Red Hat Ceph Storage는 안정적인 네트워크 연결에 크게 의존합니다. Red Hat Ceph Storage 노드는 네트워크를 사용하여 서로 통신합니다. 네트워킹 문제로 인해 Ceph OSD와 같은 여러 문제가 발생할 수 있습니다(예: 실행 중 또는 **down** 으로 잘못 보고되는 경우). 네트워킹 문제로 인해 Ceph Monitor의 시계가 정지하는 오류가 발생할 수 있습니다. 또한 패킷 손실, 대기 시간이 길거나 제한된 대역폭이 클러스터 성능 및 안정성에 영향을 줄 수 있습니다.

사전 요구 사항

- 노드에 대한 루트 수준 액세스입니다.

절차

1. **net-tools** 및 **telnet** 패키지를 설치하면 Ceph 스토리지 클러스터에서 발생할 수 있는 네트워크 문제를 해결할 때 도움이 될 수 있습니다.

   예제

   ```
   [root@host01 ~]# dnf install net-tools
   [root@host01 ~]# dnf install telnet
   ```

2. **cephadm** 쉘에 로그인하고 Ceph 구성 파일의 **public_network** 매개변수에 올바른 값이 포함되어 있는지 확인합니다.

   예제

   ```
   [ceph: root@host01 /]# cat /etc/ceph/ceph.conf
   # minimal ceph.conf for 57bddb48-ee04-11eb-9962-001a4a000672
   [global]
    fsid = 57bddb48-ee04-11eb-9962-001a4a000672
    mon_host = [v2:10.74.249.26:3300/0,v1:10.74.249.26:6789/0]
   [v2:10.74.249.163:3300/0,v1:10.74.249.163:6789/0]
   [v2:10.74.254.129:3300/0,v1:10.74.254.129:6789/0]
   [mon.host01]
   public network = 10.74.248.0/21
   ```

3. 쉘을 종료하고 네트워크 인터페이스가 시작되었는지 확인합니다.

   예제

   ```
   [root@host01 ~]# ip link list
   1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN mode
   DEFAULT group default qlen 1000
   ```

```
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
2: ens3: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP mode
DEFAULT group default qlen 1000
    link/ether 00:1a:4a:00:06:72 brd ff:ff:ff:ff:ff:ff
```

4. *짧은 호스트 이름을 사용하여 Ceph 노드가 서로 연결할 수 있는지 확인합니다. 스토리지 클러스터의 각 노드에서 이 값을 확인합니다.*

   구문

   ```
   ping SHORT_HOST_NAME
   ```

   예제

   ```
   [root@host01 ~]# ping host02
   ```

5. *방화벽을 사용하는 경우 Ceph 노드가 적절한 포트에서 서로 연결할 수 있는지 확인합니다.* ***firewall-cmd*** *및* ***telnet*** *툴은 포트 상태를 확인하고 포트가 각각 열려 있는 경우 다음을 수행합니다.*

   구문

   ```
   firewall-cmd --info-zone=ZONE
   telnet IP_ADDRESS PORT
   ```

   예제

   ```
   [root@host01 ~]# firewall-cmd --info-zone=public
   public (active)
     target: default
     icmp-block-inversion: no
     interfaces: ens3
     sources:
     services: ceph ceph-mon cockpit dhcpv6-client ssh
     ports: 9283/tcp 8443/tcp 9093/tcp 9094/tcp 3000/tcp 9100/tcp 9095/tcp
     protocols:
     masquerade: no
     forward-ports:
     source-ports:
     icmp-blocks:
     rich rules:

   [root@host01 ~]# telnet 192.168.0.22 9100
   ```

6. *인터페이스 카운터에 오류가 없는지 확인합니다. 노드 간 네트워크 연결 대기 시간이 예상되었으며 패킷 손실이 없는지 확인합니다.*

   a. ***ethtool*** *명령 사용:*

      구문

      ```
      ethtool -S INTERFACE
      ```

예제

```
[root@host01 ~]# ethtool -S ens3 | grep errors
NIC statistics:
     rx_fcs_errors: 0
     rx_align_errors: 0
     rx_frame_too_long_errors: 0
     rx_in_length_errors: 0
     rx_out_length_errors: 0
     tx_mac_errors: 0
     tx_carrier_sense_errors: 0
     tx_errors: 0
     rx_errors: 0
```

b. ***ifconfig*** *명령 사용:*

예제

```
[root@host01 ~]# ifconfig
ens3: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 10.74.249.26  netmask 255.255.248.0  broadcast 10.74.255.255
        inet6 fe80::21a:4aff:fe00:672  prefixlen 64  scopeid 0x20<link>
        inet6 2620:52:0:4af8:21a:4aff:fe00:672  prefixlen 64  scopeid 0x0<global>
        ether 00:1a:4a:00:06:72  txqueuelen 1000  (Ethernet)
        RX packets 150549316  bytes 56759897541 (52.8 GiB)
        RX errors 0  dropped 176924  overruns 0  frame 0
        TX packets 55584046  bytes 62111365424 (57.8 GiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING>  mtu 65536
        inet 127.0.0.1  netmask 255.0.0.0
        inet6 ::1  prefixlen 128  scopeid 0x10<host>
        loop  txqueuelen 1000  (Local Loopback)
        RX packets 9373290  bytes 16044697815 (14.9 GiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 9373290  bytes 16044697815 (14.9 GiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

c. ***netstat*** *명령 사용:*

예제

```
[root@host01 ~]# netstat -ai
Kernel Interface table
Iface             MTU    RX-OK RX-ERR RX-DRP RX-OVR    TX-OK TX-ERR TX-DRP TX-
OVR Flg
ens3             1500 311847720      0 364903 0      114341918      0      0      0 BMRU
lo              65536 19577001       0      0 0       19577001       0      0      0 LRU
```

7. *성능 문제의 경우 대기 시간을 확인하고 스토리지 클러스터의 모든 노드 간 네트워크 대역폭을 확인하고* ***iperf3*** *툴을 사용합니다.* ***iperf3*** *도구는 서버와 클라이언트 간의 간단한 점-투-포인트 네트워크 대역폭 테스트를 수행합니다.*

   a. *대역폭을 확인하려는 Red Hat Ceph Storage 노드에* ***iperf3*** *패키지를 설치합니다.*

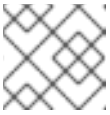
예제

```
[root@host01 ~]# dnf install iperf3
```

b. *Red Hat Ceph Storage 노드에서* ***iperf3*** *서버를 시작합니다.*

예제

```
[root@host01 ~]# iperf3 -s
-----------------------------------------------------------
Server listening on 5201
-----------------------------------------------------------
```



참고

기본 포트는 5201이지만 ***-P*** *명령 인수를 사용하여 설정할 수 있습니다.*

c. *다른 Red Hat Ceph Storage 노드에서* ***iperf3*** *클라이언트를 시작합니다.*

예제

```
[root@host02 ~]# iperf3 -c mon
Connecting to host mon, port 5201
[  4] local xx.x.xxx.xx port 52270 connected to xx.x.xxx.xx port 5201
[ ID] Interval           Transfer     Bandwidth       Retr  Cwnd
[  4]   0.00-1.00   sec   114 MBytes   954 Mbits/sec    0    409 KBytes
[  4]   1.00-2.00   sec   113 MBytes   945 Mbits/sec    0    409 KBytes
[  4]   2.00-3.00   sec   112 MBytes   943 Mbits/sec    0    454 KBytes
[  4]   3.00-4.00   sec   112 MBytes   941 Mbits/sec    0    471 KBytes
[  4]   4.00-5.00   sec   112 MBytes   940 Mbits/sec    0    471 KBytes
[  4]   5.00-6.00   sec   113 MBytes   945 Mbits/sec    0    471 KBytes
[  4]   6.00-7.00   sec   112 MBytes   937 Mbits/sec    0    488 KBytes
[  4]   7.00-8.00   sec   113 MBytes   947 Mbits/sec    0    520 KBytes
[  4]   8.00-9.00   sec   112 MBytes   939 Mbits/sec    0    520 KBytes
[  4]   9.00-10.00  sec   112 MBytes   939 Mbits/sec    0    520 KBytes
- - - - - - - - - - - - - - - - - - - - - - - - -
[ ID] Interval           Transfer     Bandwidth       Retr
[  4]   0.00-10.00  sec  1.10 GBytes   943 Mbits/sec    0             sender
[  4]   0.00-10.00  sec  1.10 GBytes   941 Mbits/sec                  receiver

iperf Done.
```

이 출력은 테스트 중에 다시 전송(후부) 하지 않고 Red Hat Ceph Storage 노드 간 네트워크 대역폭 1.1 Gbits/second 를 보여줍니다.

스토리지 클러스터에 있는 모든 노드 간에 네트워크 대역폭을 확인하는 것이 좋습니다.

8.

모든 노드에 동일한 네트워크 상호 연결 속도가 있는지 확인합니다. 연결된 노드의 속도가 느려지면 더 빠르게 연결된 노드가 느려질 수 있습니다. 또한 상호 스위치 링크가 연결된 노드의 집

계된 대역폭을 처리할 수 있는지 확인합니다.

구문

```
ethtool INTERFACE
```

예제

```
[root@host01 ~]# ethtool ens3
Settings for ens3:
Supported ports: [ TP ]
Supported link modes:   10baseT/Half 10baseT/Full
                        100baseT/Half 100baseT/Full
                        1000baseT/Half 1000baseT/Full
Supported pause frame use: No
Supports auto-negotiation: Yes
Supported FEC modes: Not reported
Advertised link modes:  10baseT/Half 10baseT/Full
                        100baseT/Half 100baseT/Full
                        1000baseT/Half 1000baseT/Full
Advertised pause frame use: Symmetric
Advertised auto-negotiation: Yes
Advertised FEC modes: Not reported
Link partner advertised link modes:  10baseT/Half 10baseT/Full
                                     100baseT/Half 100baseT/Full
                                     1000baseT/Full
Link partner advertised pause frame use: Symmetric
Link partner advertised auto-negotiation: Yes
Link partner advertised FEC modes: Not reported
Speed: 1000Mb/s (1)
Duplex: Full (2)
Port: Twisted Pair
PHYAD: 1
Transceiver: internal
Auto-negotiation: on
MDI-X: off
Supports Wake-on: g
Wake-on: d
Current message level: 0x000000ff (255)
       drv probe link timer ifdown ifup rx_err tx_err
Link detected: yes (3)
```

추가 리소스

- 자세한 내용은 고객 포털의 기본 네트워크 문제 해결 솔루션을 참조하십시오.
- "ethtool" 명령이란 무엇이며 어떻게 네트워크 장치 및 인터페이스에 대한 정보를 얻을 수 있는지 확인하십시오.
- 자세한 내용은 RHEL 네트워크 인터페이스에서 고객 포털에서 패킷 솔루션을 삭제하십시오.
- 자세한 내용은 고객 포털의 Red Hat Ceph Storage에서 사용할 수 있는 성능 벤치마킹 툴이란 무엇입니까?
- 자세한 내용은 고객 포털에서 네트워킹 문제 해결과 관련된 지식베이스 문서 및 솔루션을 참조하십시오.

3.3. 기본 CHRONY NTP 문제 해결

이 섹션에는 기본 chrony NTP 문제 해결 단계가 포함되어 있습니다.

사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.
- Ceph Monitor 노드에 대한 루트 수준 액세스입니다.

절차

1. chronyd 데몬이 Ceph Monitor 호스트에서 실행 중인지 확인합니다.

예제

```
[root@mon ~]# systemctl status chronyd
```

2. **chronyd** 가 실행 중이 아닌 경우 다음을 활성화하고 시작합니다.

예제

```
[root@mon ~]# systemctl enable chronyd
[root@mon ~]# systemctl start chronyd
```

3. **chronyd** 가 클럭을 올바르게 동기화하는지 확인합니다.

예제

```
[root@mon ~]# chronyc sources
[root@mon ~]# chronyc sourcestats
[root@mon ~]# chronyc tracking
```

추가 리소스

- 고급 **chrony NTP** 문제 해결 단계를 위해 **Red Hat Customer Portal**에서 **chrony** 문제 해결 방법을 참조하십시오.
- 자세한 내용은 **Red Hat Ceph Storage Troubleshooting Guide** 의 **Clock skew** 섹션을 참조하십시오.
- 자세한 내용은 **chrony**가 동기화되었는지 확인 섹션을 참조하십시오.

4장. CEPH 모니터 문제 해결

이 장에서는 Ceph 모니터와 관련된 가장 일반적인 오류를 수정하는 방법을 설명합니다.

4.1. 사전 요구 사항

- 네트워크 연결을 확인합니다.

4.2. 대부분의 CEPH 모니터 오류

다음 표에는 ceph health detail 명령에서 반환되거나 Ceph 로그에 포함된 가장 일반적인 오류 메시지가 나열되어 있습니다. 표에서는 오류를 설명하고 문제를 해결하기 위한 특정 절차를 가리키는 해당 섹션에 대한 링크를 제공합니다.

4.2.1. 사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.

4.2.2. Ceph Monitor 오류 메시지

일반적인 Ceph 모니터 오류 메시지 테이블 및 잠재적인 수정 사항.

오류 메시지	참조
HEALTH_WARN	
Mon.X가 다운됨(quorum 외)	Ceph Monitor가 쿼럼 없음
시계 스kew	시계 스kew
저장소가 너무 커서 있습니다.	Ceph 모니터 저장소가 너무 큽니다.

4.2.3. Ceph 로그의 일반적인 Ceph Monitor 오류 메시지

Ceph 로그에 있는 일반적인 Ceph 모니터 오류 메시지 테이블 및 잠재적인 수정 사항에 대한 링크입니다.

오류 메시지	로그 파일	참조
시계 스kew	기본 클러스터 로그	시계 스kew
시계가 동기화되지 않음	기본 클러스터 로그	시계 스kew
손상: 기록 중 오류	로그 모니터링	Ceph Monitor가 쿼럼 없음 Ceph Monitor 저장소 복구
손상: 1개의 파일이 누락됨	로그 모니터링	Ceph Monitor가 쿼럼 없음 Ceph Monitor 저장소 복구
catch 신호(Bus 오류)	로그 모니터링	Ceph Monitor가 쿼럼 없음

4.2.4. Ceph Monitor가 쿼럼 없음

*하나 이상의 **Ceph** 모니터가 **down** 으로 표시되지만 다른 **Ceph** 모니터는 여전히 쿼럼을 구성할 수 있습니다. 또한 **ceph health detail** 명령은 다음과 유사한 오류 메시지를 반환합니다.*

```
HEALTH_WARN 1 mons down, quorum 1,2 mon.b,mon.c
mon.a (rank 0) addr 127.0.0.1:6789/0 is down (out of quorum)
```

<그 Means>

***Ceph**는 다양한 이유로 **Ceph** 모니터를 **down** 으로 표시합니다.*

***ceph-mon** 데몬이 실행 중이 아닌 경우 손상된 저장소가 있거나 다른 오류가 데몬을 시작하지 못하는 것입니다. 또한 **/var/** 파티션이 가득 차 있을 수 있습니다. 결과적으로 **ceph-mon** 은 기본적으로 **/var/lib/ceph/mon-SHORT_HOST_NAME/store.db** 에 있는 저장소에 작업을 수행할 수 없습니다.*

***ceph-mon** 데몬이 실행 중이지만 **Ceph Monitor**가 쿼럼 상태가 아닌 경우 문제의 원인은 **Ceph Monitor** 상태에 따라 달라집니다.*

- *Ceph 모니터가 예상보다 오래 걸리는 경우 다른 **Ceph** 모니터를 찾을 수 없습니다. 이 문제는 네트워킹 문제로 인해 발생하거나 **Ceph Monitor** 맵(**monmap**)이 있을 수 있으며 잘못된 **IP** 주소의 다른 **Ceph** 모니터에 도달하려고 할 수 있습니다. 또는 **monmap** 이 최신 상태인 경우 **Ceph** 모니터의 시계가 동기화되지 않을 수 있습니다.*
- *Ceph 모니터가 예상보다 오래 선택되는 경우 **Ceph** 모니터의 시계가 동기화되지 않을 수 있습니다.*

- *Ceph Monitor가 동기화 에서 선택 및 뒤로 상태를 변경하는 경우 클러스터 상태가 진행 중입니다. 즉, 동기화 프로세스에서 처리할 수 있는 것보다 새 맵을 더 빠르게 생성합니다.*

- *Ceph 모니터가 자체적으로 리더 또는 펩론으로 표시하는 경우, 나머지 클러스터에서 는 쿼럼이라고 생각되는 반면 나머지 클러스터는 그렇지 않은지 확인합니다. 이 문제는 실패한 클럭 동기화로 인해 발생할 수 있습니다.*

이 문제를 해결하기 위해

1. *ceph-mon 데몬이 실행 중인지 확인합니다. 그렇지 않은 경우 다음을 시작합니다.*

   구문

   ```
   systemctl status ceph-mon@HOST_NAME
   systemctl start ceph-mon@HOST_NAME
   ```

   HOST_NAME 을 데몬이 실행 중인 호스트의 짧은 이름으로 교체합니다. 확실하지 않은 경우 hostname -s 명령을 사용하십시오.

2. *ceph-mon 을 시작할 수 없는 경우 ceph-mon 데몬 의 단계를 따르십시오.*

3. *ceph-mon 데몬을 시작할 수 있지만 down 으로 표시된 경우 ceph-mon 데몬이 실행 중이지만 'down'으로 표시된 단계를 따르십시오.*

ceph-mon 데몬을 시작할 수 없습니다.

1. *기본적으로 /var/log/ceph/ceph-mon.HOST_NAME.log에 있는 해당 Ceph Monitor 로그 를 확인합니다.*

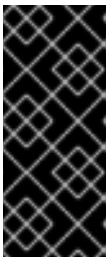
2. *로그에 다음 항목과 유사한 오류 메시지가 포함된 경우 Ceph Monitor에 손상된 저장소가 있을 수 있습니다.*

```
Corruption: error in middle of record
Corruption: 1 missing files; example: /var/lib/ceph/mon/mon.0/store.db/1234567.ldb
```

이 문제를 해결하려면 **Ceph Monitor**를 교체하십시오. 실패한 모니터 교체를 참조하십시오.

3. 로그에 다음과 유사한 오류 메시지가 포함된 경우 **/var/** 파티션이 꽉 찰 수 있습니다. **/var/** 에서 불필요한 데이터를 삭제합니다.

```
Caught signal (Bus error)
```



중요

Monitor 디렉토리에서 수동으로 데이터를 삭제하지 마십시오. 대신 **ceph-monstore-tool** 을 사용하여 압축합니다. 자세한 내용은 **Ceph Monitor** 저장소 업그레이드를 참조하십시오.

4. 다른 오류 메시지가 표시되면 지원 티켓을 엽니다. 자세한 내용은 **Red Hat** 지원 문의를 참조하십시오.

ceph-mon 데몬이 실행 중이지만 **down**으로 표시됩니다.

1. 쿼럼이 없는 **Ceph** 모니터 호스트에서 **mon_status** 명령을 사용하여 해당 상태를 확인합니다.

```
[root@mon ~]# ceph daemon ID mon_status
```

ID 를 **Ceph** 모니터의 **ID**로 바꿉니다. 예를 들면 다음과 같습니다.

```
[ceph: root@host01 /]# ceph daemon mon.host01 mon_status
```

2. 상태가 **probing** 이면 **mon_status** 출력에서 다른 **Ceph** 모니터의 위치를 확인합니다.

   a. 주소가 올바르지 않으면 **Ceph Monitor**에 잘못된 **Ceph Monitor** 맵**(monmap)**이 있습니다. 이 문제를 해결하려면 **Ceph** 모니터 맵 삽입을 참조하십시오.

   b. 주소가 올바르면 **Ceph Monitor** 시계가 동기화되었는지 확인합니다. 자세한 내용은 시계 스케이프를 참조하십시오. 또한 네트워킹 문제 해결을 참조하십시오. 자세한 내용은 네트

워킹 문제 해결을 참조하십시오.

3. *상태가 선택 되면 **Ceph Monitor** 시계가 동기화되었는지 확인합니다. 자세한 내용은 시계 스케이프를 참조하십시오.*

4. *동기화 선택에서 상태가 변경되면 지원 티켓을 엽니다. 자세한 내용은 **Red Hat** 지원 문의를 참조하십시오.*

5. ***Ceph** 모니터가 리더 또는 펩 론인 경우 **Ceph** 모니터 클럭이 동기화되었는지 확인합니다. 자세한 내용은 시계 스케이프를 참조하십시오. 시계를 동기화해도 문제가 해결되지 않는 경우 지원 티켓을 엽니다. 자세한 내용은 **Red Hat** 지원 문의를 참조하십시오.*

추가 리소스

- ***Ceph Monitor** 상태 이해를참조하십시오.*
- ***Red Hat Ceph Storage** 관리 가이드의 시작, 중지, 다시 시작 섹션 .*
- ***Red Hat Ceph Storage** 관리 가이드의 **Ceph** 관리소켓 사용 섹션 .*

4.2.5. 시계 스kew

***Ceph** 모니터는 쿼럼이 아니며 **ceph** 상태 세부 명령 출력에 다음과 같은 오류 메시지가 포함되어 있습니다.*

```
mon.a (rank 0) addr 127.0.0.1:6789/0 is down (out of quorum)
mon.a addr 127.0.0.1:6789/0 clock skew 0.08235s > max 0.05s (latency 0.0045s)
```

*또한 **Ceph** 로그에 다음과 유사한 오류 메시지가 포함됩니다.*

```
2022-05-04 07:28:32.035795 7f806062e700 0 log [WRN] : mon.a 127.0.0.1:6789/0 clock skew 0.14s
> max 0.05s
2022-05-04 04:31:25.773235 7f4997663700 0 log [WRN] : message from mon.1 was stamped
0.186257s in the future, clocks not synchronized
```

<그 Means>

클럭 skew 오류 메시지는 Ceph Monitor의 시계가 동기화되지 않았음을 나타냅니다. Ceph 모니터는 시간 정밀도에 따라 다르며 클럭이 동기화되지 않는 경우 예기치 않게 동작하기 때문에 클럭 동기화가 중요합니다.

mon_clock_drift_allowed 매개변수는 허용되는 클럭 간의 차이를 결정합니다. 기본적으로 이 매개 변수는 0.05초로 설정됩니다.

중요

이전 테스트 없이 mon_clock_drift_allowed 의 기본값을 변경하지 마십시오. 이 값을 변경하면 일반적으로 Ceph Monitor 및 Ceph Storage 클러스터의 안정성에 영향을 줄 수 있습니다.

시계 스kew 오류가 발생하면 구성된 경우 네트워크 문제 또는 chrony Network Time Protocol (NTP) 동기화의 문제가 있습니다. 또한 가상 머신에 배포된 Ceph Monitor에서 시간 동기화가 제대로 작동하지 않습니다.

이 문제를 해결하기 위해

1. 네트워크가 올바르게 작동하는지 확인합니다. 자세한 내용은 네트워킹 문제 해결을 참조하십시오. NTP에 chrony를 사용하는 경우 자세한 내용은 Basic chrony NTP 문제 해결 섹션을 참조하십시오.

2. 원격 NTP 서버를 사용하는 경우 네트워크에 자체 chrony NTP 서버를 배포하는 것이 좋습니다. 자세한 내용은 Red Hat Enterprise Linux 8 의 기본 시스템 설정 구성 의 Chrony Suite를 사용하여 NTP 구성 장을 참조하십시오.

참고

Ceph는 5분마다 시간 동기화를 평가하므로 문제를 수정하고 클럭 오차 메시지를 지우는 사이에 지연이 발생합니다.

추가 리소스

- Ceph Monitor 상태 이해
- Ceph Monitor가 쿼럼 없음

4.2.6. Ceph 모니터 저장소가 너무 큽니다.

ceph health 명령은 다음과 유사한 오류 메시지를 반환합니다.

```
mon.ceph1 store is getting too big! 48031 MB >= 15360 MB -- 62% avail
```

<그 Means>

Ceph Monitors 저장소는 사실 항목을 키-값 쌍으로 저장하는 **LevelDB** 데이터베이스입니다. 데이터베이스에는 클러스터 맵이 포함되어 있으며 기본적으로 **/var/lib/ceph/mon/CLUSTER_NAME-SHORT_HOST_NAME/store.db** 에 있습니다.

대규모 모니터 저장소를 쿼리하는 데 시간이 걸릴 수 있습니다. 결과적으로 클라이언트 쿼리에 응답하는 데 **Ceph** 모니터가 지연될 수 있습니다.

또한 **/var/** 파티션이 가득 차면 **Ceph Monitor**에서 저장소에 쓰기 작업을 수행할 수 없으며 종료됩니다. 이 문제 해결에 대한 자세한 내용은 **Ceph Monitor**가 쿼럼 상태임을 참조하십시오.

이 문제를 해결하기 위해

1. 데이터베이스 크기를 확인합니다.

```
du -sch /var/lib/ceph/mon/CLUSTER_NAME-SHORT_HOST_NAME/store.db
```

클러스터 이름과 **ceph-mon** 이 실행 중인 호스트의 짧은 호스트 이름을 지정합니다.

예제

```
# du -sch /var/lib/ceph/mon/ceph-host1/store.db
47G     /var/lib/ceph/mon/ceph-ceph1/store.db/
47G     total
```

2. **Ceph** 모니터 저장소를 압축합니다. 자세한 내용은 **Ceph Monitor Store**를 참조하십시오.

추가 리소스

- Ceph Monitor가 쿼럼 없음

4.2.7. Ceph Monitor 상태 이해

mon_status 명령은 다음과 같은 Ceph 모니터에 대한 정보를 반환합니다.

- 상태
- 랭크
- 선택 사항
- 모니터 맵(monmap)

Ceph 모니터가 쿼럼을 형성할 수 있는 경우 ceph 명령줄 유틸리티와 함께 mon_status 를 사용합니다.

Ceph 모니터가 쿼럼을 형성할 수 없지만 ceph-mon 데몬이 실행 중인 경우 관리 소켓을 사용하여 mon_status 를 실행합니다.

mon_status의 출력 예

```
{
    "name": "mon.3",
    "rank": 2,
    "state": "peon",
    "election_epoch": 96,
    "quorum": [
        1,
        2
    ],
    "outside_quorum": [],
    "extra_probe_peers": [],
    "sync_provider": [],
    "monmap": {
```

```
        "epoch": 1,
        "fsid": "d5552d32-9d1d-436c-8db1-ab5fc2c63cd0",
        "modified": "0.000000",
        "created": "0.000000",
        "mons": [
            {
                "rank": 0,
                "name": "mon.1",
                "addr": "172.25.1.10:6789\/0"
            },
            {
                "rank": 1,
                "name": "mon.2",
                "addr": "172.25.1.12:6789\/0"
            },
            {
                "rank": 2,
                "name": "mon.3",
                "addr": "172.25.1.13:6789\/0"
            }
        ]
    }
}
```

Ceph Monitor 상태

리더

*선택 단계에서 **Ceph** 모니터는 리더를 선택합니다. 리더는 순위가 가장 높은 **Ceph Monitor**이며 값이 가장 낮은 순위입니다. 위의 예에서 리더는 **mon.1** 입니다.*

peon

***peons**는 리더가 아닌 쿼럼의 **Ceph** 모니터입니다. 리더가 실패하면 가장 높은 순위를 가진 펩론이 새로운 리더가됩니다.*

probing

*다른 **Ceph** 모니터를 찾고 있는 경우 **Ceph** 모니터가 검색 상태이기 때문입니다. 예를 들어 **Ceph Monitor**를 시작한 후에는 **Ceph Monitor** 맵**(monmap)**에 지정된 충분한 **Ceph Monitor**를 찾을 때까지 계속 진행되어 쿼럼을 형성합니다.*

electing

***Ceph** 모니터는 리더를 선택하는 과정에 있는 경우 선택 상태에 있습니다. 일반적으로 이 상태는 빠르게 변경됩니다.*

동기화

다른 Ceph 모니터와 동기화되어 쿼럼에 조인하는 경우 Ceph 모니터가 동기화 상태입니다. Ceph Monitor가 작아지면 동기화 프로세스의 속도가 빨라집니다. 따라서 큰 저장소가 있는 경우 동기화에 시간이 더 오래 걸립니다.

추가 리소스

- 자세한 내용은 Red Hat Ceph Storage 5의 관리 가이드에서 Ceph 관리 소켓 사용 섹션을 참조하십시오.

4.2.8. 추가 리소스

- Red Hat Ceph Storage 문제 해결 가이드에서 4.2.2절. “Ceph Monitor 오류 메시지” 를 참조하십시오.
- Red Hat Ceph Storage 문제 해결 가이드에서 4.2.3절. “Ceph 로그의 일반적인 Ceph Monitor 오류 메시지” 를 참조하십시오.

4.3. MONMAP삽입

Ceph 모니터에 오래되거나 손상된 Ceph Monitor 맵(monmap)이 있는 경우 잘못된 IP 주소의 다른 Ceph 모니터에 도달하려고 하므로 쿼럼에 참여할 수 없습니다.

이 문제를 해결하는 가장 안전한 방법은 다른 Ceph Monitor에서 실제 Ceph 모니터 맵을 가져와 삽입하는 것입니다.

참고

이 작업은 Ceph Monitor에서 유지하는 기존 Ceph Monitor 맵을 덮어씁니다.

다음 절차에서는 다른 Ceph 모니터가 쿼럼을 형성하거나 하나 이상의 Ceph 모니터에 올바른 Ceph Monitor 맵이 있는 경우 Ceph Monitor 맵을 삽입하는 방법을 보여줍니다. 모든 Ceph 모니터에 손상된 저장소가 있으므로 Ceph Monitor 맵도 참조하십시오. https://access.redhat.com/documentation/en-us/red_hat_ceph_storage/5/html-single/troubleshooting_guide/#recovering-the-ceph-monitor-store

사전 요구 사항

- ***Ceph** 모니터 맵에 액세스합니다.*

- ***Ceph Monitor** 노드에 대한 루트 수준 액세스입니다.*

절차

1. *나머지 **Ceph** 모니터가 쿼럼을 형성할 수 있는 경우 **ceph mon getmap** 명령을 사용하여 **Ceph Monitor** 맵을 가져옵니다.*

   예제

   ```
   [ceph: root@host01 /]# ceph mon getmap -o /tmp/monmap
   ```

2. *나머지 **Ceph** 모니터가 쿼럼을 형성할 수 없고 올바른 **Ceph Monitor** 맵을 사용하여 하나 이상의 **Ceph Monitor**가 있는 경우 해당 **Ceph Monitor**에서 복사합니다.*

   a. *__Ceph Monitor__ 맵을 복사하려는 __Ceph Monitor__를 중지하십시오.*

      구문

      ```
      systemctl stop ceph-mon@HOST_NAME
      ```

      *예를 들어 **host01** 짧은 호스트 이름이 있는 호스트에서 실행 중인 **Ceph Monitor**를 중지하려면 다음을 수행합니다.*

      예제

```
[root@mon ~]# systemctl stop ceph-mon@host01
```

b. ***Ceph** 모니터 맵을 복사합니다.*

구문

```
ceph-mon -i ID --extract-monmap /tmp/monmap
```

***ID** 를 다음에서 **Ceph Monitor** 맵을 복사하려는 **Ceph Monitor ID**로 바꿉니다.*

예제

```
[ceph: root@host01 /]#  ceph-mon -i mon.a  --extract-monmap /tmp/monmap
```

3. *손상되거나 오래된 **Ceph** 모니터 맵을 사용하여 **Ceph** 모니터를 중지합니다.*

구문

```
systemctl stop ceph-mon@HOST_NAME
```

*예를 들어 **host01** 짧은 호스트 이름이 있는 호스트에서 실행 중인 **Ceph Monitor**를 중지하려면 다음을 수행합니다.*

예제

```
[root@mon ~]# systemctl stop ceph-mon@host01
```

4. ***Ceph** 모니터 맵 삽입:*

구문

```
ceph-mon -i ID --inject-monmap /tmp/monmap
```

***ID** 를 **Ceph Monitor**의 **ID**로 교체하거나 오래된 **Ceph** 모니터 맵으로 바꿉니다.*

예제

```
[root@mon ~]# ceph-mon -i mon.host01 --inject-monmap /tmp/monmap
```

5. ***Ceph** 모니터를 시작합니다.*

예제

```
[root@mon ~]# systemctl start ceph-mon@host01
```

다른 **Ceph Monitor**에서 **Ceph Monitor** 맵을 복사한 경우 해당 **Ceph Monitor**도 시작합니다.

예제

```
[root@mon ~]# systemctl start ceph-mon@host01
```

추가 리소스

- **Ceph Monitor**가 쿼럼이 아닌 것을 확인합니다.
- **Ceph Monitor** 저장소 복구참조

4.4. 실패한 모니터 교체

Ceph 모니터에 손상된 저장소가 있는 경우 스토리지 클러스터에서 모니터를 교체할 수 있습니다.

사전 요구 사항

- 실행 중인 **Red Hat Ceph Storage** 클러스터.
- 쿼럼을 형성할 수 있습니다.
- **Ceph Monitor** 노드에 대한 루트 수준 액세스.

절차

1. 모니터 호스트에서 기본적으로 **/var/lib/ceph/mon/CLUSTER_NAME-SHORT_HOST_NAME** 에 있는 모니터 저장소를 제거합니다.

```
rm -rf /var/lib/ceph/mon/CLUSTER_NAME-SHORT_HOST_NAME
```

Monitor 호스트의 짧은 호스트 이름과 클러스터 이름을 지정합니다. 예를 들어 remote 라는 클러스터에서 host1 에서 실행 중인 모니터의 모니터 저장소를 제거하려면 다음을 수행합니다.

```
[root@mon ~]# rm -rf /var/lib/ceph/mon/remote-host1
```

2. *Monitor 맵에서 모니터를 제거합니다(monmap).*

```
ceph mon remove SHORT_HOST_NAME --cluster CLUSTER_NAME
```

Monitor 호스트의 짧은 호스트 이름과 클러스터 이름을 지정합니다. 예를 들어 remote 라는 클러스터에서 host1 에서 실행 중인 모니터를 제거하려면 다음을 수행합니다.

```
[ceph: root@host01 /]# ceph mon remove host01 --cluster remote
```

3. *Monitor 호스트의 기본 파일 시스템 또는 하드웨어와 관련된 문제를 해결합니다.*

추가 리소스

- *자세한 내용은 Ceph Monitor is out of quorum 를 참조하십시오.*

4.5. 모니터 저장소 압축

모니터 저장소 크기가 크게 증가하면 다음을 압축할 수 있습니다.

- *ceph tell 명령을 사용하여 동적으로 설정합니다.*
- *ceph-mon 데몬 시작 시*
- *ceph-mon 데몬이 실행되지 않는 경우 ceph-monstore-tool 을 사용합니다. 이전에 언급된 방법에서 모니터 저장소를 압축하지 못하거나 모니터가 쿼럼 상태가 되고 해당 로그에 Caught 신호(Bus 오류) 오류 메시지가 포함된 경우 이 방법을 사용합니다.*

중요

클러스터가 **active+clean** 상태가 아니거나 리밸런싱 프로세스 중에 저장 크기 변경 사항을 모니터링합니다. 따라서 리밸런싱이 완료되면 모니터 저장소를 압축합니다. 또한 배치 그룹이 **active+clean** 상태인지 확인합니다.

사전 요구 사항

- 실행 중인 **Red Hat Ceph Storage** 클러스터.
- **Ceph Monitor** 노드에 대한 루트 수준 액세스입니다.

절차

1. **ceph-mon** 데몬이 실행되는 경우 모니터 저장소를 압축하려면 다음을 수행합니다.

구문

```
ceph tell mon.HOST_NAME compact
```

2. **HOST_NAME** 을 **ceph-mon** 이 실행 중인 호스트의 짧은 호스트 이름으로 교체합니다. 확실하지 않은 경우 **hostname -s** 명령을 사용하십시오.

예제

```
[ceph: root@host01 /]# ceph tell mon.host01 compact
```

3. **[mon]** 섹션의 **Ceph** 구성에 다음 매개 변수를 추가합니다.

```
[mon]
mon_compact_on_start = true
```

4. **ceph-mon** 데몬을 다시 시작합니다.

구문

```
systemctl restart ceph-mon@HOST_NAME
```

HOST_NAME 을 데몬이 실행 중인 호스트의 짧은 이름으로 교체합니다. 확실하지 않은 경우 **hostname -s** 명령을 사용하십시오.

예제

```
[root@mon ~]# systemctl restart ceph-mon@host01
```

5. 모니터가 쿼럼을 형성했는지 확인합니다.

```
[ceph: root@host01 /]# ceph mon stat
```

6. 필요한 경우 다른 모니터에서 이러한 단계를 반복합니다.

참고

시작하기 전에 **ceph-test** 패키지가 설치되어 있는지 확인합니다.

7. 큰 저장소의 **ceph-mon** 데몬이 실행되고 있지 않은지 확인합니다. 필요한 경우 데몬을 중지합니다.

구문

```
[root@mon ~]# systemctl status ceph-mon@HOST_NAME
[root@mon ~]# systemctl stop ceph-mon@HOST_NAME
```

***HOST_NAME** 을 데몬이 실행 중인 호스트의 짧은 이름으로 교체합니다. 확실하지 않은 경우 **hostname -s** 명령을 사용하십시오.*

예제

```
[root@mon ~]# systemctl status ceph-mon@host01
[root@mon ~]# systemctl stop ceph-mon@host01
```

8. *모니터 저장소를 압축합니다.*

구문

```
ceph-monstore-tool /var/lib/ceph/mon/mon.HOST_NAME compact
```

***HOST_NAME** 을 **Monitor** 호스트의 짧은 호스트 이름으로 교체합니다.*

예제

```
[ceph: root@host01 /]# ceph-monstore-tool /var/lib/ceph/mon/mon.host01 compact
```

9. **ceph-mon 을 다시 시작합니다.**

구문

```
systemctl start ceph-mon@HOST_NAME
```

예제

```
[root@mon ~]# systemctl start ceph-mon@host01
```

추가 리소스

- **Ceph 모니터 저장소가 너무 큽니다.**
- **Ceph Monitor가 쿼럼이 아닌 것을 확인합니다.**

4.6. CEPH 관리자의 포트 열기

ceph-mgr 데몬은 ceph-osd 데몬과 동일한 범위의 포트의 OSD에서 배치 그룹 정보를 수신합니다. 이러한 포트가 열려 있지 않으면 클러스터는 HEALTH_OK 에서 HEALTH_WARN 로 devolve를 사용하며 PGs가 PGs unknown 의 백분율로 알 수 없음을 나타냅니다.

사전 요구 사항

- *실행 중인 **Red Hat Ceph Storage** 클러스터.*

- **Ceph Manager**에 대한 루트 수준 액세스.

절차

1. 이 문제를 해결하려면 **ceph-mgr** 데몬을 실행하는 각 호스트의 경우 포트 **6800-7300** 을 엽니다.

   예제

   ```
   [root@ceph-mgr] # firewall-cmd --add-port 6800-7300/tcp
   [root@ceph-mgr] # firewall-cmd --add-port 6800-7300/tcp --permanent
   ```

2. **ceph-mgr** 데몬을 다시 시작합니다.

4.7. CEPH MONITOR 저장소 복구

Ceph 모니터는 클러스터 맵을 **LevelDB**와 같은 키-값 저장소에 저장합니다. 저장소가 모니터에서 손상된 경우 모니터는 예기치 않게 종료되고 다시 시작되지 않습니다. **Ceph** 로그에 다음과 같은 오류가 포함될 수 있습니다.

```
Corruption: error in middle of record
Corruption: 1 missing files; e.g.: /var/lib/ceph/mon/mon.0/store.db/1234567.ldb
```

Red Hat Ceph Storage 클러스터는 **3**개 이상의 **Ceph Monitor**를 사용하므로, 오류가 발생하면 다른 모니터로 교체할 수 있습니다. 그러나 특정 상황에서는 모든 **Ceph** 모니터에 손상된 저장소를 가질 수 있습니다. 예를 들어 **Ceph Monitor** 노드가 디스크 또는 파일 시스템 설정을 잘못 구성한 경우 정전으로 인해 기본 파일 시스템이 손상될 수 있습니다.

모든 **Ceph** 모니터에 손상이 있는 경우 **ceph-monstore-tool** 및 **ceph-objectstore-tool** 이라는 유틸리티를 사용하여 **OSD** 노드에 저장된 정보로 복구할 수 있습니다.

중요

이러한 절차는 다음 정보를 복구할 수 없습니다.

- 메타데이터 데몬 서버(MDS) 키링 및 맵
- 배치 그룹 설정:
  - ceph pg set_full_ratio 명령을 사용하여 전체 비율
  - ceph pg set_ nearfull _ratio 명령을 사용하여 설정된 가까운full 비율

중요

이전 백업에서 Ceph Monitor 저장소를 복원하지 마십시오. 다음 단계를 사용하여 현재 클러스터 상태에서 Ceph Monitor 저장소를 다시 빌드하고 해당 위치에서 복원합니다.

4.7.1. BlueStore를 사용할 때 Ceph Monitor 저장소 복구

모든 Ceph Monitor에서 Ceph Monitor 저장소가 손상되고 BlueStore 백엔드를 사용하는 경우 다음 절차를 따르십시오.

컨테이너화된 환경에서는 이 방법으로 Ceph 리포지토리를 연결하고 컨테이너화되지 않은 Ceph 모니터로 먼저 복원해야 합니다.

주의

이 절차에서는 데이터 손실을 초래할 수 있습니다. 이 절차의 어떤 단계에 대해 잘 모르는 경우 복구 프로세스에 대한 Red Hat 기술 지원에 문의하십시오.

사전 요구 사항

- *모든 **OSD** 컨테이너가 중지됩니다.*
- *역할을 기반으로 **Ceph** 노드에서 **Ceph** 리포지토리를 활성화합니다.*
- ***ceph-test** 및 **rsync** 패키지가 **OSD** 및 모니터 노드에 설치됩니다.*
- ***ceph-mon** 패키지는 **Monitor** 노드에 설치됩니다.*
- ***ceph-osd** 패키지는 **OSD** 노드에 설치됩니다.*

절차

1. ***Ceph** 데이터가 있는 모든 디스크를 임시 위치에 마운트합니다. 모든 **OSD** 노드에 대해 이 단계를 반복합니다.*

   a. ***ceph-volume** 명령을 사용하여 데이터 파티션을 나열합니다.*

   예제

   ```
   [ceph: root@host01 /]# ceph-volume lvm list
   ```

   b. *데이터 파티션을 임시 위치에 마운트합니다.*

   구문

   ```
   mount -t tmpfs tmpfs /var/lib/ceph/osd/ceph-$i
   ```

c.

***SELinux** 컨텍스트를 복원합니다.*

구문

```
for i in {OSD_ID}; do restorecon /var/lib/ceph/osd/ceph-$i; done
```

***OSD_ID** 를 **OSD** 노드의 **Ceph OSD ID**의 숫자 공백으로 구분된 목록으로 교체합니다.*

d.

*소유자와 그룹을 **ceph:ceph** 로 변경합니다.*

구문

```
for i in {OSD_ID}; do chown -R ceph:ceph /var/lib/ceph/osd/ceph-$i; done
```

***OSD_ID** 를 **OSD** 노드의 **Ceph OSD ID**의 숫자 공백으로 구분된 목록으로 교체합니다.*

중요

update-mon-db 명령에서 **Monitor** 데이터베이스에 추가 **db** 및 **db.slow** 디렉토리를 사용하도록 하는 버그로 인해 이러한 디렉터리도 복사해야 합니다. 이렇게 하려면 다음을 수행합니다.

1. 컨테이너 외부의 임시 위치를 준비하여 **OSD** 데이터베이스를 마운트 및 액세스하고 **Ceph** 모니터를 복원하는 데 필요한 **OSD** 맵을 추출합니다.

   구문

   ```
   ceph-bluestore-tool --cluster=ceph prime-osd-dir --dev OSD-DATA --path /var/lib/ceph/osd/ceph-OSD-ID
   ```

   OSD-DATA 를 **OSD** 데이터의 볼륨 그룹(VG) 또는 논리 볼륨(LV) 경로로 바꾸고 **OSD-ID** 를 **OSD**의 **ID**로 바꿉니다.

2. **BlueStore** 데이터베이스와 **block.db** 사이에 심볼릭 링크를 만듭니다.

   구문

   ```
   ln -snf BLUESTORE DATABASE /var/lib/ceph/osd/ceph-OSD-ID/block.db
   ```

   BLUESTORE-DATABASE 를 **BlueStore** 데이터베이스의 **Volume Group(VG)** 또는 **LV(Logical Volume)** 경로로 바꾸고 **OSD-ID** 를 **OSD ID**로 바꿉니다.

2. 손상된 저장소와 함께 **Ceph Monitor** 노드에서 다음 명령을 사용합니다. 모든 노드의 모든

***OSD*에 대해 이를 반복합니다.**

a.

*모든 **OSD** 노드에서 클러스터 맵을 수집합니다.*

예제

```
[root@host01 ~]# cd /root/
[root@host01 ~]# ms=/tmp/monstore/
[root@host01 ~]# db=/root/db/
[root@host01 ~]# db_slow=/root/db.slow/

[root@host01 ~]# mkdir $ms
[root@host01 ~]# for host in $osd_nodes; do
        echo "$host"
        rsync -avz $ms $host:$ms
        rsync -avz $db $host:$db
        rsync -avz $db_slow $host:$db_slow

        rm -rf $ms
        rm -rf $db
        rm -rf $db_slow

        sh -t $host <<EOF
          for osd in /var/lib/ceph/osd/ceph-*; do
            ceph-objectstore-tool --type bluestore --data-path \$osd --op update-mon-db
--mon-store-path $ms

          done
        EOF

            rsync -avz $host:$ms $ms
            rsync -avz $host:$db $db
            rsync -avz $host:$db_slow $db_slow
        done
```

b.

적절한 기능을 설정합니다.

예제

```
[ceph: root@host01 /]# ceph-authtool /etc/ceph/ceph.client.admin.keyring -n mon. --cap
mon 'allow *' --gen-key
```

```
[ceph: root@host01 /]# cat /etc/ceph/ceph.client.admin.keyring
 [mon.]
   key = AQCleqldWqm5IhAAgZQbEzoShkZV42RiQVffnA==
   caps mon = "allow *"
 [client.admin]
   key = AQCmAKld8J05KxAArOWeRAw63gAwwZO5o75ZNQ==
   auid = 0
   caps mds = "allow *"
   caps mgr = "allow *"
   caps mon = "allow *"
   caps osd = "allow *"
```

c. ***db* 및 *db.slow* 디렉토리에서 모든 *sst* 파일을 임시 위치로 이동합니다.**

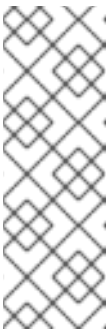
예제

```
[ceph: root@host01 /]# mv /root/db/*.sst /root/db.slow/*.sst /tmp/monstore/store.db
```

d. **수집된 맵에서 *Monitor* 저장소를 다시 빌드합니다.**

예제

```
[ceph: root@host01 /]# ceph-monstore-tool /tmp/monstore rebuild -- --keyring
/etc/ceph/ceph.client.admin
```



참고

이 명령을 사용하면 *OSD*에서 추출된 인증 키 및 *ceph-monstore-tool* 명령줄에 지정된 키 키만 *Ceph* 인증 데이터베이스에 있습니다. 클라이언트, *Ceph Manager, Ceph Object Gateway* 등과 같은 기타 모든 인증 키를 다시 생성하거나 가져와야 하므로 해당 클라이언트가 클러스터에 액세스할 수 있습니다.

e. *손상된 저장소를 백업합니다. 모든 **Ceph Monitor** 노드에 대해 이 단계를 반복합니다.*

구문

```
mv /var/lib/ceph/mon/ceph-HOSTNAME/store.db
/var/lib/ceph/mon/ceph-HOSTNAME/store.db.corrupted
```

***HOSTNAME** 을 **Ceph Monitor** 노드의 호스트 이름으로 교체합니다.*

f. *손상된 저장소를 교체합니다. 모든 **Ceph Monitor** 노드에 대해 이 단계를 반복합니다.*

구문

```
scp -r /tmp/monstore/store.db HOSTNAME:/var/lib/ceph/mon/ceph-HOSTNAME/
```

***HOSTNAME** 을 **Monitor** 노드의 호스트 이름으로 교체합니다.*

g. *새 저장소의 소유자를 변경합니다. 모든 **Ceph Monitor** 노드에 대해 이 단계를 반복합니다.*

구문

```
chown -R ceph:ceph /var/lib/ceph/mon/ceph-HOSTNAME/store.db
```

***HOSTNAME** 을 **Ceph Monitor** 노드의 호스트 이름으로 교체합니다.*

3. *임시 마운트된 모든 **OSD**를 모든 노드에서 마운트 해제합니다.*

예제

```
[root@host01 ~]# umount /var/lib/ceph/osd/ceph-*
```

4. *모든 **Ceph Monitor** 데몬을 시작합니다.*

```
[root@host01 ~]# systemctl start ceph-mon *
```

5. *모니터가 쿼럼을 형성할 수 있는지 확인합니다.*

구문

```
ceph -s
```

***HOSTNAME** 을 **Ceph Monitor** 노드의 호스트 이름으로 교체합니다.*

6. ***Ceph Manager** 인증 키를 가져오고 모든 **Ceph Manager** 프로세스를 시작합니다.*

구문

```
ceph auth import -i /etc/ceph/ceph.mgr.HOSTNAME.keyring
systemctl start ceph-mgr@HOSTNAME
```

HOSTNAME 을 Ceph Manager 노드의 호스트 이름으로 교체합니다.

7. *모든 OSD 노드에서 모든 OSD 프로세스를 시작합니다.*

예제

```
[root@host01 ~]# systemctl start ceph-osd *
```

8. *OSD가 서비스로 돌아가 있는지 확인합니다.*

예제

```
[ceph: root@host01 /]# ceph -s
```

추가 리소스

- *Ceph 노드를 CDN(Content Delivery Network)에 등록하는 방법에 대한 자세한 내용은 Red Hat Ceph Storage 설치 가이드의 CDN에 Red Hat Ceph Storage 노드 등록 및 연결 섹션을 참조하십시오.*

4.8. 추가 리소스

- *네트워크 관련 문제는 Red Hat Ceph Storage 문제 해결 가이드에서 3장. 네트워킹 문제 해결 를 참조하십시오.*

5장. CEPH OSD 문제 해결

이 장에서는 Ceph OSD와 관련된 가장 일반적인 오류를 수정하는 방법을 설명합니다.

5.1. 사전 요구 사항

- 네트워크 연결을 확인합니다. 자세한 내용은 네트워킹 문제 해결을 참조하십시오.
- ceph health 명령을 사용하여 모니터에 쿼럼이 있는지 확인합니다. 명령에서 상태 상태 (HEALTH_OK,HEALTH_WARN 또는 HEALTH_ERR)를 반환하는 경우 모니터는 쿼럼을 형성할 수 있습니다. 그렇지 않은 경우 먼저 모니터 문제를 해결합니다. 자세한 내용은 Ceph 모니터 문제 해결을 참조하십시오. ceph 상태에 대한 자세한 내용은 Ceph 상태 이해를 참조하십시오.
- 필요한 경우 리밸런싱 프로세스를 중지하여 시간과 리소스를 절약합니다. 자세한 내용은 중지 및 재조정 을 참조하십시오.

5.2. 대부분의 CEPH OSD 오류

다음 표에는 ceph health detail 명령에서 반환되거나 Ceph 로그에 포함된 가장 일반적인 오류 메시지가 나열되어 있습니다. 표에서는 오류를 설명하고 문제를 해결하기 위한 특정 절차를 가리키는 해당 섹션에 대한 링크를 제공합니다.

5.2.1. 사전 요구 사항

- Ceph OSD 노드에 대한 루트 수준 액세스입니다.

5.2.2. Ceph OSD 오류 메시지

일반적인 Ceph OSD 오류 메시지 테이블 및 잠재적인 수정 사항.

오류 메시지	참조
HEALTH_ERR	
전체 **osds**	전체 OSD
HEALTH_WARN	

오류 메시지	참조
가까운 전체 **osds**	nearfull OSD
OSD가 다운됨	OSD 다운 플로팅 OSD
요청이 차단됩니다.	느린 요청 또는 요청이 차단됨
느린 요청	느린 요청 또는 요청이 차단됨

5.2.3. Ceph 로그의 일반적인 Ceph OSD 오류 메시지

***Ceph** 로그에 있는 일반적인 **Ceph OSD** 오류 메시지와 잠재적인 수정 사항에 대한 링크입니다.*

오류 메시지	로그 파일	참조
heartbeat_check: osd.X의 응답이 없음	기본 클러스터 로그	플로팅 OSD
잘못 표시되어 있습니다.	기본 클러스터 로그	플로팅 OSD
OSD에는 느린 요청이 있습니다.	기본 클러스터 로그	느린 요청 또는 요청이 차단됨
FAILED assert(0 == "취약점 시간 초과**")**	OSD 로그	OSD 다운

5.2.4. 전체 OSD

***ceph health detail** 명령은 다음과 유사한 오류 메시지를 반환합니다.*

```
HEALTH_ERR 1 full osds
osd.3 is full at 95%
```

<그 Means>

***Ceph**에서는 클라이언트가 전체 **OSD** 노드에서 **I/O** 작업을 수행하여 데이터 손실을 방지합니다. 클러스터가 **mon _osd_full_ratio** 매개 변수에서 설정한 용량에 도달하면 **HEALTH_ERR** 전체 **osds** 메시지를 반환합니다. 기본적으로 이 매개변수는 **0.95** 로 설정되어 클러스터 용량의 **95%**를 의미합니다.*

이 문제를 해결하기 위해

원시 스토리지의 백분율(%RAW USED)이 사용되었는지 확인합니다.

```
ceph df
```

%RAW USED 가 70-75%를 초과하는 경우 다음을 수행할 수 있습니다.

- 불필요한 데이터를 삭제합니다. 이는 프로덕션 다운 타임을 방지하기 위한 단기 솔루션입니다.
- 새 OSD 노드를 추가하여 클러스터를 확장합니다. 이는 Red Hat에서 권장하는 장기적인 솔루션입니다.

추가 리소스

- Red Hat Ceph Storage 문제 해결 가이드 의 가까운 OSD.
- 자세한 내용은 전체 스토리지 클러스터에서 데이터 삭제를 참조하십시오.

5.2.5. 전체 OSD 가까운

ceph health detail 명령은 다음과 유사한 오류 메시지를 반환합니다.

```
HEALTH_WARN 1 nearfull osds
osd.2 is near full at 85%
```

<그 Means>

클러스터가 mon osd nearfull 비율 defaults 매개변수가 설정한 용량에 도달하면 Ceph가 nearfull osds 메시지를 반환합니다. 기본적으로 이 매개변수는 0.85 로 설정되어 클러스터 용량의 85%를 의미합니다.

Ceph는 CRUSH 계층 구조에 따라 데이터를 최상의 방식으로 배포하지만 동일한 배포를 보장할 수 없습니다. 11개의 데이터 배포 및 가까운full osds 메시지의 주요 원인은 다음과 같습니다.

- OSD는 클러스터의 OSD 노드 간에 균형을 맞추지 않습니다. 즉, 일부 OSD 노드는 다른 OSD보다 훨씬 더 많은 OSD를 호스팅하거나 CRUSH 맵의 일부 OSD 가중치가 용량에 적합하지

않습니다.

- *OSD 수, 사용 사례, OSD당 대상 PG(배치 그룹) 수에 따라 배치 그룹(PG) 수가 적절하지 않습니다.*
- *클러스터는 CRUSH 튜닝 가능 항목을 부적절하게 사용합니다.*
- *OSD의 백엔드 스토리지는 거의 가득 차 있습니다.*

이 문제를 해결하기 위해:

1. *PG 개수가 충분한지 확인하고 필요한 경우 늘립니다.*
2. *CRUSH 튜닝 가능 항목을 클러스터 버전에 최적으로 사용하는지 확인하고, 그러지 않으면 조정합니다.*
3. *사용률로 OSD의 가중치를 변경합니다.*
4. *OSD에서 사용하는 디스크에 남아 있는 공간 크기를 결정합니다.*

   a. *일반적으로 OSD가 얼마나 많은 공간을 사용하는지 확인하려면 다음을 수행하십시오.*

   ```
   [ceph: root@host01 /]# ceph osd df
   ```

   b. *특정 노드에서 공간 OSD의 크기를 확인합니다. 전체 OSD가 포함된 노드에서 다음 명령을 사용합니다.*

   ```
   df
   ```

   c. *필요한 경우 새 OSD 노드를 추가합니다.*

추가 리소스

- *전체 OSD*

- *Red Hat Ceph Storage 5용 스토리지 전략 가이드 의 사용률에 따른 OSD의 Weight 설정 섹션을 참조하십시오.*

- *자세한 내용은 Red Hat Ceph Storage 5의 Storage Strategies 가이드의 CRUSH Tunables 섹션을 참조하십시오. Red Hat Ceph Storage의 OSD 에서 CRUSH 맵 튜닝 가능 수정 사항이 Red Hat 고객 포털의 OSD 간 PG 배포에 미치는 영향을 테스트하는 방법에서 참조하십시오.*

- *자세한 내용은 배치 그룹 승격 을 참조하십시오.*

5.2.6. OSD 다운

ceph health detail 명령은 다음과 유사한 오류를 반환합니다.

```
HEALTH_WARN 1/3 in osds are down
```

<그 Means>

서비스 장애 또는 다른 OSD와의 통신 문제로 인해 ceph-osd 프로세스 중 하나를 사용할 수 없습니다. 그 결과 남아 있는 ceph-osd 데몬에서 모니터에 오류가 발생했습니다.

ceph-osd 데몬이 실행되지 않으면 기본 OSD 드라이브 또는 파일 시스템이 손상되었거나 누락된 인증 키와 같은 기타 오류가 데몬이 시작되지 않습니다.

대부분의 경우 네트워킹 문제로 인해 ceph-osd 데몬이 실행 중이지만 여전히 down 으로 표시되는 상황이 발생합니다.

이 문제를 해결하기 위해

1. *다운 된 OSD를 확인합니다.*

```
[ceph: root@host01 /]# ceph health detail
HEALTH_WARN 1/3 in osds are down
osd.0 is down since epoch 23, last address 192.168.106.220:6800/11080
```

2. **ceph-osd** 데몬을 다시 시작합니다.

```
[root@host01 ~]# systemctl restart ceph-osd@OSD_NUMBER
```

OSD_NUMBER 를 다운 된 **OSD**의 **ID**로 바꿉니다. 예를 들면 다음과 같습니다.

```
[root@host01 ~]# systemctl restart ceph-osd@0
```

a. **ceph-osd** 를 시작할 수 없는 경우 **ceph-osd** 데몬의 단계를 따르십시오.

b. **ceph-osd** 데몬을 시작할 수 있지만 **down** 으로 표시된 경우 **The ceph-osd** 데몬의 단계를 따르십시오. 그러나 여전히 **'down'**으로 표시됩니다.

ceph-osd 데몬을 시작할 수 없음

1. 다수의 **OSD**(일반적으로 **12**개)가 포함된 노드가 있는 경우 기본 최대 스레드 수(**PID** 수)가 충분한지 확인합니다. 자세한 내용은 **PID** 수 증가를 참조하십시오.

2. **OSD** 데이터 및 저널 파티션이 제대로 마운트되었는지 확인합니다. **ceph-volume lvm list** 명령을 사용하여 **Ceph Storage** 클러스터와 연결된 모든 장치 및 볼륨을 나열한 다음 제대로 마운트되었는지 수동으로 검사할 수 있습니다. 자세한 내용은 **mount(8)** 매뉴얼 페이지를 참조하십시오.

3. **ERROR: missing keyring**이 있는 경우 인증 오류 메시지에 **requirement**를 사용할 수 없습니다. **OSD**는 누락된 인증 키입니다.

4. **ERROR: unable to open OSD superblock on /var/lib/ceph/osd/ceph-1** 오류 메시지가 표시되면 **ceph-osd** 데몬에서 기본 파일 시스템을 읽을 수 없습니다. 이 오류 문제를 해결하고 수정하는 방법에 대한 지침은 다음 단계를 참조하십시오.

a. 해당 로그 파일을 확인하여 오류 원인을 확인합니다. 기본적으로 **Ceph**는 **/var/log/ceph/** 디렉터리에 로그 파일을 저장합니다.

b. **EIO** 오류 메시지는 기본 디스크의 실패를 나타냅니다. 이 문제를 해결하려면 기본 **OSD** 디스크를 교체하십시오. 자세한 내용은 **OSD** 드라이브 교체를 참조하십시오.

c. 로그에 다음과 같은 기타 **FAILED assert** 오류가 포함된 경우 지원 티켓을 엽니다. *자세한 내용은* ***Red Hat*** *지원 문의를 참조하십시오*.

```
FAILED assert(0 == "hit suicide timeout")
```

5. 기본 파일 시스템 또는 디스크에 오류가 있는지 **dmesg** 출력에서 확인합니다.

```
dmesg
```

a. 다음과 유사한 **-5** 오류 메시지는 기본 **XFS** 파일 시스템의 손상을 나타냅니다. 이 문제를 해결하는 방법에 대한 자세한 내용은 **Red Hat Customer Portal** *의* ***"xfs_log_force: error -5 returned"****의 의미를 참조하십시오*.

```
xfs_log_force: error -5 returned
```

b. **dmesg** 출력에 **SCSI** 오류 메시지가 포함된 경우 **Red Hat Customer Portal**의 ***SCSI*** *오류 코드* ***Solution*** **find** *솔루션을 참조하여 문제를 해결하는 가장 좋은 방법을 결정합니다.*

c. 또는 기본 파일 시스템을 수정할 수 없는 경우 **OSD** 드라이브를 교체합니다. *자세한 내용은* ***OSD*** *드라이브 교체를 참조하십시오*.

6. **OSD**가 다음과 같은 세그멘테이션 오류로 인해 실패한 경우 필요한 정보를 수집하고 지원 티켓을 엽니다. *자세한 내용은* ***Red Hat*** *지원 문의를 참조하십시오*.

```
Caught signal (Segmentation fault)
```

ceph-osd *가 실행 중이지만 여전히* **down**으로 표시됩니다.

1. 해당 로그 파일을 확인하여 오류 원인을 확인합니다. 기본적으로 **Ceph**는 **/var/log/ceph/** 디렉터리에 로그 파일을 저장합니다.

a. 로그에 다음 항목과 유사한 오류 메시지가 포함된 경우 ***Flapping OSD*** *를 참조하십시오*.

```
wrongly marked me down
heartbeat_check: no reply from osd.2 since back
```

b.

다른 오류가 표시되면 지원 티켓을 엽니다. 자세한 내용은 **Red Hat** 지원 문의를 참조하십시오.

추가 리소스

- OSD 분리
- 오래된 배치 그룹

5.2.7. OSD 분리

ceph -w | grep osds 명령은 **OSD**를 **down** 으로 반복적으로 표시한 다음 짧은 시간 내에 다시 가동 합니다.

```
ceph -w | grep osds
2022-05-05 06:27:20.810535 mon.0 [INF] osdmap e609: 9 osds: 8 up, 9 in
2022-05-05 06:27:24.120611 mon.0 [INF] osdmap e611: 9 osds: 7 up, 9 in
2022-05-05 06:27:25.975622 mon.0 [INF] HEALTH_WARN; 118 pgs stale; 2/9 in osds are down
2022-05-05 06:27:27.489790 mon.0 [INF] osdmap e614: 9 osds: 6 up, 9 in
2022-05-05 06:27:36.540000 mon.0 [INF] osdmap e616: 9 osds: 7 up, 9 in
2022-05-05 06:27:39.681913 mon.0 [INF] osdmap e618: 9 osds: 8 up, 9 in
2022-05-05 06:27:43.269401 mon.0 [INF] osdmap e620: 9 osds: 9 up, 9 in
2022-05-05 06:27:54.884426 mon.0 [INF] osdmap e622: 9 osds: 8 up, 9 in
2022-05-05 06:27:57.398706 mon.0 [INF] osdmap e624: 9 osds: 7 up, 9 in
2022-05-05 06:27:59.669841 mon.0 [INF] osdmap e625: 9 osds: 6 up, 9 in
2022-05-05 06:28:07.043677 mon.0 [INF] osdmap e628: 9 osds: 7 up, 9 in
2022-05-05 06:28:10.512331 mon.0 [INF] osdmap e630: 9 osds: 8 up, 9 in
2022-05-05 06:28:12.670923 mon.0 [INF] osdmap e631: 9 osds: 9 up, 9 in
```

또한 **Ceph** 로그에는 다음과 유사한 오류 메시지가 포함되어 있습니다.

```
2022-05-25 03:44:06.510583 osd.50 127.0.0.1:6801/149046 18992 : cluster [WRN] map e600547
wrongly marked me down
```

```
2022-05-25 19:00:08.906864 7fa2a0033700 -1 osd.254 609110 heartbeat_check: no reply from
osd.2 since back 2021-07-25 19:00:07.444113 front 2021-07-25 18:59:48.311935 (cutoff 2021-07-25
18:59:48.906862)
```

<그 Means>

OSD를 플래핑하는 주요 원인은 다음과 같습니다.

-

*스크럽 또는 복구와 같은 특정 스토리지 클러스터 작업에는 대규모 인덱스 또는 대규모 배치 그룹이 있는 오브젝트에 대해 이러한 작업을 수행하는 경우 비정상 시간이 걸립니다. 일반적으로 이러한 작업이 완료되면 푸핑 **OSD** 문제가 해결됩니다.*

- *기본 물리적 하드웨어에 문제가 있습니다. 이 경우 **ceph health detail** 명령에서도 느린 요청 오류 메시지를 반환합니다.*

- *네트워크에 문제가 있습니다.*

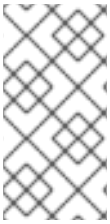
***Ceph OSD**는 스토리지 클러스터의 개인 네트워크가 실패하는 상황을 관리할 수 없으며, 공용 클라이언트 쪽 네트워크에 있는 대기 시간이 상당히 중요합니다.*

***Ceph OSD**는 개인 네트워크를 사용하여 하트비트 패킷을 서로 보내어 실행 중이며 임을 나타냅니다. 개인 스토리지 클러스터 네트워크가 제대로 작동하지 않으면 **OSD**가 하트비트 패킷을 보내고 수신할 수 없습니다. 결과적으로 이들은 **Ceph** 모니터의 다운 된 상태로 서로 보고하고, 이를 **up** 으로 표시합니다.*

***Ceph** 구성 파일의 다음 매개 변수는 이 동작에 영향을 미칩니다.*

매개변수	설명	기본값
osd_heartbeat_grace_time	OSD를 Ceph 모니터로 보고하기 전에 반환할 하트비트 패킷을 대기하는 OSD 시간입니다.	20초
mon_osd_min_down_reporters	Ceph 모니터가 OSD를 **down** 으로 표시하기 전에 다른 OSD를 **아래로**보고해야 하는 OSD 수	2

*이 표는 기본 구성에서 하나의 **OSD**만 종료 중인 첫 번째 **OSD**에 대해 **3**개의 개별 보고를 수행한 경우 **OSD**를 아래로 표시합니다. 경우에 따라 하나의 호스트가 네트워크 문제가 발생하는 경우 전체 클러스터에 **OSD**가 소모되는 경우가 있습니다. 호스트에 있는 **OSD**가 클러스터의 다른 **OSD**를 **down** 으로 보고하기 때문입니다.*



참고

*회전 **OSD** 시나리오에는 **OSD** 프로세스가 시작된 후 즉시 종료되는 상황이 발생하지 않습니다.*

이 문제를 해결하기 위해

1. ***ceph health detail** 명령의 출력을 다시 확인합니다. 느린 요청 오류 메시지가 포함된 경우 이 문제 해결 방법에 대한 세부 정보를 참조하십시오.*

```
ceph health detail
HEALTH_WARN 30 requests are blocked > 32 sec; 3 osds have slow requests
30 ops are blocked > 268435 sec
1 ops are blocked > 268435 sec on osd.11
1 ops are blocked > 268435 sec on osd.18
28 ops are blocked > 268435 sec on osd.39
3 osds have slow requests
```

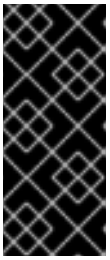
2. *다운 된 노드와 노드가 있는 **OSD**를 확인합니다.*

```
ceph osd tree | grep down
```

3. ***OSD**가 포함된 노드에서 네트워크 문제를 해결합니다. 자세한 내용은 네트워킹 문제 해결을 참조하십시오.*

4. *또는 **no up** 및 **nodown** 플래그를 설정하여 **OSD**를 **down** 으로 표시하도록 임시로 강제 실행할 수 있습니다.*

```
ceph osd set noup
ceph osd set nodown
```



중요

***noup** 및 **nodown** 플래그를 사용하면 문제의 근본 원인을 해결할 수 없지만 **OSD**만 회전하지 않습니다. 지원 티켓을 열려면 자세한 내용은 서비스 관련 **Red Hat** 지원 연락처 섹션을 참조하십시오.*

중요

***OSD**를 분리하면 네트워크 스위치 수준에서 **Ceph OSD** 노드의 **MTU** 잘못 구성 또는 둘 다로 인해 발생할 수 있습니다. 이 문제를 해결하려면 코어를 포함하여 모든 스토리지 클러스터 노드에서 **MTU**를 일정한 크기로 설정하고 예정된 다운타임을 사용하여 네트워크 스위치에 액세스합니다. 이 설정을 변경하면 네트워크 내에서 문제가 숨겨지며 실제 네트워크 불일치를 해결하지 않으므로 **osd heartbeat min** 크기를 조정하지 마십시오.*

추가 리소스

- 자세한 내용은 **Red Hat Ceph Storage Architecture Guide**의 **Ceph 하트비트** 섹션을 참조하십시오.

- **Red Hat Ceph Storage Troubleshooting Guide** 의 **Slow 요청 또는 요청이 차단된** 섹션을 참조하십시오.

5.2.8. 느린 요청 또는 요청이 차단됨

ceph-osd 데몬은 요청에 응답하는 속도가 느리고 **ceph** 상태 세부 정보 명령에서 다음과 유사한 오류 메시지를 반환합니다.

```
HEALTH_WARN 30 requests are blocked > 32 sec; 3 osds have slow requests
30 ops are blocked > 268435 sec
1 ops are blocked > 268435 sec on osd.11
1 ops are blocked > 268435 sec on osd.18
28 ops are blocked > 268435 sec on osd.39
3 osds have slow requests
```

또한 **Ceph** 로그에 다음과 유사한 오류 메시지가 포함됩니다.

```
2022-05-24 13:18:10.024659 osd.1 127.0.0.1:6812/3032 9 : cluster [WRN] 6 slow requests, 6
included below; oldest blocked for > 61.758455 secs
```

```
2022-05-25 03:44:06.510583 osd.50 [WRN] slow request 30.005692 seconds old, received at {date-
time}: osd_op(client.4240.0:8 benchmark_data_ceph-1_39426_object7 [write 0~4194304]
0.69848840) v4 currently waiting for subops from [610]
```

<그 Means>

느린 요청이 있는 **OSD**는 모든 **OSD**에서 **osd_op_complaint_time** 매개변수로 정의된 시간 내에 큐에서 초당 **I/O** 작업**(IOPS)**을 제공할 수 없습니다. 기본적으로 이 매개변수는 **30**초로 설정됩니다.

OSD에서 요청 속도가 느린 주요 원인은 다음과 같습니다.

- 디스크 드라이브, 호스트, 랙 또는 네트워크 스위치와 같은 기본 하드웨어 문제

- 네트워크에 문제가 있습니다. 이러한 문제는 일반적으로 **OSD** 플래핑과 연결되어 있습니다. 자세한 내용은 **Flapping OSD** 를 참조하십시오.

- *시스템 로드*

*다음 표에서는 느린 요청 유형을 보여줍니다. **dump_historic_ops** 관리 소켓 명령을 사용하여 느린 요청 유형을 확인합니다. 관리 소켓에 대한 자세한 내용은 **Red Hat** Ceph Storage 5 관리 가이드의 Ceph 관리 소켓 사용 섹션을 참조하십시오.*

느린 요청 유형	설명
rw 잠금 대기	OSD가 작업에 대한 배치 그룹에 대한 잠금을 획득하기를 기다리고 있습니다.
하위 프로세스 대기	OSD에서 복제 OSD가 작업을 저널에 적용할 때까지 대기 중입니다.
플래그 포인트가 도달하지 않았습니다.	OSD는 주요 작업 이정표에 도달하지 않았습니다.
성능 저하 오브젝트 대기	OSD가 지정된 수의 오브젝트를 아직 복제하지 않았습니다.

이 문제를 해결하기 위해

1. *느린 또는 차단 요청이 있는 **OSD**가 디스크 드라이브, 호스트, 랙 또는 네트워크 스위치와 같은 일반적인 하드웨어를 공유하는지 확인합니다.*

2. ***OSD**가 디스크를 공유하는 경우:*

   a. ***smartmontools** 유틸리티를 사용하여 디스크 또는 로그의 상태를 확인하여 디스크의 오류를 확인합니다.*

   참고

   ***smartmontools** 유틸리티는 **smartmontools** 패키지에 포함되어 있습니다.*

   b. ***OSD** 디스크에서 **I/O** 대기 보고서(**%iowai**)를 얻으려면 **iostat** 유틸리티를 사용하여 디스크가 과도한 부하인지 확인합니다.*

참고

iostat 유틸리티는 **tested** 패키지에 포함되어 있습니다.

3. **OSD**가 노드를 다른 서비스와 공유하는 경우:

   a. **RAM** 및 **CPU** 사용률 확인

   b. **netstat** 유틸리티를 사용하여 **NIC**(네트워크 인터페이스 컨트롤러)의 네트워크 통계를 확인하고 모든 네트워킹 문제를 해결합니다. 자세한 내용은 네트워크 문제 해결을 참조하십시오.

4. **OSD**가 랙을 공유하는 경우 랙의 네트워크 스위치를 확인합니다. 예를 들어 점보 프레임을 사용하는 경우 경로의 **NIC**에 점보 프레임이 설정되어 있는지 확인합니다.

5. 느린 요청이 있는 **OSD**에서 공유하는 공통 하드웨어를 확인하거나 하드웨어 및 네트워킹 문제를 해결 및 수정할 수 없는 경우 지원 티켓을 엽니다. 자세한 내용은 **Red Hat** 지원 문의를 참조하십시오.

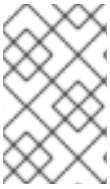
추가 리소스

- 자세한 내용은 **Red Hat Ceph Storage** 관리 가이드의 **Ceph** 관리소켓 사용 섹션을 참조하십시오.

5.3. 중지 및 재조정 시작

OSD가 실패하거나 중지하면 **CRUSH** 알고리즘은 나머지 **OSD**에서 데이터를 재배포하는 리밸런싱 프로세스를 자동으로 시작합니다.

리밸런싱은 시간과 리소스가 필요할 수 있으므로 문제 해결 또는 **OSD** 유지 관리 중에 재조정을 중지하는 것이 좋습니다.



참고

중지된 **OSD** 내의 배치 그룹은 문제 해결 및 유지 관리 중에 성능이 저하 됩니다.

사전 요구 사항

- ***Ceph Monitor** 노드에 대한 루트 수준 액세스입니다.*

절차

1. ***Cephadm** 쉘에 로그인합니다.*

   예제

   ```
   [root@host01 ~]# cephadm shell
   ```

2. ***OSD**를 중지하기 전에 **noout** 플래그를 설정합니다.*

   예제

   ```
   [ceph: root@host01 /]# ceph osd set noout
   ```

3. *문제 해결 또는 유지 관리를 마치면 **noout** 플래그를 해제하여 리밸런싱을 시작합니다.*

   예제

   ```
   [ceph: root@host01 /]# ceph osd unset noout
   ```

추가 리소스

- ***Red Hat Ceph Storage** 아키텍처 가이드 의 리밸런스 및 복구 섹션.*

5.4. OSD 데이터 파티션 마운트

***OSD** 데이터 파티션이 올바르게 마운트되지 않으면 **ceph-osd** 데몬을 시작할 수 없습니다. 파티션이 예상대로 마운트되지 않은 경우 이 섹션의 단계에 따라 마운트합니다.*

사전 요구 사항

- ***ceph-osd** 데몬에 액세스할 수 있습니다.*
- ***Ceph Monitor** 노드에 대한 루트 수준 액세스입니다.*

절차

1. *파티션을 마운트합니다.*

   구문

   ```
   mount -o noatime PARTITION /var/lib/ceph/osd/CLUSTER_NAME-OSD_NUMBER
   ```

   ***PARTITION** 을 **OSD** 데이터 전용 **OSD** 드라이브의 파티션 경로로 교체합니다. 클러스터 이름과 **OSD** 번호를 지정합니다.*

   예제

   ```
   [root@host01 ~]# mount -o noatime /dev/sdd1 /var/lib/ceph/osd/ceph-0
   ```

2. 실패한 **ceph-osd** 데몬을 시작합니다.

구문

```
systemctl start ceph-osd@OSD_NUMBER
```

OSD_NUMBER 를 **OSD ID**로 교체합니다.

예제

```
[root@host01 ~]# systemctl start ceph-osd@0
```

추가 리소스

- 자세한 내용은 **Red Hat Ceph Storage** 문제 해결 가이드에서 **Down OSD** 를 참조하십시오.

5.5. OSD 드라이브 교체

Ceph는 내결함성을 위해 설계되었습니다. 즉, 데이터가 손실되지 않고 성능이 저하된 상태에서 작동할 수 있습니다. 따라서 데이터 스토리지 드라이브가 실패하더라도 **Ceph**가 작동할 수 있습니다. 실패한 드라이브의 컨텍스트에서 **degraded** 상태는 다른 **OSD**에 저장된 데이터의 추가 복사본이 클러스터의 다른 **OSD**로 자동 다시 입력됨을 의미합니다. 그러나 이러한 경우 오류가 발생한 **OSD** 드라이브를 교체하고 **OSD**를 수동으로 다시 생성합니다.

드라이브가 실패하면 **Ceph**에서 **OSD**를 **down** 으로 보고합니다.

```
HEALTH_WARN 1/3 in osds are down
osd.0 is down since epoch 23, last address 192.168.106.220:6800/11080
```

참고

Ceph는 네트워킹 또는 권한 문제로 인해 OSD를 down 으로 표시할 수 있습니다. 자세한 내용은 OSD 다운 을 참조하십시오.

최신 서버는 일반적으로 hot-swappable 드라이브로 배포되므로 실패한 드라이브를 끌어서 노드를 중단하지 않고 새 드라이브로 교체할 수 있습니다. 전체 절차에는 다음 단계가 포함됩니다.

1. Ceph 클러스터에서 OSD를 제거합니다. 자세한 내용은 Ceph Cluster 프로시저에서 OSD 제거를 참조하십시오.

2. 드라이브를 교체하십시오. 자세한 내용은 물리적 드라이브 교체 섹션을 참조하십시오.

3. 클러스터에 OSD를 추가합니다. 자세한 내용은 Ceph 클러스터에 OSD 추가를 참조하십시오.

사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.
- Ceph Monitor 노드에 대한 루트 수준 액세스입니다.
- OSD가 한 개 이상 다운 되었습니다.

Ceph 클러스터에서 OSD 제거

1. Cephadm 쉘에 로그인합니다.

예제

```
[root@host01 ~]# cephadm shell
```

2. *다운 된 **OSD**를 확인합니다.*

예제

```
[ceph: root@host01 /]# ceph osd tree | grep -i down
ID  CLASS  WEIGHT   TYPE NAME          STATUS  REWEIGHT  PRI-AFF
 0   hdd 0.00999        osd.0     down  1.00000          1.00000
```

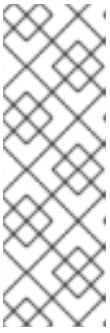
3. *클러스터가 데이터를 재조정하고 다른 **OSD**에 복사하기 위해 **OSD**를 밖으로 표시합니다.*

구문

```
ceph osd out OSD_ID.
```

예제

```
[ceph: root@host01 /]# ceph osd out osd.0
marked out osd.0.
```



참고

***OSD**가 다운되면 **Ceph**는 **mon_osd_ down _ out _interval** 매개변수에 따라 **OSD**에서 하트비트 패킷을 수신하지 않는 경우 **600**초 후에 자동으로 표시합니다. 이 경우 필요한 사본 수가 클러스터 내에 존재하는지 확인하기 위해 실패한 **OSD**에서 실패한 **OSD**를 백필하기 시작합니다. 클러스터가 다시 입력되는 동안 클러스터가 성능이 저하됩니다.*

4. *실패한 **OSD**가 다시 입력되었는지 확인합니다.*

예제

```
[ceph: root@host01 /]# ceph -w | grep backfill
2022-05-02 04:48:03.403872 mon.0 [INF] pgmap v10293282: 431 pgs: 1
active+undersized+degraded+remapped+backfilling, 28 active+undersized+degraded, 49
active+undersized+degraded+remapped+wait_backfill, 59 stale+active+clean, 294
active+clean; 72347 MB data, 101302 MB used, 1624 GB / 1722 GB avail; 227 kB/s rd, 1358
B/s wr, 12 op/s; 10626/35917 objects degraded (29.585%); 6757/35917 objects misplaced
(18.813%); 63500 kB/s, 15 objects/s recovering
2022-05-02 04:48:04.414397 mon.0 [INF] pgmap v10293283: 431 pgs: 2
active+undersized+degraded+remapped+backfilling, 75
active+undersized+degraded+remapped+wait_backfill, 59 stale+active+clean, 295
active+clean; 72347 MB data, 101398 MB used, 1623 GB / 1722 GB avail; 969 kB/s rd, 6778
B/s wr, 32 op/s; 10626/35917 objects degraded (29.585%); 10580/35917 objects misplaced
(29.457%); 125 MB/s, 31 objects/s recovering
2022-05-02 04:48:00.380063 osd.1 [INF] 0.6f starting backfill to osd.0 from (0'0,0'0] MAX to
2521'166639
2022-05-02 04:48:00.380139 osd.1 [INF] 0.48 starting backfill to osd.0 from (0'0,0'0] MAX to
2513'43079
2022-05-02 04:48:00.380260 osd.1 [INF] 0.d starting backfill to osd.0 from (0'0,0'0] MAX to
2513'136847
2022-05-02 04:48:00.380849 osd.1 [INF] 0.71 starting backfill to osd.0 from (0'0,0'0] MAX to
2331'28496
2022-05-02 04:48:00.381027 osd.1 [INF] 0.51 starting backfill to osd.0 from (0'0,0'0] MAX to
2513'87544
```

*배치 그룹 상태가 **active+clean** 에서 활성 **,** 일부 저하된 오브젝트**,** 마지막으로 마이그레이션이 완료되면 **active +clean** 으로 변경되는 것을 확인해야 합니다.*

5. ***OSD**를 중지합니다.*

구문

```
ceph orch daemon stop OSD_ID
```

예제

```
[ceph: root@host01 /]# ceph orch daemon stop osd.0
```

6. *스토리지 클러스터에서 **OSD**를 제거합니다.*

구문

```
ceph orch osd rm OSD_ID --replace
```

예제

```
[ceph: root@host01 /]# ceph orch osd rm 0 --replace
```

***OSD_ID** 가 보존됩니다.*

물리적 드라이브 교체

물리 드라이브 교체에 대한 자세한 내용은 하드웨어 노드의 설명서를 참조하십시오.

1. *드라이브가 **hot-swappable**인 경우 오류가 발생한 드라이브를 새 드라이브로 교체합니다.*

2. *드라이브가 핫 스왑이 아니며 노드에 여러 **OSD**가 포함된 경우 전체 노드를 종료하고 물리적 드라이브를 교체해야 할 수 있습니다. 클러스터가 다시 입력되지 않도록 합니다. 자세한 내용은 **Red Hat Ceph Storage** 문제 해결 가이드 의 중지 및 재조정 장을 참조하십시오.*

3. *드라이브가 /dev/ 디렉토리 아래에 표시되면 드라이브 경로를 기록해 둡니다.*

4. *OSD를 수동으로 추가하려면 OSD 드라이브를 찾아 디스크를 포맷하십시오.*

Ceph 클러스터에 OSD 추가

1. *새 드라이브가 삽입되면 다음 옵션을 사용하여 OSD를 배포할 수 있습니다.*

- *--unmanaged 매개 변수가 설정되지 않은 경우 OSD는 Ceph Orchestrator에서 자동으로 배포합니다.*

  예제

  ```
  [ceph: root@host01 /]# ceph orch apply osd --all-available-devices
  ```

- *unmanaged 매개 변수가 true 로 설정된 사용 가능한 모든 장치에 OSD를 배포합니다.*

  예제

  ```
  [ceph: root@host01 /]# ceph orch apply osd --all-available-devices --unmanaged=true
  ```

- *특정 장치 및 호스트에서 OSD를 배포합니다.*

  예제

  ```
  [ceph: root@host01 /]# ceph orch daemon add osd host02:/dev/sdb
  ```

2. **CRUSH 계층 구조가 정확한지 확인합니다.**

예제

```
[ceph: root@host01 /]# ceph osd tree
```

추가 리소스

- **Red Hat Ceph Storage Operations Guide의 사용 가능한 모든 장치에서 Ceph OSD 배포 섹션을** 참조하십시오.
- **Red Hat Ceph Storage 운영 가이드의 특정 장치 및 호스트에서 Ceph OSD 배포 섹션을 참조하십시오.**
- **Red Hat Ceph Storage 문제 해결 가이드의 Down OSD 섹션을 참조하십시오.**
- **Red Hat Ceph Storage 설치 가이드를 참조하십시오.**

5.6. PID 수 증가

Ceph OSD가 12개 이상인 노드가 있는 경우 특히 복구 중에 기본 최대 스레드 수(PID 수)가 충분하지 않을 수 있습니다. 결과적으로 일부 ceph-osd 데몬은 종료하여 다시 시작하지 못할 수 있습니다. 이 경우 허용되는 최대 스레드 수를 늘립니다.

절차

임시로 번호를 늘리려면 다음을 수행하십시오.

```
[root@mon ~]# sysctl -w kernel.pid.max=4194303
```

수를 영구적으로 늘리려면 **/etc/sysctl.conf** 파일을 다음과 같이 업데이트합니다.

```
kernel.pid.max = 4194303
```

5.7. 전체 스토리지 클러스터에서 데이터 삭제

Ceph에서는 **mon_osd_full_ratio** 매개 변수에서 지정한 용량에 도달한 **OSD**의 **I/O** 작업을 자동으로 방지하고 전체 **osds** 오류 메시지를 반환합니다.

다음 절차에서는 불필요한 데이터를 삭제하여 이 오류를 해결하는 방법을 설명합니다.

참고

mon_osd_full_ratio 매개변수는 클러스터를 생성할 때 **full_ratio** 매개변수의 값을 설정합니다. **mon_osd_full_ratio** 의 값은 나중에 변경할 수 없습니다. **full_ratio** 값을 일시적으로 늘리려면 대신 **set-full-ratio** 를 늘리십시오.

사전 요구 사항

- **Ceph Monitor** 노드에 대한 루트 수준 액세스입니다.

절차

1. **Cephadm** 쉘에 로그인합니다.

   예제

   ```
   [root@host01 ~]# cephadm shell
   ```

2. **full_ratio** 의 현재 값을 결정하십시오. 기본적으로 **0.95** 로 설정되어 있습니다.

   ```
   [ceph: root@host01 /]# ceph osd dump | grep -i full
   full_ratio 0.95
   ```

3. 일시적으로 **set-full-ratio** 값을 **0.97** 로 늘립니다.

```
[ceph: root@host01 /]# ceph osd set-full-ratio 0.97
```

중요

Red Hat은 **set-full-ratio** 를 **0.97**보다 큰 값으로 설정하지 않는 것이 좋습니다. 이 매개 변수를 더 높은 값으로 설정하면 복구 프로세스가 더 어려워집니다. 결과적으로 전체 **OSD**를 전혀 복구하지 못할 수 있습니다.

4. 매개 변수를 **0.97** 로 성공적으로 설정했는지 확인합니다.

```
[ceph: root@host01 /]# ceph osd dump | grep -i full
full_ratio 0.97
```

5. 클러스터 상태를 모니터링합니다.

```
[ceph: root@host01 /]# ceph -w
```

클러스터가 상태를 완전히 완전히 변경하는 즉시 불필요한 데이터를 삭제합니다.

6. **full_ratio** 의 값을 **0.95** 로 다시 설정합니다.

```
[ceph: root@host01 /]# ceph osd set-full-ratio 0.95
```

7. 매개 변수를 **0.95** 로 성공적으로 설정했는지 확인합니다.

```
[ceph: root@host01 /]# ceph osd dump | grep -i full
full_ratio 0.95
```

추가 리소스

- **Red Hat Ceph Storage** 문제 해결 가이드 의 전체 **OSD** 섹션 .
- **Red Hat Ceph Storage** 문제 해결 가이드 의 전체 **OSD** 가까운 섹션 .

6장. 다중 사이트 CEPH 오브젝트 게이트웨이 문제 해결

이 장에서는 다중 사이트 Ceph 개체 게이트웨이 구성 및 운영 조건과 관련된 가장 일반적인 오류를 수정하는 방법을 설명합니다.

6.1. 사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.
- 실행 중인 Ceph 오브젝트 게이트웨이.

6.2. CEPH OBJECT GATEWAY에 대한 오류 코드 정의

Ceph Object Gateway 로그에는 환경의 상태 문제를 해결하는 데 도움이 되는 오류 및 경고 메시지가 포함되어 있습니다. 몇 가지 일반적인 해결 방법이 아래에 나열되어 있으며 이러한 해결 사항은 다음과 같습니다.

일반적인 오류 메시지

data_sync: ERROR: 동기화 작업에서 반환된 오류

이는 하위 수준 버킷 동기화 프로세스에서 오류를 반환했다고 발생하는 높은 수준의 데이터 동기화 프로세스입니다. 이 메시지는 중복되어 있습니다. 버킷 동기화 오류가 로그에 표시됩니다.

데이터 동기화: **ERROR: failed to sync object: BUCKET_NAME:_OBJECT_NAME**

프로세스가 원격 게이트웨이에서 **HTTP**를 통해 필요한 오브젝트를 가져오지 못하거나 프로세스가 **RADOS**에 해당 오브젝트를 쓸 수 없어 다시 시도합니다.

데이터 동기화: **ERROR: failure in sync, backing out (sync_status=2)**

위의 조건 중 하나를 반영하는 낮은 수준의 메시지는 특히 데이터가 동기화되기 전에 삭제되었다는 메시지를 반영하여 **-2 ENOENT** 상태를 표시합니다.

데이터 동기화: **ERROR: failure in sync, backing out (sync_status=-5)**

위 조건 중 하나를 반영하는 낮은 수준의 메시지는 특히 해당 오브젝트를 **RADOS**에 쓰지 못하여 **-5 EIO** 를 표시합니다.

오류: 원격 데이터 로그 정보를 가져오지 못했습니다. **ret=11**

이는 다른 게이트웨이 의 오류 조건을 반영하는 **libcurl** 의 일반 오류 코드입니다. 기본적으로 다시 시도합니다.

메타 동기화: 오류: (2) 이러한 파일 또는 디렉토리 없이 mdlog 정보를 읽을 수 없습니다

mdlog의 shard가 생성되지 않았으므로 동기화할 항목이 없습니다.

오류 메시지 동기화

오브젝트 동기화 실패

프로세스가 원격 게이트웨이에서 HTTP를 통해 이 오브젝트를 가져오지 못하거나 RADOS에 해당 오브젝트를 쓰지 못하며 다시 시도합니다.

버킷 인스턴스를 동기화하지 못했습니다. (11) 리소스를 일시적으로 사용할 수 없음

기본 영역과 보조 영역 간의 연결 문제입니다.

버킷 인스턴스를 동기화하지 못했습니다. (125) 작업이 취소됨

동일한 RADOS 오브젝트에 대한 쓰기 작업 사이에 racing 조건이 있습니다.

추가 리소스

- *추가 지원을 받으려면 Red Hat 지원팀 에 문의하십시오.*

6.3. 다중 사이트 CEPH 오브젝트 게이트웨이 동기화

다중 사이트 동기화는 다른 영역에서 변경 로그를 읽습니다. 메타데이터 및 데이터 로그에서 동기화 진행 상황을 대략적으로 보려면 다음 명령을 사용할 수 있습니다.

예제

```
[ceph: root@host01 /]# radosgw-admin sync status
```

이 명령은 소스 영역 뒤에 있는 로그 shard(있는 경우)를 나열합니다.

위에 실행한 동기화 상태 결과가 로그 shard 뒤에 있는 경우 X 에 대해 shard-id를 대체합니다.

구문

```
radosgw-admin data sync status --shard-id=X --source-zone=ZONE_NAME
```

예제

```
[ceph: root@host01 /]# radosgw-admin data sync status --shard-id=27 --source-zone=us-east
{
  "shard_id": 27,
  "marker": {
        "status": "incremental-sync",
        "marker": "1_1534494893.816775_131867195.1",
        "next_step_marker": "",
        "total_entries": 1,
        "pos": 0,
        "timestamp": "0.000000"
  },
   "pending_buckets": [],
   "recovering_buckets": [
        "pro-registry:4ed07bb2-a80b-4c69-aa15-fdc17ae6f5f2.314303.1:26"
   ]
}
```

동기화 옆에 있는 버킷과 이전 오류로 인해 다시 시도할 버킷이 출력됩니다.

*다음 명령을 사용하여 개별 버킷의 상태를 검사하고 **X** 의 버킷 **ID**를 대체합니다.*

구문

```
radosgw-admin bucket sync status --bucket=X.
```

X 를 버킷의 ID 번호로 바꿉니다.

그 결과 소스 영역 뒤에 있는 버킷 인덱스 로그 shard가 표시됩니다.

동기화의 일반적인 오류는 enterpriseUSY입니다. 즉, 동기화가 이미 진행 중인 경우 다른 게이트웨이에서 동기화가 이미 진행 중임을 나타냅니다. 다음 명령을 사용하여 읽을 수 있는 동기화 오류 로그에 기록된 오류를 읽습니다.

```
radosgw-admin sync error list
```

동기화 프로세스가 성공할 때까지 다시 시도합니다. 여전히 개입이 필요할 수 있는 오류가 발생할 수 있습니다.

6.3.1. 멀티 사이트 Ceph Object Gateway 데이터 동기화에 대한 성능 카운터

다음 성능 카운터는 데이터 동기화를 측정하기 위해 Ceph Object Gateway의 다중 사이트 구성에 사용할 수 있습니다.

- *poll_latency 는 원격 복제 로그에 대한 요청의 대기 시간을 측정합니다.*
- *fetch_bytes 는 데이터 동기화에서 가져온 오브젝트 및 바이트 수를 측정합니다.*

ceph --admin-daemon 명령을 사용하여 성능 카운터의 현재 지표 데이터를 확인합니다.

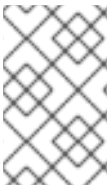
구문

```
ceph --admin-daemon /var/run/ceph/ceph-client.rgw.RGW_ID.asok perf dump data-sync-from-ZONE_NAME
```

예제

```
[ceph: root@host01 /]# ceph --admin-daemon /var/run/ceph/ceph-client.rgw.host02-
rgw0.103.94309060818504.asok perf dump data-sync-from-us-west

{
    "data-sync-from-us-west": {
        "fetch bytes": {
            "avgcount": 54,
            "sum": 54526039885
        },
        "fetch not modified": 7,
        "fetch errors": 0,
        "poll latency": {
            "avgcount": 41,
            "sum": 2.533653367,
            "avgtime": 0.061796423
        },
        "poll errors": 0
    }
}
```



참고

데몬을 실행하는 노드에서 **ceph --admin-daemon** 명령을 실행해야 합니다.

추가 리소스

- 성능 카운터에 대한 자세한 내용은 **Red Hat Ceph Storage** 관리 가이드 의 **Ceph** 성능 카운터 장을 참조하십시오.

7장. CEPH ISCSI 게이트웨이 문제 해결 (LIMITED AVAILABILITY)

스토리지 관리자는 Ceph iSCSI 게이트웨이를 사용할 때 발생할 수 있는 대부분의 일반적인 오류 문제를 해결할 수 있습니다. 다음은 발생할 수 있는 몇 가지 일반적인 오류입니다.

- iSCSI 로그인 문제.
- VMware ESXi에서 다양한 연결 실패를 보고합니다.
- 시간 초과 오류.

참고

이 기술은 제한적인 가용성입니다. 자세한 내용은 사용 중단된 기능 장을 참조하십시오.

7.1. 사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.
- 실행 중인 Ceph iSCSI 게이트웨이.
- 네트워크 연결을 확인합니다.

7.2. VMWARE ESXI에서 스토리지 오류가 발생하는 손실된 연결에 대한 정보 수집

시스템 및 디스크 정보를 수집하면 연결이 손실되어 스토리지 장애를 유발하는 iSCSI 대상이 무엇인지 결정하는 데 도움이 됩니다. 필요한 경우 Ceph iSCSI 게이트웨이 문제 해결을 지원하기 위해 이 정보를 Red Hat의 글로벌 지원 서비스에 제공할 수도 있습니다.

사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.

- ***iSCSI** 대상인 실행 중인 **Ceph iSCSI** 게이트웨이.*
- ***iSCSI** 이니시에이터의 실행 중인 **VMware ESXi** 환경.*
- ***VMware ESXi** 노드에 대한 루트 수준 액세스.*

절차

1. ***VWware ESXi** 노드에서 커널 로그를 엽니다.*

```
[root@esx:~]# more /var/log/vmkernel.log
```

2. ***VMware ESXi** 커널 로그에서 다음 오류 메시지에서 정보를 수집합니다.*

예제

```
2022-05-30T11:07:07.570Z cpu32:66506)iscsi_vmk:
iscsivmk_ConnRxNotifyFailure: Sess [ISID: 00023d000005 TARGET:
iqn.2017-12.com.redhat.iscsi-gw:ceph-igw TPGT: 3 TSIH: 0]
```

*이 메시지에서 **ISID** 번호, **CACHEGET** 이름, 대상 포털 그룹 태그(**TPGT**) 번호를 기록해 둡니다. 이 예제에서는 다음과 같습니다.*

```
ISID: 00023d000005
TARGET: iqn.2017-12.com.redhat.iscsi-gw:ceph-igw
TPGT: 3
```

예제

```
2022-05-30T11:07:07.570Z cpu32:66506)iscsi_vmk:
iscsivmk_ConnRxNotifyFailure: vmhba64:CH:4 T:0 CN:0: Connection rx
notifying failure: Failed to Receive. State=Bound
```

*이 메시지에서 어댑터 채널(**CH**) 번호를 기록해 둡니다. 이 예제에서는 다음과 같습니다.*

```
vmhba64:CH:4 T:0
```

3. ***Ceph iSCSI** 게이트웨이 노드의 원격 주소를 찾으려면 다음을 수행합니다.*

```
[root@esx:~]# esxcli iscsi session connection list
```

예제

```
...
vmhba64,iqn.2017-12.com.redhat.iscsi-gw:ceph-igw,00023d000003,0
   Adapter: vmhba64
   Target: iqn.2017-12.com.redhat.iscsi-gw:ceph-igw (1)
   ISID: 00023d000003 (2)
   CID: 0
   DataDigest: NONE
   HeaderDigest: NONE
   IFMarker: false
   IFMarkerInterval: 0
   MaxRecvDataSegmentLength: 131072
   MaxTransmitDataSegmentLength: 262144
   OFMarker: false
   OFMarkerInterval: 0
   ConnectionAddress: 10.2.132.2
   RemoteAddress: 10.2.132.2 (3)
   LocalAddress: 10.2.128.77
   SessionCreateTime: 03/28/18 21:45:19
   ConnectionCreateTime: 03/28/18 21:45:19
   ConnectionStartTime: 03/28/18 21:45:19
   State: xpt_wait
...
```

*명령 출력에서 **ISID** 값과 이전에 수집된 **CACHE GET** 이름 값과 일치한 다음 **RemoteAddress** 값을 기록합니다. 이 예에서는 다음이 있습니다.*

```
Target: iqn.2017-12.com.redhat.iscsi-gw:ceph-igw
ISID: 00023d000003
RemoteAddress: 10.2.132.2
```

*이제 **Ceph iSCSI** 게이트웨이 노드에서 더 많은 정보를 수집하여 문제를 추가로 해결할 수 있습니다.*

a. *__RemoteAddress__ 값이 언급된 __Ceph iSCSI__ 게이트웨이 노드에서 __sosreport__ 를 실행하여 시스템 정보를 수집합니다.*

```
[root@igw ~]# sosreport
```

4. *__dead__ 상태로 전환된 디스크를 찾으려면 다음을 수행합니다.*

```
[root@esx:~]# esxcli storage nmp device list
```

예제

```
...
iqn.1998-01.com.vmware:d04-nmgjd-pa-zyc-sv039-rh2288h-xnh-732d78fd-
00023d000004,iqn.2017-12.com.redhat.iscsi-gw:ceph-igw,t,3-
naa.60014054a5d46697f85498e9a257567c
   Runtime Name: vmhba64:C4:T0:L4 (1)
   Device: naa.60014054a5d46697f85498e9a257567c (2)
   Device Display Name: LIO-ORG iSCSI Disk
(naa.60014054a5d46697f85498e9a257567c)
   Group State: dead (3)
   Array Priority: 0
   Storage Array Type Path Config:
{TPG_id=3,TPG_state=ANO,RTP_id=3,RTP_health=DOWN} (4)
   Path Selection Policy Path Config: {non-current path; rank: 0}
...
```

명령 출력에서 __CH__ 번호와 이전에 수집한 __TPGT__ 번호와 일치한 다음 __Device__ 값을 기록합니다. 이 예제에서는 다음과 같습니다.

```
vmhba64:C4:T0
Device: naa.60014054a5d46697f85498e9a257567c
TPG_id=3
```

*장치 이름을 사용하면 각 **iSCSI** 디스크에 대한 추가 정보를 **dead** 상태로 수집할 수 있습니다.*

a. ***iSCSI** 디스크에 대한 자세한 정보를 수집합니다.*

구문

```
esxcli storage nmp path list -d ISCSI_DISK_DEVICE >
/tmp/esxcli_storage_nmp_path_list.txt
esxcli storage core device list -d ISCSI_DISK_DEVICE >
/tmp/esxcli_storage_core_device_list.txt
```

예제

```
[root@esx:~]# esxcli storage nmp path list -d naa.60014054a5d46697f85498e9a257567c
> /tmp/esxcli_storage_nmp_path_list.txt
[root@esx:~]# esxcli storage core device list -d
naa.60014054a5d46697f85498e9a257567c > /tmp/esxcli_storage_core_device_list.txt
```

5. ***VMware ESXi** 환경에서 추가 정보를 수집합니다.*

```
[root@esx:~]# esxcli storage vmfs extent list > /tmp/esxcli_storage_vmfs_extent_list.txt
[root@esx:~]# esxcli storage filesystem list > /tmp/esxcli_storage_filesystem_list.txt
[root@esx:~]# esxcli iscsi session list > /tmp/esxcli_iscsi_session_list.txt
[root@esx:~]# esxcli iscsi session connection list >
/tmp/esxcli_iscsi_session_connection_list.txt
```

6. *가능한 **iSCSI** 로그인 문제가 있는지 확인합니다.*

- ***iSCSI** 로그인 데이터가 전송되지 않았습니다.*

- *iSCSI 로그인 시간 초과 또는 포털 그룹을 찾지 못했습니까?*

추가 리소스

- *Red Hat 글로벌 지원 서비스에 대한 sosreport를 생성하는 Red Hat의 지식베이스 솔루션을 참조하십시오.*
- *Red Hat 글로벌 지원 서비스용 파일 업로드 시 Red Hat의 지식베이스 솔루션을 참조하십시오.*
- *고객 포털에서 Red Hat 지원 케이스 를 여는 방법은 무엇입니까?*

7.3. 데이터가 전송되지 않았기 때문에 ISCSI 로그인 실패 확인

iSCSI 게이트웨이 노드에서 기본적으로 /var/log/messages 에 시스템 로그에 일반 로그인 협상 실패 메시지가 표시될 수 있습니다.

예제

```
Apr  2 23:17:05 osd1 kernel: rx_data returned 0, expecting 48.
Apr  2 23:17:05 osd1 kernel: iSCSI Login negotiation failed.
```

시스템이 이 상태에 있는 동안 이 절차에 설명된 대로 시스템 정보 수집을 시작합니다.

사전 요구 사항

- *실행 중인 Red Hat Ceph Storage 클러스터.*
- *iSCSI 대상인 실행 중인 Ceph iSCSI 게이트웨이.*

- ***iSCSI** 이니시에이터의 실행 중인 **VMware ESXi** 환경.*
- ***Ceph iSCSI** 게이트웨이 노드에 대한 루트 수준 액세스.*
- ***VMware ESXi** 노드에 대한 루트 수준 액세스.*

절차

1. *추가 로깅을 활성화합니다.*

```
[root@igw ~]# echo "iscsi_target_mod +p" > /sys/kernel/debug/dynamic_debug/control
[root@igw ~]# echo "target_core_mod +p" > /sys/kernel/debug/dynamic_debug/control
```

2. *추가 디버깅 정보가 시스템 로그를 채울 때까지 몇 분 정도 기다립니다.*
3. *추가 로깅을 비활성화합니다.*

```
[root@igw ~]# echo "iscsi_target_mod -p" > /sys/kernel/debug/dynamic_debug/control
[root@igw ~]# echo "target_core_mod -p" > /sys/kernel/debug/dynamic_debug/control
```

4. ***sosreport** 를 실행하여 시스템 정보를 수집합니다.*

```
[root@igw ~]# sosreport
```

5. ***Ceph iSCSI** 게이트웨이 및 **VMware ESXi** 노드의 네트워크 트래픽을 동시에 캡처합니다.*

구문

```
tcpdump -s0 -i NETWORK_INTERFACE -w OUTPUT_FILE_PATH
```

예제

```
[root@igw ~]# tcpdump -s 0 -i eth0 -w /tmp/igw-eth0-tcpdump.pcap
```

참고

*포트 **3260**에서 트래픽을 찾습니다.*

a.

*네트워크 패킷 캡처 파일이 클 수 있으므로 파일을 **Red Hat** 글로벌 지원 서비스에 업로드하기 전에 **iSCSI** 대상 및 이니시에이터에서 **tcpdump** 출력을 압축할 수 있습니다.*

구문

```
gzip OUTPUT_FILE_PATH
```

예제

```
[root@igw ~]# gzip /tmp/igw-eth0-tcpdump.pcap
```

6.

***VMware ESXi** 환경에서 추가 정보를 수집합니다.*

```
[root@esx:~]# esxcli iscsi session list > /tmp/esxcli_iscsi_session_list.txt
[root@esx:~]# esxcli iscsi session connection list >
/tmp/esxcli_iscsi_session_connection_list.txt
```

a.

*각 **iSCSI** 디스크에 대한 추가 정보를 나열하고 수집합니다.*

구문

```
esxcli storage nmp path list -d ISCSI_DISK_DEVICE >
/tmp/esxcli_storage_nmp_path_list.txt
```

예제

```
[root@esx:~]# esxcli storage nmp device list
[root@esx:~]# esxcli storage nmp path list -d naa.60014054a5d46697f85498e9a257567c
> /tmp/esxcli_storage_nmp_path_list.txt
[root@esx:~]# esxcli storage core device list -d
naa.60014054a5d46697f85498e9a257567c > /tmp/esxcli_storage_core_device_list.txt
```

추가 리소스

- ***Red Hat** 글로벌 지원 서비스에 대한 **sosreport**를 생성하는 **Red Hat**의 지식베이스 솔루션을 참조하십시오.*
- ***Red Hat** 글로벌 지원 서비스용 파일 업로드 시 **Red Hat**의 지식베이스 솔루션을 참조하십시오.*
- *자세한 내용은 **tcpdump**를 사용하여 네트워크 패킷을 캡처하는 방법에 대한 **Red Hat**의 지식베이스 솔루션을 참조하십시오.*
- *고객 포털에서 **Red Hat** 지원 케이스 를 여는 방법은 무엇입니까?*

7.4. 시간 초과 또는 포털 그룹을 찾을 수 없기 때문에 ISCSI 로그인 실패 확인

***iSCSI** 게이트웨이 노드의 시간 초과가 표시되거나 시스템 로그에서 대상 포털 그룹 메시지를 찾을 수 없는 경우 기본적으로 **/var/log/messages** 가 표시됩니다.*

예제

```
Mar 28 00:29:01 osd2 kernel: iSCSI Login timeout on Network Portal 10.2.132.2:3260
```

또는

예제

```
Mar 23 20:25:39 osd1 kernel: Unable to locate Target Portal Group on iqn.2017-12.com.redhat.iscsi-gw:ceph-igw
```

시스템이 이 상태에 있는 동안 이 절차에 설명된 대로 시스템 정보 수집을 시작합니다.

사전 요구 사항

- *실행 중인 **Red Hat Ceph Storage** 클러스터.*
- *실행 중인 **Ceph iSCSI** 게이트웨이.*
- ***Ceph iSCSI** 게이트웨이 노드에 대한 루트 수준 액세스.*

절차

1. *대기 중인 작업의 덤프를 활성화하고 파일에 씁니다.*

```
[root@igw ~]# dmesg -c ; echo w > /proc/sysrq-trigger ; dmesg -c > /tmp/waiting-tasks.txt
```

2. 다음 메시지에 대한 대기 중인 작업 목록을 검토합니다.

- **iscsit_tpg_disable_portal_group**
- **core_tmr_abort_task**
- **transport_generic_free_cmd**

이러한 메시지가 대기 중인 작업 목록에 표시되면 **tcmu-runner** 서비스에서 문제가 발생했음을 나타냅니다. **tcmu-runner** 서비스가 제대로 다시 시작되지 않았거나 **tcmu-runner** 서비스가 충돌했을 수 있습니다.

3. **tcmu-runner** 서비스가 실행 중인지 확인합니다.

```
[root@igw ~]# systemctl status tcmu-runner
```

a. **tcmu-runner** 서비스가 실행되고 있지 않은 경우 **tcmu-runner** 서비스를 다시 시작하기 전에 **rbd-target-gw** 서비스를 중지하십시오.

```
[root@igw ~]# systemctl stop rbd-target-gw
[root@igw ~]# systemctl stop tcmu-runner
[root@igw ~]# systemctl start tcmu-runner
[root@igw ~]# systemctl start rbd-target-gw
```

중요

Ceph iSCSI 게이트웨이를 중지하면 **tcmu-runner** 서비스가 중단된 동안 **IO**가 정지되지 않습니다.

b. **tcmu-runner** 서비스가 실행 중인 경우 새 버그가 될 수 있습니다. 새로운 **Red Hat** 지원 케이스를 개설합니다.

추가 리소스

- **Red Hat** 글로벌 지원 서비스에 대한 **sosreport**를 생성하는 **Red Hat**의 지식베이스 솔루션을 참조하십시오.

- *Red Hat 글로벌 지원 서비스용 파일 업로드 시 Red Hat의 지식베이스 솔루션을 참조하십시오.*

- *고객 포털에서 Red Hat 지원 케이스 를 여는 방법은 무엇입니까?*

7.5. 시간 초과 명령 오류

시스템 로그에 SCSI 명령이 실패한 경우 Ceph iSCSI 게이트웨이에서 명령 시간 초과 오류를 보고할 수 있습니다.

예제

```
Mar 23 20:03:14 igw tcmu-runner: 2018-03-23 20:03:14.052 2513 [ERROR]
tcmu_rbd_handle_timedout_cmd:669 rbd/rbd.gw1lun011: Timing out cmd.
```

또는

예제

```
Mar 23 20:03:14 igw tcmu-runner: tcmu_notify_conn_lost:176 rbd/rbd.gw1lun011: Handler
connection lost (lock state 1)
```

<그 Means>

처리 대기 중인 다른 작업이 있을 수 있으므로 응답이 적시에 수신되지 않았기 때문에 SCSI 명령이 시간 초과될 수 있습니다. 이러한 오류 메시지의 또 다른 이유는 비정상적인 Red Hat Ceph Storage 클러스터와 관련이 있을 수 있습니다.

이 문제를 해결하기 위해

1. 작업을 유지할 수 있는 대기 중인 작업이 있는지 확인합니다.

2. **Red Hat Ceph Storage** 클러스터의 상태를 확인합니다.

3. **Ceph iSCSI** 게이트웨이 노드에서 **iSCSI** 이니시에이터 노드로 경로의 각 장치에서 시스템 정보를 수집합니다.

추가 리소스

- 대기 시간 초과로 인해 **iSCSI** 로그인 오류 확인 또는 대기 중인 작업을 보는 방법에 대한 자세한 내용은 **Red Hat Ceph Storage** 문제 해결 가이드 의 포털 그룹 섹션을 참조하십시오.

- 스토리지 클러스터 상태를 확인하는 방법에 대한 자세한 내용은 **Red Hat Ceph Storage** 문제 해결 가이드 의 스토리지 클러스터 상태 진단 섹션을 참조하십시오.

- 필요한 정보를 수집하는 방법에 대한 자세한 내용은 **Red Hat Ceph Storage** 문제 해결 가이드 의 **VMware ESXi**에서 스토리지 실패를 유발하는 연결에 대한 정보 수집을 참조하십시오.

7.6. 중단 작업 오류

Ceph iSCSI 게이트웨이에서 시스템 로그에 중단 작업 오류를 보고할 수 있습니다.

예제

```
Apr  1 14:23:58 igw kernel: ABORT_TASK: Found referenced iSCSI task_tag: 1085531
```

<그 **Means**>

실패한 스위치 또는 잘못된 포트와 같은 다른 네트워크 중단으로 인해 이러한 유형의 오류 메시지가 발생할 수 있습니다. 또 다른 가능성은 비정상적인 **Red Hat Ceph Storage** 클러스터입니다.

이 문제를 해결하기 위해

1. 환경의 네트워크 중단을 확인합니다.

2. Red Hat Ceph Storage 클러스터의 상태를 확인합니다.

3. Ceph iSCSI 게이트웨이 노드에서 iSCSI 이니시에이터 노드로 경로의 각 장치에서 시스템 정보를 수집합니다.

추가 리소스

- 스토리지 클러스터 상태를 확인하는 방법에 대한 자세한 내용은 Red Hat Ceph Storage 문제 해결 가이드 의 스토리지 클러스터 상태 진단 섹션을 참조하십시오.

- 필요한 정보를 수집하는 방법에 대한 자세한 내용은 Red Hat Ceph Storage 문제 해결 가이드 의 VMware ESXi에서 스토리지 실패를 유발하는 연결에 대한 정보 수집을 참조하십시오.

7.7. 추가 리소스

- Ceph iSCSI 게이트웨이에 대한 자세한 내용은 Red Hat Ceph Storage Block Device 가이드 를 참조하십시오.

- 자세한 내용은 3장. 네트워킹 문제 해결 를 참조하십시오.

8장. CEPH 배치 그룹 문제 해결

이 섹션에는 Ceph PG(배치 배치 그룹)와 관련된 가장 일반적인 오류를 수정하는 방법에 대한 정보가 포함되어 있습니다.

8.1. 사전 요구 사항

- 네트워크 연결을 확인합니다.
- 모니터가 쿼럼을 형성할 수 있는지 확인합니다.
- 모든 정상 OSD가 up 및 in, 및 backfilling 및 복구 프로세스가 완료되었는지 확인합니다.

8.2. 대부분의 일반적인 CEPH 배치 그룹 오류

다음 표에는 ceph 상태 세부 명령에서 반환되는 일반적인 오류 메시지가 나열되어 있습니다. 표는 오류를 설명하고 문제를 해결하기 위한 특정 절차를 가리키는 해당 섹션에 대한 링크를 제공합니다.

또한 최적의 상태가 아닌 배치 그룹을 나열할 수 있습니다. 자세한 내용은 8.3절. “오래된,비활성 또는 불명확 상태에 있는 배치 그룹 나열 ”를 참조하십시오.

8.2.1. 사전 요구 사항

- 실행 중인 Red Hat Ceph Storage 클러스터.
- 실행 중인 Ceph 오브젝트 게이트웨이.

8.2.2. 배치 그룹 오류 메시지

일반적인 배치 그룹 오류 메시지 테이블 및 잠재적인 수정 사항.

오류 메시지	참조
HEALTH_ERR	

오류 메시지	참조
PGS 아래로	배치 그룹이 **다운**됨
PGS 일관되지 않음	일관성 없는 배치 그룹
crub 오류	일관성 없는 배치 그룹
HEALTH_WARN	
PGS stale	오래된 배치 그룹
unfound	unfound 오브젝트

8.2.3. 오래된 배치 그룹

***ceph health** 명령은 일부 **PG(배치 그룹)**를 오래된 것으로 나열합니다.*

> *HEALTH_WARN 24 pgs stale; 3/300 in osds are down*

<그 Means>

*모니터는 배치 그룹 작동 세트의 기본 **OSD**에서 또는 기본 **OSD**가 기본 **OSD**가 다운 되었다고 보고되는 경우 배치 그룹을 오래된 것으로 표시합니다.*

*일반적으로 **PG**는 스토리지 클러스터를 시작한 후 피어링 프로세스가 완료될 때까지 이전 상태로 들어갑니다. 그러나 **PG**가 예상보다 오래 동안 오래 상태로 유지되는 경우 해당 **PG**의 기본 **OSD**가 다운 되었거나 **PG** 통계를 모니터에 보고하지 않음을 나타낼 수 있습니다. 오래된 **PG**를 저장하는 기본 **OSD**가 백업 되면 **Ceph**가 **PG**를 복구하기 시작합니다.*

***mon_osd_report_timeout** 설정은 **OSD**에서 **PGs** 통계를 모니터에 보고하는 빈도를 결정합니다. 기본적으로 이 매개 변수는 **0.5** 로 설정되어 **OSD**가 초당 절반씩 통계를 보고합니다.*

이 문제를 해결하기 위해

1. *오래된 **PG**와 **OSD** 가 저장된 배치를 식별합니다. 오류 메시지에는 다음 예와 유사한 정보가 포함됩니다.*

예제

```
[ceph: root@host01 /]# ceph health detail
HEALTH_WARN 24 pgs stale; 3/300 in osds are down
...
pg 2.5 is stuck stale+active+remapped, last acting [2,0]
...
osd.10 is down since epoch 23, last address 192.168.106.220:6800/11080
osd.11 is down since epoch 13, last address 192.168.106.220:6803/11539
osd.12 is down since epoch 24, last address 192.168.106.220:6806/11861
```

2. **down** 으로 표시된 **OSD**를 사용하여 문제를 해결합니다. 자세한 내용은 **OSD 다운** 을 참조하십시오.

추가 리소스

- **Red Hat Ceph Storage 5** 관리 가이드 의 **Monitoring Placement Group Sets** 섹션

8.2.4. 일관성 없는 배치 그룹

일부 배치 그룹은 활성 + 정리 + 일관되지 않음으로 표시되고 **ceph** 상태 세부 정보는 다음과 유사한 오류 메시지를 반환합니다.

```
HEALTH_ERR 1 pgs inconsistent; 2 scrub errors
pg 0.6 is active+clean+inconsistent, acting [0,1,2]
2 scrub errors
```

<그 Means>

Ceph가 배치 그룹에서 하나 이상의 개체 복제본에서 불일치를 감지하면 배치 그룹이 일관되지 않은 것으로 표시됩니다. 가장 일반적인 불일치는 다음과 같습니다.

- 오브젝트의 크기가 잘못된 것입니다.
- 복구가 완료된 후 하나의 복제본에서 개체가 누락됩니다.

대부분의 경우 배치 그룹 내에서 불일치를 제거하는 동안 오류가 발생합니다.

이 문제를 해결하기 위해

1. **Cephadm** 셸에 로그인합니다.

예제

```
[root@host01 ~]# cephadm shell
```

2. 일관되지 않은 상태에 있는 배치 그룹을 확인합니다.

```
[ceph: root@host01 /]# ceph health detail
HEALTH_ERR 1 pgs inconsistent; 2 scrub errors
pg 0.6 is active+clean+inconsistent, acting [0,1,2]
2 scrub errors
```

3. 배치 그룹이 일관되지 않은 이유를 확인합니다.

   a. 배치 그룹에서 딥 스크럽 프로세스를 시작합니다.

   구문

   ```
   ceph pg deep-scrub ID
   ```

   일관성 없는 배치 그룹의 **ID** 로 **ID**를 바꿉니다. 예를 들면 다음과 같습니다.

   ```
   [ceph: root@host01 /]# ceph pg deep-scrub 0.6
   instructing pg 0.6 on osd.0 to deep-scrub
   ```

   b. **ceph -w** 의 출력을 검색하여 해당 배치 그룹과 관련된 모든 메시지를 검색합니다.

   구문

```
ceph -w | grep ID
```

일관성 없는 배치 그룹의 **ID** 로 **ID**를 바꿉니다. 예를 들면 다음과 같습니다.

```
[ceph: root@host01 /]# ceph -w | grep 0.6
2022-05-26 01:35:36.778215 osd.106 [ERR] 0.6 deep-scrub stat mismatch, got 636/635
objects, 0/0 clones, 0/0 dirty, 0/0 omap, 0/0 hit_set_archive, 0/0 whiteouts,
1855455/1854371 bytes.
2022-05-26 01:35:36.788334 osd.106 [ERR] 0.6 deep-scrub 1 errors
```

4. 출력에 다음 항목과 유사한 오류 메시지가 포함된 경우 일관성 없는 배치 그룹을 복구할 수 있습니다. 자세한 내용은 일관성 없는 배치 그룹 복구를 참조하십시오.

구문

```
PG.ID shard OSD: soid OBJECT missing attr , missing attr _ATTRIBUTE_TYPE
PG.ID shard OSD: soid OBJECT digest 0 != known digest DIGEST, size 0 != known size
SIZE
PG.ID shard OSD: soid OBJECT size 0 != known size SIZE
PG.ID deep-scrub stat mismatch, got MISMATCH
PG.ID shard OSD: soid OBJECT candidate had a read error, digest 0 != known digest
DIGEST
```

5. 출력에 다음 항목과 유사한 오류 메시지가 포함된 경우 데이터를 유실할 수 있으므로 일관성 없는 배치 그룹을 복구하는 것은 안전하지 않습니다. 이 상황에서 지원 티켓을 엽니 다. 자세한 내용은 **Red Hat** 지원 문의를 참조하십시오.

```
PG.ID shard OSD: soid OBJECT digest DIGEST != known digest DIGEST
PG.ID shard OSD: soid OBJECT omap_digest DIGEST != known omap_digest DIGEST
```

추가 리소스

- **Red Hat Ceph Storage** 문제 해결 가이드 의 목록 배치 그룹 불일치 를 참조하십시오.

- Red Hat Ceph Storage Architecture Guide의 Ceph 데이터 무결성 섹션을 참조하십시오.
- Red Hat Ceph Storage 구성 가이드 의 OSD 수정 섹션을 참조하십시오.

8.2.5. 불명확한 배치 그룹

ceph health 명령은 다음과 유사한 오류 메시지를 반환합니다.

```
HEALTH_WARN 197 pgs stuck unclean
```

<그 Means>

Ceph 구성 파일의 mon_pg_stuck_threshold 매개변수에 지정된 초 동안 active+clean 상태를 달성하지 않은 경우 Ceph는 배치 그룹을 unclean 으로 표시합니다. mon_pg_stuck_threshold 의 기본값은 300 초입니다.

배치 그룹이 불명확한 경우 osd_pool_default_size 매개변수에 지정된 횟수를 복제하지 않는 오브젝트를 포함합니다. osd_pool_default_size 의 기본값은 3 이며, 이는 Ceph가 세 개의 복제본을 생성함을 의미합니다.

일반적으로 불명확 한 배치 그룹은 일부 OSD가 다운 될 수 있음을 나타냅니다.

이 문제를 해결하기 위해

1. 다운 된 OSD를 확인합니다.

```
[ceph: root@host01 /]# ceph osd tree
```

2. OSD를 사용하여 문제를 해결합니다. 자세한 내용은 OSD 다운 을 참조하십시오.

추가 리소스

- 오래된 비활성 또는 불명확 상태에 있는 배치 그룹 나열.

8.2.6. 비활성 배치 그룹

ceph health 명령은 다음과 유사한 오류 메시지를 반환합니다.

```
HEALTH_WARN 197 pgs stuck inactive
```

<그 Means>

Ceph 구성 파일의 mon_pg_stuck_threshold 매개변수에 지정된 초 동안 활성 상태가 아닌 경우 Ceph는 배치 그룹을 비활성 으로 표시합니다. mon_pg_stuck_threshold 의 기본값은 300 초입니다.

일반적으로 비활성 배치 그룹은 일부 OSD가 다운 될 수 있음을 나타냅니다.

이 문제를 해결하기 위해

1. *다운 된 OSD를 확인합니다.*

```
# ceph osd tree
```

2. *OSD를 사용하여 문제를 해결합니다.*

추가 리소스

- *오래된 비활성 또는 불명확 상태에 있는 배치 그룹 나열*
- *자세한 내용은 OSD 다운 을 참조하십시오.*

8.2.7. 배치 그룹이 다운됨

ceph 상태 세부 정보는 일부 배치 그룹이 다운 되었음을 보고합니다.

```
HEALTH_ERR 7 pgs degraded; 12 pgs down; 12 pgs peering; 1 pgs recovering; 6 pgs stuck
unclean; 114/3300 degraded (3.455%); 1/3 in osds are down
...
pg 0.5 is down+peering
pg 1.4 is down+peering
...
osd.1 is down since epoch 69, last address 192.168.106.220:6801/8651
```

<그 Means>

*경우에 따라 피어링 프로세스를 차단할 수 있으므로 배치 그룹이 활성 상태이고 사용 불가능합니다. 일반적으로 **OSD** 실패로 인해 피어링 오류가 발생합니다.*

이 문제를 해결하기 위해

피어링 프로세스를 차단하는 작업을 확인합니다.

구문

```
ceph pg ID query
```

***ID** 를 다운 된 배치 그룹의 **ID**로 바꿉니다.*

예제

```
[ceph: root@host01 /]#  ceph pg 0.5 query

{ "state": "down+peering",
  ...
  "recovery_state": [
      { "name": "Started\/Primary\/Peering\/GetInfo",
        "enter_time": "2021-08-06 14:40:16.169679",
        "requested_info_from": []},
      { "name": "Started\/Primary\/Peering",
        "enter_time": "2021-08-06 14:40:16.169659",
        "probing_osds": [
              0,
              1],
        "blocked": "peering is blocked due to down osds",
        "down_osds_we_would_probe": [
              1],
        "peering_blocked_by": [
              { "osd": 1,
                "current_lost_at": 0,
                "comment": "starting or marking this osd lost may let us proceed"}]},
      { "name": "Started",
        "enter_time": "2021-08-06 14:40:16.169513"}
  ]
}
```

recovery_state 섹션에는 피어 프로세스가 차단되는 이유에 대한 정보가 포함되어 있습니다.

- osds 오류 메시지로 인해 출력에 피어링이 차단된 경우 OSD 다운 을 참조하십시오.
- 다른 오류 메시지가 표시되면 지원 티켓을 엽니다. 자세한 내용은 Red Hat 지원 서비스 문의를 참조하십시오.

추가 리소스

- Red Hat Ceph Storage 관리 가이드의 CephOSD 피어링 섹션 .

8.2.8. unfound 오브젝트

ceph health 명령은 unfound 키워드를 포함하는 다음 오류 메시지를 반환합니다.

```
HEALTH_WARN 1 pgs degraded; 78/3778 unfound (2.065%)
```

<그 Means>

이러한 오브젝트 또는 최신 복사본이 존재하지만 해당 복사본을 찾을 수 없는 경우 Ceph는 개체를 unfound 로 표시합니다. 결과적으로 Ceph는 이러한 오브젝트를 복구하고 복구 프로세스를 진행할 수 없습니다.

예제 종료

배치 그룹은 osd.1 및 osd.2 에 데이터를 저장합니다.

1. OSD.1 이 다운 됩니다.
2. OSD.2 에서는 일부 쓰기 작업을 처리합니다.
3. OSD.1 이 가동 됩니다.

4. *osd.1 과 osd.2 간의 피어링 프로세스가 시작되고 osd.1 에서 누락된 개체가 복구를 위해 대기 중입니다.*

5. *Ceph가 새 오브젝트를 복사하기 전에 osd.2 가 다운 됩니다.*

그 결과 osd.1 은 이러한 오브젝트가 존재하지만 오브젝트 복사본이 있는 OSD가 없습니다.

이 시나리오에서는 Ceph가 오류가 발생한 노드에 다시 액세스할 수 있을 때까지 대기 중이며 unfound 개체는 복구 프로세스를 차단합니다.

이 문제를 해결하기 위해

1. *Cephadm 쉘에 로그인합니다.*

   예제

```
[root@host01 ~]# cephadm shell
```

2. *어떤 배치 그룹에 unfound 오브젝트가 포함되어 있는지 확인합니다.*

```
[ceph: root@host01 /]# ceph health detail
HEALTH_WARN 1 pgs recovering; 1 pgs stuck unclean; recovery 5/937611 objects
degraded (0.001%); 1/312537 unfound (0.000%)
pg 3.8a5 is stuck unclean for 803946.712780, current state active+recovering, last acting
[320,248,0]
pg 3.8a5 is active+recovering, acting [320,248,0], 1 unfound
recovery 5/937611 objects degraded (0.001%); **1/312537 unfound (0.000%)**
```

3. *배치 그룹에 대한 추가 정보를 나열합니다.*

   구문

```
ceph pg ID query
```

ID 를 unfound 오브젝트가 포함된 배치 그룹의 ID로 바꿉니다.

예제

```
[ceph: root@host01 /]# ceph pg 3.8a5 query
{ "state": "active+recovering",
  "epoch": 10741,
  "up": [
        320,
        248,
        0],
  "acting": [
        320,
        248,
        0],
<snip>
  "recovery_state": [
        { "name": "Started\/Primary\/Active",
          "enter_time": "2021-08-28 19:30:12.058136",
          "might_have_unfound": [
                { "osd": "0",
                  "status": "already probed"},
                { "osd": "248",
                  "status": "already probed"},
                { "osd": "301",
                  "status": "already probed"},
                { "osd": "362",
                  "status": "already probed"},
                { "osd": "395",
                  "status": "already probed"},
                { "osd": "429",
                  "status": "osd is down"}],
          "recovery_progress": { "backfill_targets": [],
              "waiting_on_backfill": [],
              "last_backfill_started": "0\/\/0\/\/-1",
              "backfill_info": { "begin": "0\/\/0\/\/-1",
                  "end": "0\/\/0\/\/-1",
                  "objects": []},
              "peer_backfill_info": [],
              "backfills_in_flight": [],
              "recovering": [],
              "pg_backend": { "pull_from_peer": [],
                  "pushing": []}},
          "scrub": { "scrubber.epoch_start": "0",
              "scrubber.active": 0,
```

```
            "scrubber.block_writes": 0,
            "scrubber.finalizing": 0,
            "scrubber.waiting_on": 0,
            "scrubber.waiting_on_whom": []}},
        { "name": "Started",
          "enter_time": "2021-08-28 19:30:11.044020"}],
```

might_have_unfound 섹션에는 **Ceph**가 **unfound** 오브젝트를 검색하려고 하는 **OSD**가 포함되어 있습니다.

- 이미 프로브된 상태는 **Ceph**가 해당 **OSD**에서 **unfound** 오브젝트를 찾을 수 없음을 나타냅니다.
- **osd**가 **down** 상태이면 **Ceph**가 해당 **OSD**에 연결할 수 없음을 나타냅니다.

4. **down** 으로 표시된 **OSD**의 문제를 해결합니다. 자세한 내용은 **OSD 다운** 을 참조하십시오.

5. **OSD**가 다운 되는 문제를 해결할 수 없는 경우 지원 티켓을 엽니다. 자세한 내용은 **Red Hat 지원 문의**를 참조하십시오.

8.3. 오래된,비활성 또는 불명확 상태에 있는 배치 그룹 나열

실패 후 배치 그룹은 성능 저하 또는 피어링 과 같은 상태를 입력합니다. 이 상태는 오류 복구 프로세스를 통해 정상적인 진행을 나타냅니다.

그러나 배치 그룹이 예상보다 오랜 시간 동안 이러한 상태 중 하나에 남아 있으면 더 큰 문제가 있음을 나타낼 수 있습니다. 배치 그룹이 최적이 아닌 상태로 고정되는 경우 모니터 보고서입니다.

Ceph 구성 파일의 **mon_pg_stuck_threshold** 옵션은 배치 그룹이 비활성, 비정적 또는 오래된 것으로 간주되는 시간 이후의 시간(초)을 결정합니다.

다음 표에는 이러한 상태가 짧은 설명과 함께 나열되어 있습니다.

상태	무엇을 의미합니까?	대부분의 일반적인 원인	참조
비활성	PG는 읽기/쓰기 요청을 처리할 수 없습니다.	• 피어링 문제	비활성 배치 그룹
unclean	PG에는 원하는 횟수만큼 복제되지 않은 오브젝트가 포함되어 있습니다. PG가 복구되지 않는 문제가 있습니다.	• **unfound** 오브젝트 • OSD가 **다운됨** • 잘못된 설정	불명확한 배치 그룹
stale	PG의 상태는 **ceph-osd** 데몬에 의해 업데이트되지 않았습니다.	• OSD가 **다운됨**	오래된 배치 그룹

사전 요구 사항

- *실행 중인 **Red Hat Ceph Storage** 클러스터.*
- *노드에 대한 루트 수준 액세스입니다.*

절차

1. ***Cephadm** 쉘에 로그인합니다.*

   예제

   ```
   [root@host01 ~]# cephadm shell
   ```

2. *정지된 **PG**를 나열합니다.*

   예제

```
[ceph: root@host01 /]# ceph pg dump_stuck inactive
[ceph: root@host01 /]# ceph pg dump_stuck unclean
[ceph: root@host01 /]# ceph pg dump_stuck stale
```

추가 리소스

- **Red Hat Ceph Storage** 관리 가이드의 배치 그룹 상태 섹션을 참조하십시오.

8.4. 배치 그룹 불일치 나열

rados 유틸리티를 사용하여 다양한 오브젝트 복제본의 불일치를 나열합니다. **--format=json-pretty** 옵션을 사용하여 자세한 출력을 나열합니다.

이 섹션에서는 다음 목록을 설명합니다.

- 풀에 일관되지 않는 배치 그룹
- 배치 그룹의 일관성 없는 오브젝트
- 배치 그룹에 일관성 없는 스냅샷 세트

사전 요구 사항

- 실행 중인 **Red Hat Ceph Storage** 클러스터가 정상 상태입니다.
- 노드에 대한 루트 수준 액세스입니다.

절차

- 풀의 일치하지 않는 배치 그룹을 모두 나열합니다.

구문

```
rados list-inconsistent-pg POOL --format=json-pretty
```

예제

```
[ceph: root@host01 /]# rados list-inconsistent-pg data --format=json-pretty
[0.6]
```

- ***ID**가 있는 배치 그룹의 일관성 없는 오브젝트를 나열합니다.*

구문

```
rados list-inconsistent-obj PLACEMENT_GROUP_ID
```

예제

```
[ceph: root@host01 /]# rados list-inconsistent-obj 0.6
{
    "epoch": 14,
    "inconsistents": [
        {
            "object": {
                "name": "image1",
                "nspace": "",
                "locator": "",
                "snap": "head",
                "version": 1
            },
            "errors": [
                "data_digest_mismatch",
                "size_mismatch"
```

```
            ],
            "union_shard_errors": [
                "data_digest_mismatch_oi",
                "size_mismatch_oi"
            ],
            "selected_object_info": "0:602f83fe:::foo:head(16'1 client.4110.0:1
dirty|data_digest|omap_digest s 968 uv 1 dd e978e67f od ffffffff alloc_hint [0 0 0])",
            "shards": [
                {
                    "osd": 0,
                    "errors": [],
                    "size": 968,
                    "omap_digest": "0xffffffff",
                    "data_digest": "0xe978e67f"
                },
                {
                    "osd": 1,
                    "errors": [],
                    "size": 968,
                    "omap_digest": "0xffffffff",
                    "data_digest": "0xe978e67f"
                },
                {
                    "osd": 2,
                    "errors": [
                        "data_digest_mismatch_oi",
                        "size_mismatch_oi"
                    ],
                    "size": 0,
                    "omap_digest": "0xffffffff",
                    "data_digest": "0xffffffff"
                }
            ]
        }
    ]
}
```


다음 필드는 불일치의 원인을 결정하는 데 중요합니다.

- ***name:** 일관되지 않은 복제본이 있는 오브젝트의 이름입니다.*
- ***nSpace:** 풀의 논리적인 분리 네임스페이스입니다. 기본적으로 비어 있습니다.*
- ***locator:** 배치에 오브젝트 이름 대신 사용되는 키입니다.*

- *snap: 오브젝트의 스냅샷 ID입니다. 오브젝트의 쓰기 가능한 유일한 버전을 head 라고 합니다. 오브젝트가 복제본인 경우 이 필드에는 순차적 ID가 포함됩니다.*

- *버전: 일관되지 않은 복제본이 있는 오브젝트의 버전 ID입니다. 개체에 대한 각 쓰기 작업은 해당 개체를 늘립니다.*

- *Error: 어떤 shard 또는 shard가 잘못되었는지 확인하지 않고 shard 간 불일치를 나타내는 오류 목록입니다. 오류를 추가로 조사하려면 shard 배열을 참조하십시오.*

  - *data_digest_mismatch: 하나의 OSD에서 읽은 복제본의 다이제스트는 다른 OSD와 다릅니다.*

  - *size_mismatch: 복제본 또는 head 오브젝트의 크기가 예상과 일치하지 않습니다.*

  - *read_error:이 오류는 불일치로 인해 디스크 오류로 인해 발생할 가능성이 있음을 나타냅니다.*

- *union_shard_error: shard와 관련된 모든 오류의 조합입니다. 이러한 오류는 잘못된 shard에 연결됩니다. oi 로 끝나는 오류는 결함이 있는 오브젝트의 정보를 선택한 오브젝트와 비교해야 함을 나타냅니다. 오류를 추가로 조사하려면 shard 배열을 참조하십시오.*

  위 예제에서 osd.2 에 저장된 오브젝트 복제본에는 osd.0 및 osd.1 에 저장된 복제본과 다른 다이제스트가 있습니다. 특히, 복제본의 다이제스트는 osd.2 의 shard 읽기에서 계산되는 0xffffffff ff가 아니라 0xe978e67f 입니다. 또한 osd.2 에서 읽은 복제본 크기는 0이며 osd.0 및 osd.1 은 968입니다.

- *일관되지 않은 스냅샷 세트를 나열합니다.*

구문

```
rados list-inconsistent-snapset PLACEMENT_GROUP_ID
```

예제

```
[ceph: root@host01 /]# rados list-inconsistent-snapset 0.23 --format=json-pretty
{
    "epoch": 64,
    "inconsistents": [
        {
            "name": "obj5",
            "nspace": "",
            "locator": "",
            "snap": "0x00000001",
            "headless": true
        },
        {
            "name": "obj5",
            "nspace": "",
            "locator": "",
            "snap": "0x00000002",
            "headless": true
        },
        {
            "name": "obj5",
            "nspace": "",
            "locator": "",
            "snap": "head",
            "ss_attr_missing": true,
            "extra_clones": true,
            "extra clones": [
                2,
                1
            ]
        }
    ]
```

이 명령은 다음 오류를 반환합니다.

- ***ss_attr_missing:** 하나 이상의 속성이 누락되었습니다. 속성은 키-값 쌍 목록으로 설정된 스냅샷에 대한 정보입니다.*
- ***ss_attr_corrupted:** 하나 이상의 속성이 디코딩되지 않습니다.*
- ***clone_missing:** 클론이 없습니다.*

- **snapset_mismatch:** 스냅샷 세트는 그 자체로 일관되지 않습니다.
- **head_mismatch:** 스냅샷 세트는 헤드 가 존재하거나 그렇지 않음을 나타내지만, **scrub** 결과는 그렇지 않으면 보고됩니다.
- 제목: 스냅샷 세트의 헤드 가 누락되어 있습니다.
- **size_mismatch:** 복제본 또는 **head** 오브젝트의 크기가 예상과 일치하지 않습니다.

추가 리소스

- **Red Hat Ceph Storage** 문제 해결 가이드 의 일관성 없는 배치 그룹 섹션.
- **Red Hat Ceph Storage** 문제 해결 가이드에서 일관성 없는 배치 그룹 복구 섹션 .

8.5. 일관성 없는 배치 그룹 복구

딥 스크럽하는 동안 오류가 있기 때문에 일부 배치 그룹에는 불일치가 포함될 수 있습니다. **Ceph**에서 이러한 배치 그룹을 일관되지 않음으로 보고합니다.

```
HEALTH_ERR 1 pgs inconsistent; 2 scrub errors
pg 0.6 is active+clean+inconsistent, acting [0,1,2]
2 scrub errors
```

주의

특정 불일치만 복구할 수 있습니다.

Ceph 로그에 다음 오류가 포함된 경우 배치 그룹을 복구하지 마십시오.

```
_PG_._ID_ shard _OSD_: soid _OBJECT_ digest _DIGEST_ != known digest _DIGEST_
_PG_._ID_ shard _OSD_: soid _OBJECT_ omap_digest _DIGEST_ != known omap_digest
_DIGEST_
```

대신 지원 티켓을 엽니 다. 자세한 내용은 **Red Hat** 지원 문의를 참조하십시오.

사전 요구 사항

- **Ceph Monitor** 노드에 대한 루트 수준 액세스입니다.

절차

- 일관성 없는 배치 그룹을 복구합니다.

  구문

  ```
  ceph pg repair ID
  ```

  ID 를 일관성 없는 배치 그룹의 **ID**로 바꿉니다.

추가 리소스

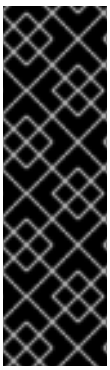
- **Red Hat Ceph Storage Troubleshooting Guide** 의 **Inconsistent placement groups** 섹션을 참조하십시오.
- 목록 배치 그룹 불일치 **Red Hat Ceph Storage** 문제 해결 가이드를 참조하십시오.

8.6. 배치 그룹 증가

PG(배치 그룹**)** 수가 충분하지 않으면 **Ceph** 클러스터 및 데이터 배포 성능에 영향을 미칩니다. 이는 가까운 전체 **osds** 오류 메시지의 주요 원인 중 하나입니다.

*권장되는 비율은 **OSD**당 **100~300**개의 **PG**입니다. 이 비율은 클러스터에 **OSD**를 추가할 때 감소할 수 있습니다.*

***pg_num** 및 **pgp_num** 매개변수는 **PG** 개수를 결정합니다. 이러한 매개변수는 각 풀마다 구성되므로 **PG** 수가 낮은 각 풀은 개별적으로 조정해야 합니다.*



중요

***PG** 수를 늘리는 것은 **Ceph** 클러스터에서 수행할 수 있는 가장 집중적인 프로세스입니다. 이 프로세스는 속도가 느리고 체계적인 방식으로 수행되지 않으면 심각한 성능 영향을 미칠 수 있습니다. **pgp_num** 을 늘리면 프로세스를 중지하거나 되돌릴 수 없으며 완료해야 합니다. 비즈니스 중요한 처리 시간 할당 외부에서 **PG** 수를 늘리고 모든 클라이언트에 잠재적인 성능에 미치는 영향에 대해 경고하는 것이 좋습니다. 클러스터가 **HEALTH_ERR** 상태인 경우 **PG** 수를 변경하지 마십시오.*

사전 요구 사항

- *실행 중인 **Red Hat Ceph Storage** 클러스터가 정상 상태입니다.*
- *노드에 대한 루트 수준 액세스입니다.*

절차

1. *개별 **OSD** 및 **OSD** 호스트에서 데이터 재배포 및 복구의 영향을 줄입니다.*

   a. ***osd max backfills,osd_recovery_max_active, osd_recovery_op_priority** 매개변수의 값을 낮춥니다.*

   ```
   [ceph: root@host01 /]# ceph tell osd.* injectargs '--osd_max_backfills 1 --
   osd_recovery_max_active 1 --osd_recovery_op_priority 1'
   ```

   b. *단순 및 딥 스크럽을 사용하지 않도록 설정합니다.*

   ```
   [ceph: root@host01 /]# ceph osd set noscrub
   [ceph: root@host01 /]# ceph osd set nodeep-scrub
   ```

2. *풀 계산기당 **Ceph PG(배치 그룹)** 를 사용하여 **pg_num** 및 **pgp_num** 매개변수의 최적의 값을 계산합니다.*

3. *원하는 값에 도달할 때까지 **pg_num** 값을 작은 증분으로 늘립니다.*

   a. *시작 증가 값을 결정합니다. **2**의 전원인 매우 낮은 값을 사용하고 클러스터에 미치는 영향을 확인할 때 늘립니다. 최적 값은 풀 크기, **OSD** 수 및 클라이언트 **I/O** 로드에 따라 달라집니다.*

   b. ***pg_num** 값을 늘립니다.*

   구문

   ```
   ceph osd pool set POOL pg_num VALUE
   ```

   풀 이름과 새 값을 지정합니다. 예를 들면 다음과 같습니다.

   예제

   ```
   [ceph: root@host01 /]# ceph osd pool set data pg_num 4
   ```

   c. *클러스터 상태를 모니터링합니다.*

   예제

   ```
   [ceph: root@host01 /]# ceph -s
   ```

PGs 상태는 creating 에서 active+clean 으로 변경됩니다. 모든 PG가 active+clean 상태가 될 때까지 기다립니다.

4. *원하는 값에 도달할 때까지 pgp_num 값을 작은 증분으로 늘립니다.*

   a. *시작 증가 값을 결정합니다. 2의 전원인 매우 낮은 값을 사용하고 클러스터에 미치는 영향을 확인할 때 늘립니다. 최적 값은 풀 크기, OSD 수 및 클라이언트 I/O 로드에 따라 달라집니다.*

   b. *pgp_num 값을 늘립니다.*

   구문

   ```
   ceph osd pool set POOL pgp_num VALUE
   ```

   풀 이름과 새 값을 지정합니다. 예를 들면 다음과 같습니다.

   ```
   [ceph: root@host01 /]# ceph osd pool set data pgp_num 4
   ```

   c. *클러스터 상태를 모니터링합니다.*

   ```
   [ceph: root@host01 /]# ceph -s
   ```

   PG 상태는 피어링,wait_backfill,백필링,복구 등을 통해 변경됩니다. 모든 PG가 active+clean 상태가 될 때까지 기다립니다.

5. *insufficient PG 수를 사용하여 모든 풀에 대해 이전 단계를 반복합니다.*

6. *osd max backfills,osd_recovery_max_active, osd_recovery_op_priority 를 기본값으로 설정합니다.*

```
[ceph: root@host01 /]# ceph tell osd.* injectargs '--osd_max_backfills 1 --
osd_recovery_max_active 3 --osd_recovery_op_priority 3'
```

7. 단순 및 딥 스크럽을 활성화합니다.

```
[ceph: root@host01 /]# ceph osd unset noscrub
[ceph: root@host01 /]# ceph osd unset nodeep-scrub
```

추가 리소스

- Nearfull OSD를 참조하십시오.
- Red Hat Ceph Storage 관리 가이드의 Monitoring Placement Group Sets 섹션을 참조하십시오.

8.7. 추가 리소스

- 자세한 내용은 3장. 네트워킹 문제 해결 를 참조하십시오.
- Ceph 모니터와 관련된 가장 일반적인 오류 문제 해결에 대한 자세한 내용은 4장. Ceph 모니터 문제 해결 를 참조하십시오.
- Ceph OSD와 관련된 가장 일반적인 오류 문제 해결에 대한 자세한 내용은 5장. Ceph OSD 문제 해결 를 참조하십시오.
- PG 자동 스케일러에 대한 자세한 내용은 Red Hat Ceph Storage Strategies Guide의 자동 확장 배치 그룹 섹션을 참조하십시오.

9장. CEPH 오브젝트 문제 해결

스토리지 관리자는 **ceph-objectstore-tool** 유틸리티를 사용하여 고급 또는 하위 오브젝트 작업을 수행할 수 있습니다. **ceph-objectstore-tool** 유틸리티는 특정 **OSD** 또는 배치 그룹 내의 오브젝트와 관련된 문제를 해결하는 데 도움이 될 수 있습니다.

중요

개체를 조작하면 복구할 수 없는 데이터 손실이 발생할 수 있습니다. **ceph-objectstore-tool** 유틸리티를 사용하기 전에 **Red Hat** 지원에 문의하십시오.

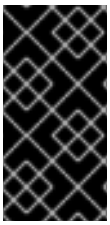
9.1. 사전 요구 사항

- 네트워크 관련 문제가 없는지 확인합니다.

9.2. 높은 수준의 오브젝트 작업 문제 해결

스토리지 관리자는 **ceph-objectstore-tool** 유틸리티를 사용하여 고급 오브젝트 작업을 수행할 수 있습니다. **ceph-objectstore-tool** 유틸리티는 다음과 같은 고급 오브젝트 작업을 지원합니다.

- 오브젝트 나열
- 손실된 오브젝트 나열
- 손실된 오브젝트 수정



중요

개체를 조작하면 복구할 수 없는 데이터 손실이 발생할 수 있습니다. **ceph-objectstore-tool** 유틸리티를 사용하기 전에 **Red Hat** 지원에 문의하십시오.

9.2.1. 사전 요구 사항

- **Ceph OSD** 노드에 대한 루트 수준 액세스입니다.

9.2.2. 오브젝트 나열

OSD에는 배치 그룹이 **0**개, 배치 그룹(**PG**) 내에 **0**개에서 많은 오브젝트를 포함할 수 있습니다. **ceph-objectstore-tool** 유틸리티를 사용하면 **OSD**에 저장된 오브젝트를 나열할 수 있습니다.

사전 요구 사항

- **Ceph OSD** 노드에 대한 루트 수준 액세스입니다.
- **ceph-osd** 데몬 중지.

절차

1. 적절한 **OSD**가 다운되었는지 확인합니다.

구문

```
systemctl status ceph-osd@OSD_NUMBER
```

예제

```
[root@host01 ~]# systemctl status ceph-osd@1
```

2. **Cephadm** 쉘에 로그인합니다.

예제

```
[root@host01 ~]# cephadm shell
```

3. *배치 그룹에 관계없이 **OSD** 내의 모든 오브젝트를 식별합니다.*

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD --op list
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 --op list
```

4. *배치 그룹 내의 모든 오브젝트를 식별합니다.*

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD --pgid PG_ID --op list
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 --pgid 0.1c
--op list
```

5. *오브젝트가 속한 **PG**를 식별합니다.*

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD --op list OBJECT_ID
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 --op list default.region
```

9.2.3. 손실된 오브젝트 수정

***ceph-objectstore-tool** 유틸리티를 사용하여 **Ceph OSD**에 저장된 손실 및 **unfound** 오브젝트를 나열하고 수정할 수 있습니다. 이 절차는 레거시 오브젝트에만 적용됩니다.*

사전 요구 사항

- ***Ceph OSD** 노드에 대한 루트 수준 액세스입니다.*
- ***ceph-osd** 데몬 중지.*

절차

1. *적절한 **OSD**가 다운되었는지 확인합니다.*

구문

```
systemctl status ceph-osd@OSD_NUMBER
```

예제

```
[root@host01 ~]# systemctl status ceph-osd@1
```

2. ***Cephadm** 셸에 로그인합니다.*

예제

```
[root@host01 ~]# cephadm shell
```

3. *손실된 기존 오브젝트를 모두 나열하려면 다음을 수행합니다.*

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD --op fix-lost --dry-run
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 --op fix-lost
--dry-run
```

4. ***ceph-objectstore-tool** 유틸리티를 사용하여 손실되거나 **unfound** 오브젝트를 수정합니다. 적절한 **situation**를 선택합니다.*

   a. *손실된 모든 오브젝트를 수정하려면 다음을 수행합니다.*

   구문

   ```
   ceph-objectstore-tool --data-path PATH_TO_OSD --op fix-lost
   ```

   예제

   ```
   [ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 --op
   fix-lost
   ```

   b. *배치 그룹 내의 손실된 오브젝트를 모두 수정하려면 다음을 수행합니다.*

   구문

   ```
   ceph-objectstore-tool --data-path PATH_TO_OSD --pgid PG_ID --op fix-lost
   ```

   예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 --pgid
0.1c --op fix-lost
```

c. **ID로 손실된 오브젝트를 수정하려면 다음을 수행합니다.**

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD --op fix-lost OBJECT_ID
```

예제

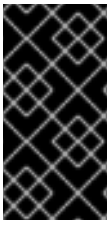
```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 --op
fix-lost default.region
```

9.3. 낮은 수준의 오브젝트 작업 문제 해결

스토리지 관리자는 **ceph-objectstore-tool** 유틸리티를 사용하여 낮은 수준의 오브젝트 작업을 수행할 수 있습니다. **ceph-objectstore-tool** 유틸리티는 다음과 같은 하위 수준 오브젝트 작업을 지원합니다.

- 오브젝트의 콘텐츠 조작
- 오브젝트 제거
- 오브젝트 맵(**OMAP**) 나열

- *OMAP 헤더 조작*
- *OMAP 키 조작*
- *오브젝트의 특성 나열*
- *오브젝트의 특성 키 조작*



중요

개체를 조작하면 복구할 수 없는 데이터 손실이 발생할 수 있습니다. ceph-objectstore-tool 유틸리티를 사용하기 전에 Red Hat 지원에 문의하십시오.

9.3.1. 사전 요구 사항

- *Ceph OSD 노드에 대한 루트 수준 액세스입니다.*

9.3.2. 개체의 콘텐츠 조작

ceph-objectstore-tool 유틸리티를 사용하면 오브젝트에서 바이트를 가져오거나 설정할 수 있습니다.

중요

개체에 바이트를 설정하면 복구할 수 없는 데이터 손실이 발생할 수 있습니다. 데이터 손실을 방지하려면 오브젝트의 백업 사본을 만듭니다.

사전 요구 사항

- *Ceph OSD 노드에 대한 루트 수준 액세스입니다.*
- *ceph-osd 데몬 중지.*

절차

1. *적절한 **OSD**가 다운되었는지 확인합니다.*

구문

```
systemctl status ceph-osd@OSD_ID
```

예제

```
[root@host01 ~]# systemctl status ceph-osd@1
```

2. ***OSD** 또는 배치 그룹(**PG**)의 오브젝트를 나열하여 오브젝트를 찾습니다.*

3. *오브젝트에서 바이트를 설정하기 전에 오브젝트의 백업 및 작업 사본을 만듭니다.*

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD --pgid PG_ID \
OBJECT \
get-bytes > OBJECT_FILE_NAME
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 --pgid 0.1c
\
'{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}'  \
```

```
get-bytes > zone_info.default.backup

[ceph: root@host01 /]#  ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 --pgid 0.1c
\
'{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}'  \
get-bytes > zone_info.default.working-copy
```

4. *작업 복사본 오브젝트 파일을 편집하고 그에 따라 오브젝트 콘텐츠를 수정합니다.*

5. *오브젝트의 바이트를 설정합니다.*

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD --pgid PG_ID \
OBJECT \
set-bytes < OBJECT_FILE_NAME
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 --pgid 0.1c
\
'{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}' \
set-bytes < zone_info.default.working-copy
```

9.3.3. 오브젝트 제거

***ceph-objectstore-tool** 유틸리티를 사용하여 오브젝트를 제거합니다. 오브젝트를 제거하면 해당 콘텐츠 및 참조가 배치 그룹(**PG**)에서 제거됩니다.*

중요

제거되면 개체를 다시 생성할 수 없습니다.

사전 요구 사항

- ***Ceph OSD** 노드에 대한 루트 수준 액세스입니다.*
- ***ceph-osd** 데몬 중지.*

절차

1. ***Cephadm** 쉘에 로그인합니다.*

   예제

   ```
   [root@host01 ~]# cephadm shell
   ```

2. *오브젝트를 제거합니다.*

   구문

   ```
   ceph-objectstore-tool --data-path PATH_TO_OSD --pgid PG_ID \
   OBJECT \
   remove
   ```

   예제

   ```
   [ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 --pgid 0.1c
   ```

```
\
'{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}' \
remove
```

9.3.4. 오브젝트 맵 나열

ceph-objectstore-tool 유틸리티를 사용하여 오브젝트 맵(OMAP)의 콘텐츠를 나열합니다. 출력에서 키 목록을 제공합니다.

사전 요구 사항

- **Ceph OSD** 노드에 대한 루트 수준 액세스입니다.
- **ceph-osd** 데몬 중지.

절차

1. 적절한 **OSD**가 다운되었는지 확인합니다.

```
systemctl status ceph-osd@OSD_ID
```

예제

```
[root@host01 ~]# systemctl status ceph-osd@1
```

2. **Cephadm** 쉘에 로그인합니다.

예제

```
[root@host01 ~]# cephadm shell
```

-

3. *오브젝트 맵을 나열합니다.*

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD --pgid PG_ID \
OBJECT \
list-omap
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 --pgid 0.1c
\
'{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}' \
list-omap
```

9.3.5. 오브젝트 맵 헤더 조작

***ceph-objectstore-tool** 유틸리티는 오브젝트 키와 연결된 값으로 오브젝트 맵(OMAP) 헤더를 출력합니다.*

사전 요구 사항

- ***Ceph OSD** 노드에 대한 루트 수준 액세스입니다.*
- ***ceph-osd** 데몬 중지.*

절차

1. *적절한 **OSD**가 다운되었는지 확인합니다.*

구문

```
systemctl status ceph-osd@OSD_ID
```

예제

```
[root@host01 ~]# systemctl status ceph-osd@1
```

2. ***Cephadm** 쉘에 로그인합니다.*

예제

```
[root@host01 ~]# cephadm shell
```

3. *오브젝트 맵 헤더를 가져옵니다.*

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD \
--pgid PG_ID OBJECT \
get-omaphdr > OBJECT_MAP_FILE_NAME
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 \
--pgid 0.1c '{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}' \
get-omaphdr > zone_info.default.omaphdr.txt
```

4. *오브젝트 맵 헤더를 설정합니다.*

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD \
--pgid PG_ID OBJECT \
get-omaphdr < OBJECT_MAP_FILE_NAME
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 \
--pgid 0.1c '{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}' \
set-omaphdr < zone_info.default.omaphdr.txt
```

9.3.6. 오브젝트 맵 키 조작

***ceph-objectstore-tool** 유틸리티를 사용하여 오브젝트 맵(OMAP) 키를 변경합니다. 데이터 경로, 배치 그룹 **ID(PG ID)**, 개체 및 **OMAP**의 키를 제공해야 합니다.*

사전 요구 사항

- ***Ceph OSD** 노드에 대한 루트 수준 액세스입니다.*
- ***ceph-osd** 데몬 중지.*

절차

- *오브젝트 맵 키를 가져옵니다.*

  구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD \
--pgid PG_ID OBJECT \
get-omap KEY > OBJECT_MAP_FILE_NAME
```

  예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 \
--pgid 0.1c '{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}'  \
get-omap "" > zone_info.default.omap.txt
```

- *오브젝트 맵 키를 설정합니다.*

  구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD \
--pgid PG_ID OBJECT \
set-omap KEY < OBJECT_MAP_FILE_NAME
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 \
--pgid 0.1c '{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}'  \
set-omap "" < zone_info.default.omap.txt
```

- *오브젝트 맵 키를 제거합니다.*

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD \
--pgid PG_ID OBJECT \
rm-omap KEY
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 \
--pgid 0.1c '{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}'  \
rm-omap ""
```

9.3.7. 오브젝트의 특성 나열

***ceph-objectstore-tool** 유틸리티를 사용하여 오브젝트의 특성을 나열합니다. 출력에서는 오브젝트의 키와 값을 제공합니다.*

사전 요구 사항

- ***Ceph OSD** 노드에 대한 루트 수준 액세스입니다.*
- ***ceph-osd** 데몬 중지.*

절차

1. *적절한 **OSD**가 다운되었는지 확인합니다.*

   구문

   ```
   systemctl status ceph-osd@OSD_ID
   ```

   예제

   ```
   [root@host01 ~]# systemctl status ceph-osd@1
   ```

2. *오브젝트의 특성을 나열합니다.*

   구문

   ```
   ceph-objectstore-tool --data-path PATH_TO_OSD \
   --pgid PG_ID OBJECT \
   list-attrs
   ```

   예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 \
--pgid 0.1c '{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}' \
list-attrs
```

9.3.8. 개체 특성 키 조작

ceph-objectstore-tool 유틸리티를 사용하여 오브젝트의 특성을 변경합니다. 오브젝트의 특성을 조작하려면 데이터 경로, 배치 그룹 **ID(PG ID)**, 오브젝트의 속성의 키가 필요합니다.

사전 요구 사항

- **Ceph OSD** 노드에 대한 루트 수준 액세스입니다.
- **ceph-osd** 데몬을 중지합니다.

절차

1. 적절한 **OSD**가 다운되었는지 확인합니다.

구문

```
systemctl status ceph-osd@OSD_ID
```

예제

```
[root@host01 ~]# systemctl status ceph-osd@1
```

2. ***Cephadm** 셸에 로그인합니다.*

예제

```
[root@host01 ~]# cephadm shell
```

3. *오브젝트의 특성을 가져옵니다.*

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD \
--pgid PG_ID OBJECT \
get-attrs KEY > OBJECT_ATTRS_FILE_NAME
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 \
--pgid 0.1c '{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}' \
get-attrs "oid" > zone_info.default.attr.txt
```

4. *오브젝트의 특성을 설정합니다.*

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD \
--pgid PG_ID OBJECT \
set-attrs KEY < OBJECT_ATTRS_FILE_NAME
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 \
--pgid 0.1c '{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}' \
set-attrs "oid" < zone_info.default.attr.txt
```

5. *오브젝트의 특성을 제거합니다.*

구문

```
ceph-objectstore-tool --data-path PATH_TO_OSD \
--pgid PG_ID OBJECT  \
rm-attrs KEY
```

예제

```
[ceph: root@host01 /]# ceph-objectstore-tool --data-path /var/lib/ceph/osd/ceph-0 \
--pgid 0.1c '{"oid":"zone_info.default","key":"","snapid":-
2,"hash":235010478,"max":0,"pool":11,"namespace":""}' \
rm-attrs "oid"
```

9.4. 추가 리소스

- Red Hat Ceph Storage 지원의 경우 Red Hat 고객 포털 을 참조하십시오.

10장. 서비스에 대한 RED HAT 지원 문의

이 가이드의 정보가 문제 해결에 도움이 되지 않은 경우 이 장에서는 Red Hat 지원 서비스에 문의하는 방법을 설명합니다.

10.1. 사전 요구 사항

- Red Hat 지원 계정.

10.2. RED HAT 지원 엔지니어에 대한 정보 제공

Red Hat Ceph Storage와 관련된 문제를 해결할 수 없는 경우 Red Hat 지원 서비스에 문의하여 문제 해결 시간을 단축할 수 있도록 지원 엔지니어가 충분한 정보를 제공합니다.

사전 요구 사항

- 노드에 대한 루트 수준 액세스입니다.
- Red Hat 지원 계정.

절차

1. Red Hat 고객 포털에서 지원 티켓을 엽니다.
2. sosreport 를 티켓에 연결하는 것이 좋습니다. 자세한 내용은 What is a sosreport and how to create one in Red Hat Enterprise Linux? 솔루션을 참조하십시오.
3. Ceph 데몬이 세그먼트 오류와 함께 실패하는 경우 사람이 읽을 수 있는 코어 덤프 파일을 생성하는 것이 좋습니다. 자세한 내용은 읽을 수 있는 코어 덤프 파일 생성을 참조하십시오.

10.3. 읽을 수 있는 코어 덤프 파일 생성

Ceph 데몬이 세그먼트화 오류와 함께 예기치 않게 종료되면 해당 오류에 대한 정보를 수집하여 Red Hat 지원 엔지니어에게 제공합니다.

*이러한 정보는 초기 조사를 가속화합니다. 또한 지원 엔지니어는 코어 덤프 파일의 정보와 **Red Hat Ceph Storage** 클러스터의 정보를 비교할 수 있습니다.*

10.3.1. 사전 요구 사항

1. ***debuginfo** 패키지가 아직 설치되지 않은 경우 설치합니다.*

   a. *다음 리포지토리를 활성화하여 필요한 **debuginfo** 패키지를 설치합니다.*

   예제

   ```
   [root@host01 ~]# subscription-manager repos --enable=rhceph-5-tools-for-rhel-8-x86_64-rpms
   [root@host01 ~]# yum --disablerepo='*' --enable=rhceph-5-tools-for-rhel-8-x86_64-debug-rpms
   ```

   리포지토리가 활성화되면 지원되는 패키지 목록에서 필요한 디버그 정보 패키지를 설치할 수 있습니다.

   ```
   ceph-base-debuginfo
   ceph-common-debuginfo
   ceph-debugsource
   ceph-fuse-debuginfo
   ceph-immutable-object-cache-debuginfo
   ceph-mds-debuginfo
   ceph-mgr-debuginfo
   ceph-mon-debuginfo
   ceph-osd-debuginfo
   ceph-radosgw-debuginfo
   cephfs-mirror-debuginfo
   ```

2. ***gdb** 패키지가 설치되어 있고 그렇지 않은 경우 설치합니다.*

   예제

```
[root@host01 ~]# dnf install gdb
```

- 10.3.2절. “컨테이너화된 배포에서 읽을 수 있는 코어 덤프 파일 생성”

10.3.2. 컨테이너화된 배포에서 읽을 수 있는 코어 덤프 파일 생성

{storage-product| 5의 코어 덤프 파일을 생성할 수 있으며 여기에는 코어 덤프 파일을 캡처하는 두 가지 시나리오가 포함됩니다.

- SIGILL, SIGTRAP, SIGABRT 또는 SIGSEGV 오류로 인해 Ceph 프로세스가 예기치 않게 종료되는 경우

또는

- 예를 들어 Ceph 프로세스 등의 디버깅 문제에서는 CPU 사이클이 크게 사용되거나 응답하지 않는 등의 작업을 수동으로 수행할 수 있습니다.

사전 요구 사항

- Ceph 컨테이너를 실행하는 컨테이너 노드에 대한 루트 수준 액세스입니다.
- 적절한 디버깅 패키지를 설치합니다.
- GNU Project Debugger (gdb) 패키지 설치
- 호스트에 8GB 이상의 RAM이 있는지 확인합니다. 호스트에 데몬이 여러 개인 경우 Red Hat 은 더 많은 RAM을 권장합니다.

절차

1. SIGILL, SIGTRAP, SIGABRT 또는 SIGSEGV 오류로 인해 Ceph 프로세스가 예기치 않게 종료되는 경우:

a.

*코어 패턴을 **Ceph** 프로세스가 실패한 컨테이너가 실행 중인 노드의 **systemd-coredump** 서비스로 설정합니다.*

예제

```
[root@mon]# echo "| /usr/lib/systemd/systemd-coredump %P %u %g %s %t %c %h %e"
> /proc/sys/kernel/core_pattern
```

b.

***Ceph** 프로세스로 인한 다음 컨테이너 오류를 확인하고 **/var/lib/systemd/coredump/** 디렉터리에서 코어 덤프 파일을 검색합니다.*

예제

```
[root@mon]# ls -ltr /var/lib/systemd/coredump
total 8232
-rw-r-----. 1 root root 8427548 Jan 22 19:24 core.ceph-
osd.167.5ede29340b6c4fe4845147f847514c12.15622.1584573794000000.xz
```

2.

***Ceph** 모니터 및 **Ceph OSD** 의 코어 덤프 파일을 수동으로 캡처하려면 다음을 수행합니다.*

a.

***MONITOR_ID** 또는 **OSD_ID** 를 가져오고 컨테이너를 입력합니다.*

구문

```
podman ps
podman exec -it MONITOR_ID_OR_OSD_ID bash
```

예제

```
[root@host01 ~]# podman ps
[root@host01 ~]# podman exec -it ceph-1ca9f6a8-d036-11ec-8263-fa163ee967ad-osd-2
bash
```

b. *컨테이너 내부에 **procps-ng** 및 **gdb** 패키지를 설치합니다.*

예제

```
[root@host01 /]# dnf install procps-ng gdb
```

c. *프로세스 **ID**를 찾습니다.*

구문

```
ps -aef | grep PROCESS | grep -v run
```

***PROCESS** 를 실행 중인 프로세스의 이름으로 교체합니다(예: **ceph-mon** 또는 **ceph-osd**).*

예제

-

```
[root@host01 /]# ps -aef | grep ceph-mon | grep -v run
ceph       15390   15266  0 18:54 ?        00:00:29 /usr/bin/ceph-mon --cluster ceph --
setroot ceph --setgroup ceph -d -i 5
ceph       18110   17985  1 19:40 ?        00:00:08 /usr/bin/ceph-mon --cluster ceph --
setroot ceph --setgroup ceph -d -i 2
```

d.

코어 덤프 파일을 생성합니다.

구문

```
gcore ID
```

***ID** 를 이전 단계에서 얻은 프로세스의 **ID**로 바꿉니다(예: **18110**):*

예제

```
[root@host01 /]# gcore 18110
warning: target file /proc/18110/cmdline contained unexpected null characters
Saved corefile core.18110
```

e.

코어 덤프 파일이 올바르게 생성되었는지 확인합니다.

예제

```
[root@host01 /]# ls -ltr
total 709772
-rw-r--r--. 1 root root 726799544 Mar 18 19:46 core.18110
```

f. ***Ceph Monitor** 컨테이너 외부에 코어 덤프 파일을 복사합니다.*

구문

```
podman cp ceph-mon-MONITOR_ID:/tmp/mon.core.MONITOR_PID /tmp
```

***MONITOR_ID** 를 **Ceph Monitor**의 **ID** 번호로 바꾸고 **MONITOR_PID** 를 프로세스 **ID** 번호로 교체합니다.*

3. *다른 **Ceph** 데몬에 대한 코어 덤프 파일을 수동으로 캡처하려면 다음을 수행합니다.*

a. ***cephadm** 셸에 로그인합니다.*

예제

```
[root@host03 ~]# cephadm shell
```

b. *데몬에 대해 **ptrace** 를 활성화합니다.*

예제

```
[ceph: root@host01 /]# ceph config set mgr mgr/cephadm/allow_ptrace true
```

c.

데몬 서비스를 재배포합니다.

구문

```
ceph orch redeploy SERVICE_ID
```

예제

```
[ceph: root@host01 /]# ceph orch redeploy mgr
[ceph: root@host01 /]# ceph orch redeploy rgw.rgw.1
```

d.

***cephadm** 쉘 을 종료하고 데몬이 배포된 호스트에 로그인합니다.*

예제

```
[ceph: root@host01 /]# exit
[root@host01 ~]# ssh root@10.0.0.11
```

e.

***DAEMON_ID** 를 가져오고 컨테이너를 입력합니다.*

예제

```
[root@host04 ~]# podman ps
[root@host04 ~]# podman exec -it ceph-1ca9f6a8-d036-11ec-8263-fa163ee967ad-rgw-rgw-1-host04 bash
```

f. ***procps-ng 및 gdb 패키지를 설치합니다.***

예제

```
[root@host04 /]# dnf install procps-ng gdb
```

g. *프로세스의* ***PID****를 가져옵니다.*

예제

```
[root@host04 /]# ps aux | grep rados
ceph           6  0.3  2.8 5334140 109052 ?      Sl   May10   5:25 /usr/bin/radosgw -n client.rgw.rgw.1.host04 -f --setuser ceph --setgroup ceph --default-log-to-file=false --default-log-to-stderr=true --default-log-stderr-prefix=debug
```

h. *코어 덤프 수집:*

구문

```
gcore PID
```

예제

```
[root@host04 /]# gcore 6
```

i. *코어 덤프 파일이 올바르게 생성되었는지 확인합니다.*

예제

```
[root@host04 /]# ls -ltr
total 108798
-rw-r--r--. 1 root root 726799544 Mar 18 19:46 core.6
```

j. *컨테이너 외부에 코어 덤프 파일을 복사합니다.*

구문

```
podman cp ceph-mon-DAEMON_ID:/tmp/mon.core.PID /tmp
```

***DAEMON_ID** 를 **Ceph** 데몬의 **ID** 번호로 교체하고 **PID** 를 프로세스 **ID** 번호로 바꿉니다.*

4. *분석을 위해 코어 덤프 파일을 **Red Hat** 지원 케이스에 업로드합니다. 자세한 내용은 **Red Hat** 지원 엔지니어에게 정보 제공을 참조하십시오.*

10.3.3. 추가 리소스

-

How to use gdb to generate a readable backtrace from an application core **solution on the Red Hat Customer Portal**

- **Red Hat** 고객 포털에서 *애플리케이션이 충돌하거나 세그멘테이션 결함 솔루션을 사용할 때 코어 파일 덤프를 활성화하는 방법*

부록 A. CEPH 하위 시스템의 기본 로깅 수준 값

다양한 Ceph 하위 시스템의 기본 로깅 수준 값 테이블입니다.

하위 시스템	로그 수준	메모리 수준
asok	1	5
auth	1	5
buffer	0	0
클라이언트	0	5
context	0	5
crush	1	5
default	0	5
Filer	0	5
bluestore	1	5
finisher	1	5
heartbeatmap	1	5
javaclient	1	5
journaler	0	5
journal	1	5
lockdep	0	5
MDS 밸런서	1	5
mds locker	1	5
MDS 로그 만료	1	5
MDS 로그	1	5
MDS migrator	1	5

하위 시스템	로그 수준	메모리 수준
mds	1	5
monc	0	5
월요일	1	5
ms	0	5
objclass	0	5
objectcacher	0	5
objecter	0	0
optracker	0	5
osd	0	5
Paxos	0	5
perfcounter	1	5
rados	0	5
rbd	0	5
rgw	1	5
throttle	1	5
timer	0	5
tp	0	5

부록 B. CEPH 클러스터의 상태 메시지

***Red Hat Ceph Storage** 클러스터에서 늘릴 수 있는 일련의 가능한 상태 메시지가 표시됩니다. 이들은 고유 식별자를 가진 상태 검사로 정의됩니다. 식별자는 도구를 사용하여 상태 점검을 감지하고 의미를 반영하는 방식으로 제시할 수 있도록 설계된**terse pseudo-human-readable** 문자열입니다.*

*표 **B.1.** 모니터*

상태 코드	설명
DAEMON_OLD_VERSION	데몬에서 이전 버전의 Ceph가 실행 중인지 경고합니다. 여러 버전이 감지되면 상태 오류가 생성됩니다.
MON_DOWN	하나 이상의 Ceph Monitor 데몬이 현재 다운되었습니다.
MON_CLOCK_SKEW	**ceph-mon** 데몬을 실행하는 노드의 클럭은 충분히 동기화되지 않습니다. **ntpd** 또는 **chrony** 를 사용하여 클럭을 동기화하여 해결합니다.
MON_MSGR2_NOT_ENABLED	**ms_bind_msgr2** 옵션이 활성화되어 있지만 클러스터의 monmap에서 v2 포트에 바인딩하도록 하나 이상의 Ceph Monitor가 구성되지 않았습니다. **ceph mon enable-msgr2** 명령을 실행하여 이 문제를 해결합니다.
MON_DISK_LOW	하나 이상의 Ceph 모니터가 디스크 공간보다 낮습니다.
MON_DISK_CRIT	하나 이상의 Ceph 모니터는 디스크 공간의 매우 낮은 수준입니다.
MON_DISK_BIG	하나 이상의 Ceph 모니터의 데이터베이스 크기가 매우 큽니다.
AUTH_INSECURE_GLOBAL_ID_RECLAIM	하나 이상의 클라이언트 또는 데몬은 Ceph 모니터에 다시 연결할 때 **global_id** 를 안전하게 회수하지 않는 스토리지 클러스터에 연결됩니다.
AUTH_INSECURE_GLOBAL_ID_RECLAIM_ALLOWED	현재 Ceph는 **auth_allow_insecure _global_ reclaim이 true 로 설정되었기 때문에 안전하지 않은 프로세스를 사용하여 모니터에 다시 연결할** 수 있도록 구성되어 있습니다.

*표 **B.2.** 관리자*

상태 코드	설명
MGR_DOWN	현재 모든 Ceph Manager 데몬이 종료되었습니다.
MGR_MODULE_DEPENDENCY	활성화된 Ceph Manager 모듈이 종속성 확인에 실패했습니다.
MGR_MODULE_ERROR	Ceph Manager 모듈에 예기치 않은 오류가 발생했습니다. 일반적으로 이는 모듈 serve 함수에서 처리되지 않은 예외가 발생했음을 의미합니다.

*표 **B.3. OSDs***

상태 코드	설명
OSD_DOWN	하나 이상의 OSD가 축소되었습니다.
OSD_CRUSH_TYPE_DOWN	특정 CRUSH 하위 트리 내의 모든 OSD가 다운된 상태(예: 호스트의 모든 OSD)가 표시됩니다. 예를 들어 OSD_HOST_DOWN 및 OSD_ROOT_DOWN
OSD_ORPHAN	OSD는 CRUSH 맵 계층에서 참조되지만 존재하지 않습니다. **ceph osd crush rm osd._OSD_ID** 명령을 실행하여 OSD를 제거합니다.
OSD_OUT_OF_ORDER_FULL	*가까운full*,*backfillfull*, full 또는 *failsafefull*의 사용률 임계값은 오름차순이 아닙니다. **ceph osd set-nearfull-ratio *RATIO*, ceph osd set-backfillfull-ratio *RATIO*** 및 **ceph osd set-full-ratio *RATIO***를 실행하여 임계값을 조정합니다.
OSD_FULL	하나 이상의 OSD가 전체 임계값을 초과했으며 스토리지 클러스터가 쓰기 작업을 수행하지 못하도록 합니다. 전체 임계값을 작은 margin **ceph osd set-full-ratio *RATIO*** 로 높여 쓰기 가용성을 복원합니다.
OSD_BACKFILLFULL	하나 이상의 OSD가 백필full 임계값을 초과하여 데이터가 이 장치로 리밸런스될 수 없도록 합니다.
OSD_NEARFULL	하나 이상의 OSD가 가까운full 임계값을 초과했습니다.
OSDMAP_FLAGS	관심 있는 하나 이상의 스토리지 클러스터 플래그가 설정되었습니다. 이러한 플래그에는 *전체*,*pauserd*,*pauserd*,*no up*,*noin*,*noout*,*norecover*,*norebalance*,*no rebalance*,*nodeep_scrub*, *notieragent* 등이 있습니다. *전체*를 제외하고, **ceph osd set *FLAG*** 및 **ceph osd unset *FLAG* 명령을 사용하여 플래그**를 지울 수 있습니다.

상태 코드	설명
OSD_FLAGS	하나 이상의 OSD 또는 CRUSH에 관심 플래그가 설정되어 있습니다. 이 플래그에는 *noup*,*nodown*,*noin*, *noout* 이 포함됩니다.
OLD_CRUSH_TUNABLES	CRUSH 맵은 매우 오래된 설정을 사용하며 업데이트해야 합니다.
OLD_CRUSH_STRAW_CALC_VERSION	CRUSH 맵은 **straw** 버킷에 대한 중간 가중치 값을 계산하기 위해 최적화되지 않은 이전 방법을 사용합니다.
CACHE_POOL_NO_HIT_SET	하나 이상의 캐시 풀은 사용률을 추적하도록 구성된 적정으로 구성되지 않으므로 계층화 에이전트가 캐시에서 플러시 및 제거될 cold 오브젝트를 식별할 수 없습니다. **ceph osd 풀 set_ *POOL_NAME* _ hit_set_type *TYPE*, ceph osd pool set *POOL_NAME* hit_set_period *PERIOD_SECONDS*, ceph osd pool set POOL_count *NUMBER_OF_HIT_SETS*를 사용하여 캐시 풀에서 적중** 세트를 구성합니다. 및 **ceph osd 풀은 *POOL_NAME* hit_set_fpp EXTRAGET *_FALSE_POSITIVE_RATE*** 명령을 설정합니다.
OSD_NO_SORTBITWISE	**sortbitwise** 플래그가 설정되어 있지 않습니다.**ceph osd set sortbitwise** 명령을 사용하여 플래그를 설정합니다.
POOL_FULL	하나 이상의 풀이 할당량에 도달했으며 더 이상 쓰기를 허용하지 않습니다. **ceph osd 풀 set-quota *POOL_NAME* max_objects *NUMBER_OF_OBJECTS*** 및 **ceph osd pool set-quota *POOL_NAME* max_bytes *BYTES*** 를 사용하여 풀 할당량을 늘리거나 사용률을 줄이기 위해 기존 데이터를 삭제합니다.
BLUEFS_SPILLOVER	BlueStore 백엔드를 사용하는 하나 이상의 OSD는 db 파티션이 할당되었지만 해당 공간에는 일반 느린 장치에 "침착"이 있습니다. **ceph config set osd bluestore_warn_on_bluefs_spillover false** 명령을 사용하여 비활성화합니다.
BLUEFS_AVAILABLE_SPACE	이 출력은 *BDEV_DB 무료,BDEV_SLOW 무료 및 available_ from_bluestore* 의 세 가지 값을 제공합니다.
BLUEFS_LOW_SPACE	BlueStore File System (BlueFS)이 사용 가능한 공간에 부족하고 사용 **가능한_from_bluestore** 가 있는 경우 BlueFS 할당 단위 크기를 줄일 수 있습니다.

상태 코드	설명
BLUESTORE_FRAGMENTATION	BlueStore는 기본 스토리지에서 사용 가능한 공간을 작동하므로 조각화됩니다. 이는 정상적이고 피할 수 없지만 과도한 조각화는 느려집니다.
BLUESTORE_LEGACY_STATFS	bluestore는 풀 단위로 내부 사용량 통계를 추적하며 하나 이상의 OSD에는 BlueStore 볼륨이 있습니다. **ceph config set global bluestore_warn_on_legacy_statfs false** 명령을 사용하여 경고를 비활성화합니다.
BLUESTORE_NO_PER_POOL_OMAP	bluestore는 풀별 omap 공간 사용률을 추적합니다. **ceph config set global bluestore_warn_on_no_per_pool_omap false** 명령을 사용하여 경고를 비활성화합니다.
BLUESTORE_NO_PER_PG_OMAP	bluestore는 PG에 의해 omap 공간 사용률을 추적합니다. **ceph config set global bluestore_warn_on_no_per_pg_omap false** 명령을 사용하여 경고를 비활성화합니다.
BLUESTORE_DISK_SIZE_MISMATCH	BlueStore를 사용하는 하나 이상의 OSD에는 물리 장치의 크기와 메타데이터를 추적하는 메타데이터 사이에 내부 불일치가 있습니다.
BLUESTORE_NO_COMPRESSION `	하나 이상의 OSD가 BlueStore 압축 플러그인을 로드할 수 없습니다. 이는 손상된 설치로 인해 **ceph-osd** 바이너리가 압축 플러그인과 일치하지 않거나 **ceph-osd** 데몬 재시작을 포함하지 않은 최근 업그레이드로 인해 발생할 수 있습니다.
BLUESTORE_SPURIOUS_READ_ERRORS	BlueStore를 사용하는 하나 이상의 OSD는 메인 장치에서 오래된 읽기 오류를 감지합니다. bluestore는 디스크 읽기를 다시 시도하여 이러한 오류로부터 복구되었습니다.

*표 **B.4.** 장치 상태*

상태 코드	설명
DEVICE_HEALTH	하나 이상의 장치가 곧 실패할 것으로 예상되며, 여기서 경고 임계값은 **mgr/devicehealth/warn_threshold** 구성 옵션에 의해 제어됩니다. 장치를 *출력* 하여 데이터를 마이그레이션하고 하드웨어를 교체합니다.

상태 코드	설명
DEVICE_HEALTH_IN_USE	하나 이상의 장치가 곧 실패할 것으로 예상되며 **mgr/devicehealth/mark_out_threshold** 를 기반으로 스토리지 클러스터의 "out"으로 표시되었지만 여전히 더 많은 PG에 참여하고 있습니다.
DEVICE_HEALTH_TOOMANY	너무 많은 장치가 곧 실패 할 것으로 예상되고 **mgr/devicehealth/self_heal** 동작이 활성화되어 있으며 모든 실용 장치를 표시하는 것이 너무 많은 OSD가 자동으로 표시되지 않도록 클러스터 **mon_osd_min_in_ratio** 비율을 초과합니다.

표 B.5. 풀 및 배치 그룹

상태 코드	설명
PG_AVAILABILITY	데이터 가용성이 감소되므로 스토리지 클러스터에서 클러스터의 일부 데이터에 대한 잠재적인 읽기 또는 쓰기 요청을 서비스할 수 없습니다.
PG_DEGRADED	일부 데이터에 대해 데이터 중복성이 감소되므로 스토리지 클러스터에 복제된 풀에 대해 원하는 개수의 복제본이 없거나 코드 조각이 삭제됩니다.
PG_RECOVERY_FULL	스토리지 클러스터에서 여유 공간이 부족하기 때문에 데이터 중복성을 줄일 수 있습니다. 특히 하나 이상의 PGs에 **recovery_too full** 플래그가 설정되어 있어 하나 이상의 OSD가 전체 임계값 이상이므로 클러스터를 마이그레이션하거나 복구할 수 없습니다.
PG_BACKFILL_FULL	스토리지 클러스터에서 사용 가능한 공간이 부족하거나 일부 데이터의 경우 데이터 중복성을 줄일 수 있습니다. 특히 하나 이상의 PGs에 **backfill_toofull** 플래그가 설정되어 있어 하나 이상의 OSD가 **백fillfull** 임계값 이상이므로 클러스터를 마이그레이션하거나 복구할 수 없습니다.
PG_DAMAGED	데이터 스크럽링은 스토리지 클러스터의 데이터 일관성, 특히 하나 이상의 PG가 일관되지 않거나 **스냅인** 업이 문제를 발견했거나, **복구** 플래그가 설정되어 있거나, 이러한 불일치의 복구가 현재 진행 중인 것으로 나타났습니다.

상태 코드	설명
OSD_SCRUB_ERRORS	최근 OSD Scrubs가 불일치를 발견했습니다.
OSD_TOO_MANY_REPAIRS	읽기 오류가 발생하고 다른 복제본을 사용할 수 있으면 즉시 오류를 복구하여 클라이언트에서 개체 데이터를 가져올 수 있습니다.
LARGE_OMAP_OBJECTS	하나 이상의 풀에는 o **sd_deep_scrub_large_omap_object_threshold 또는 osd_ deep_scrub_large_object_value_sum_threshold 또는 둘 다에 따라 결정되는 큰 omap 오브젝트** 가 포함됩니다. **ceph config set osd_deep_scrub_large_omap_object_key_threshold *KEYS*** 및 **ceph config set osd_deep_scrub_large_large_omap_object_sum_threshold *BYTES*** 명령을 사용하여 임계값을 조정합니다.
CACHE_POOL_NEAR_FULL	캐시 계층 풀은 거의 가득 차 있습니다. **ceph osd pool set *CACHE_POOL_NAME* target_max_bytes *BYTES*** 및 **ceph osd pool set *CACHE_POOL_NAME* target_max_bytes *BYTES*** 명령을 사용하여 캐시 풀 대상 크기를 조정합니다.
TOO_FEW_PGS	스토리지 클러스터에서 사용 중인 PG의 수는 OSD당 **mon_pg_warn_min_per_osd** PG의 구성 가능한 임계값보다 낮습니다.
POOL_PG_NUM_NOT_POWER_OF_TWO	하나 이상의 풀에는 2의 전원이 아닌 **pg_num** 값이 있습니다. **ceph config set global mon_warn_on_pool_pg_num_not_power_of_two false** 명령을 사용하여 경고를 비활성화합니다.
POOL_TOO_FEW_PGS	하나 이상의 풀에는 현재 풀에 저장된 데이터의 양에 따라 더 많은 PG가 있어야 합니다. **ceph osd pool set POOL_autoscale_mode off 명령을 사용하여 PG의 자동 확장을 비활성화하고, ceph os d pool set *POOL_* NAME pg_autoscale_mode를 사용하여 PG의 수를 자동으로 조정하거나 ceph osd pool set *POOL_ NAME* pg_num _NEW_PG_ NUMER 명령을 사용하여 PG의 수** 를 수동으로 설정할 수 있습니다.
TOO_MANY_PGS	스토리지 클러스터에서 사용 중인 PG의 수는 OSD당 **mon_max_pg_per_osd** PGs의 구성 가능한 임계값보다 큽니다. 하드웨어를 추가하여 클러스터에서 OSD 수를 늘립니다.

상태 코드	설명
POOL_TOO_MANY_PGS	하나 이상의 풀에는 현재 풀에 저장된 데이터의 양에 따라 더 많은 PG가 있어야 합니다. **ceph osd pool set POOL_autoscale_mode off 명령을 사용하여 PG의 자동 확장을 비활성화하고, ceph os d pool set *POOL_* NAME pg_autoscale_mode를 사용하여 PG의 수를 자동으로 조정하거나 ceph osd pool set *POOL_ NAME* pg_num _NEW_PG_ NUMER 명령을 사용하여 PG의 수**를 수동으로 설정할 수 있습니다.
POOL_TARGET_SIZE_BYTES_OVERCOMMITTED	하나 이상의 풀에는 **target_size_bytes** 속성이 설정되어 풀의 예상 크기를 추정하지만 값은 사용 가능한 총 스토리지를 초과합니다. **ceph osd pool set *POOL_NAME* target_size_bytes 0 명령을 사용하여 풀의 값을 0** 으로 설정합니다.
POOL_HAS_TARGET_SIZE_BYTES_AND_RATIO	하나 이상의 풀에 **target_size_bytes** 및 **target_size_ratio** 가 모두 풀의 예상 크기를 추정하도록 설정되어 있습니다. **ceph osd pool set *POOL_NAME* target_size_bytes 0 명령을 사용하여 풀의 값을 0** 으로 설정합니다.
TOO_FEW_OSDS	스토리지 클러스터의 OSD 수는 **o'sd_pool_default_size** 의 구성 가능한 임계값 미만입니다.
SMALLER_PGP_NUM	하나 이상의 풀에 **pgp_num** 보다 작은 pgp**_num** 값이 있습니다. 이는 일반적으로 PG 수가 배치 동작을 늘리지 않고 증가했음을 나타냅니다. **pgp_num** 을 **ceph osd pool set *POOL _ NAME* pgp_num *PG_NUM_VALUE*** 명령과 일치하도록 설정합니다.
MANY_OBJECTS_PER_PG	하나 이상의 풀은 전체 스토리지 클러스터 평균보다 훨씬 높은 PG당 평균 오브젝트 수를 갖습니다. 특정 임계값은 **mon_pg_warn_max_object_skew** 구성 값으로 제어됩니다.
POOL_APP_NOT_ENABLED	하나 이상의 오브젝트가 포함되어 있지만 특정 애플리케이션에서 사용하도록 태그가 지정되지 않은 풀이 있습니다. **rbd 풀 init *POOL_NAME*명령이 있는 애플리케이션에서 사용할 풀에** 레이블을 지정하여 이 경고를 해결합니다.
POOL_FULL	하나 이상의 풀이 할당량에 도달했습니다. 이 오류 조건을 트리거하는 임계값은 **mon_pool_quota_crit_threshold** 구성 옵션으로 제어합니다.

상태 코드	설명
POOL_NEAR_FULL	하나 이상의 풀이 구성된 완전성 임계값에 접근하고 있습니다. **ceph osd pool set-quota *POOL_NAME* max_objects *NUMBER_OF_OBJECTS*** 및 **ceph osd pool set-quota *POOL_NAME* max_bytes *BYTES*** 명령으로 풀 할당량을 조정합니다.
OBJECT_MISPLACED	스토리지 클러스터에서 하나 이상의 오브젝트가 스토리지 클러스터를 저장하려는 노드에 저장되지 않습니다. 이는 일부 최근 스토리지 클러스터 변경으로 인한 데이터 마이그레이션이 아직 완료되지 않았음을 나타냅니다.
OBJECT_UNFOUND	스토리지 클러스터에 있는 하나 이상의 오브젝트를 찾을 수 없습니다. 특히 OSD는 새 개체 또는 업데이트된 개체 복사본이 있어야 하지만 현재 온라인 상태인 OSD에서 해당 오브젝트 버전의 복사본을 찾을 수 없다는 것을 알고 있습니다.
SLOW_OPS	하나 이상의 OSD 또는 모니터 요청을 처리하는 데 시간이 오래 걸립니다. 이는 극단적인 부하, 느린 저장 장치 또는 소프트웨어 버그의 표시일 수 있습니다.
PG_NOT_SCRUBBED	하나 이상의 PG가 최근에 스크럽되지 않았습니다. PGS는 일반적으로 **osd_scrub_max_interval** 에 의해 지정된 모든 간격 내에서 스크럽됩니다. **ceph pg scrub *PG_ID*** 명령을 사용하여 scrub를 시작합니다.
PG_NOT_DEEP_SCRUBBED	하나 이상의 PG가 최근 딥 스크럽되지 않았습니다. **ceph pg deep-scrub *PG_ID*** 명령을 사용하여 scrub을 시작합니다. PGS는 일반적으로 모든 **osd_deep_scrub_interval** 초를 스크럽하며, **mon_warn_pg_not_deep_deep_scrued_ratio** 간격의 백분율이 0crub 없이 경과했을 때 이 경고가 트리거됩니다.
PG_SLOW_SNAP_TRIMMING	하나 이상의 PGs에 대한 스냅샷 트리밍 큐는 구성된 경고 임계값을 초과했습니다. 이는 최근에 많은 수의 스냅샷이 삭제되었거나 OSD가 스냅샷을 빠르게 트리밍하여 새 스냅샷 삭제 속도를 유지할 수 없음을 나타냅니다.

*표 **B.6.** 기타*

상태 코드	설명

상태 코드	설명
RECENT_CRASH	최근에 하나 이상의 Ceph 데몬이 중단되었으며, 해당 충돌은 관리자가 아직 확인하지 않았습니다.
TELEMETRY_CHANGED	Telemetry가 활성화되어 있지만 Telemetry 보고서의 콘텐츠가 그 이후로 변경되어 Telemetry 보고서가 전송되지 않습니다.
AUTH_BAD_CAPS	하나 이상의 인증 사용자에게 모니터에서 구문 분석할 수 없는 기능이 있습니다. **ceph auth *ENTITY_NAME DAEMON_TYPE CAPS*** 명령을 사용하여 사용자의 기능을 업데이트합니다.
OSD_NO_DOWN_OUT_INTERVAL	**mon_osd_down_out_interval** 옵션이 0으로 설정되어 있으므로 OSD가 실패한 후 시스템이 복구 또는 복구 작업을 자동으로 수행하지 않습니다. **ceph config global mon_warn_on_osd_down_out_zero false 명령으로 간격** 을 음소거합니다.
DASHBOARD_DEBUG	대시보드 디버그 모드가 활성화되어 있습니다. 즉, REST API 요청을 처리하는 동안 오류가 있는 경우 HTTP 오류 응답에 Python traceback이 포함됩니다. **ceph dashboard debug disable** 명령을 사용하여 디버그 모드를 비활성화합니다.