Red Hat Ceph Storage 5

하드웨어 가이드

Red Hat Ceph Storage의 하드웨어 선택 권장 사항

Last Updated: 2022-08-15

Red Hat Ceph Storage 5 하드웨어 가이드

Red Hat Ceph Storage의 하드웨어 선택 권장 사항

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 Storage와 함께 사용할 하드웨어 선택에 대한 높은 수준의 지침을 제공합니다. 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. 요약

많은 하드웨어 벤더가 이제 Ceph 최적화 서버 및 개별 워크로드 프로파일을 위해 설계된 랙 수준 솔루션을 모두 제공합니다. Red Hat은 하드웨어 선택 프로세스를 단순화하고 조직의 위험을 줄이기 위해 여러 스토리지 서버 벤더와 협력하여 다양한 클러스터 크기 및 워크로드 프로필에 대한 특정 클러스터 옵션을 테스트 및 평가했습니다. Red Hat의 정확한 방법론은 성능 테스트와 광범위한 클러스터 기능 및 크기에 대한 입증된 지침을 결합합니다. 적절한 스토리지 서버 및 랙 레벨 솔루션을 통해 Red Hat Ceph Storage는 처리량이 민감하고 비용 및 용량 중심 워크로드부터 새로운 IOPS 집약적 워크로드에 이르기까지 다양한 워크로드를 지원하는 스토리지 풀을 제공할 수 있습니다.

Red Hat Ceph Storage는 엔터프라이즈 데이터 저장 비용을 대폭 절감하고 조직이 기하급수적인 데이터 증가를 관리할 수 있도록 지원합니다. 이 소프트웨어는 퍼블릭 또는 프라이빗 클라우드 배포를 위한 강력하고 현대적인 페타바이트 규모의 스토리지 플랫폼입니다. Red Hat Ceph Storage는 엔터프라이즈 블록 및 개체 스토리지를 위한 완성도 높은 인터페이스를 제공하여 테넌트와 무관한 OpenStack® 환경을 특징으로 하는 액티브 아카이브, 리치 미디어 및 클라우드 인프라 워크로드를 위한 최적의 솔루션입니다. [1]. 통합 소프트웨어 정의 스케일 아웃 스토리지 플랫폼으로 제공되는 Red Hat Ceph Storage는 다음과 같은 기능을 제공하여 애플리케이션 혁신과 가용성을 향상시키는 데 집중할 수 있습니다.

- 수백 페타바이트 규모로 확장 [2].
- 클러스터의 단일 장애 지점 없음.
- 상용 서버 하드웨어에서 실행하여 자본 지출(CapEx) 절감.
- 자체 관리 및 자동 복구 속성으로 낮은 운영 비용(OpEx).

Red Hat Ceph Storage는 다양한 요구 사항을 충족하기 위해 수많은 업계 표준 하드웨어 구성에서 실행할 수 있습니다. Red Hat은 클러스터 설계 프로세스를 단순화하고 가속화하기 위해 참여 하드웨어 벤더를 통해 광범위한 성능 및 적합성 테스트를 수행합니다. 이 테스트를 통해 선택한 하드웨어를 부하에서 평가할 수 있으며 다양한 워크로드에 대해 필수적인 성능 및 크기 조정 데이터를 생성할 수 있습니다. 이는 Ceph 스토리지 클러스터 하드웨어 선택을 궁극적으로 간소화합니다. 이 가이드에서 설명한 대로 다중 하드웨어 벤더는 이제 IOPS, 처리량-, 비용 및 용량에 최적화된 솔루션을 사용 가능한 옵션으로 Red Hat Ceph Storage 배포에 최적화된 서버 및 랙 수준 솔루션을 제공합니다.

소프트웨어 정의 스토리지는 까다로운 애플리케이션 및 스토리지 요구 사항을 충족하기 위해 스케일 아웃 솔루션을 원하는 조직에 많은 이점을 제공합니다. 검증된 방법론과 다양한 벤더에서 수행한 광범위한 테스트를 통해 Red Hat은 모든 환경의 요구를 충족하기 위해 하드웨어를 선택하는 프로세스를 단순화합니다. 중요한 것은 이 문서에 나열된 지침과 예제 시스템은 샘플 시스템에서 프로덕션 워크로드의 영향을 수치화하기 위한 대체 방법이 아니라는 점입니다.

---

[1] Ceph는 반년 간 OpenStack 사용자 설문 조사에 따르면 OpenStack의 선도적인 스토리지였습니다.

[2] 자세한 내용은 Yahoo Cloud Object Store - Exabyte Scale의 Object Storage 를 참조하십시오.

2장. 하드웨어 선택에 대한 일반적인 원칙

스토리지 관리자는 프로덕션 Red Hat Ceph Storage 클러스터를 실행하기 위한 적절한 하드웨어를 선택해야 합니다. Red Hat Ceph Storage용 하드웨어를 선택하는 경우 다음 일반 원칙을 검토하십시오. 이러한 원칙은 시간을 절약하고, 일반적인 실수를 피하고, 비용을 절약하고, 보다 효과적인 솔루션을 얻는 데 도움이 됩니다.

2.1. 사전 요구 사항

- 계획된 Red Hat Ceph Storage 사용.
- Red Hat Enterprise Linux 인증을 통한 Linux System Administration Advance 레벨.
- Ceph 인증이 있는 스토리지 관리자입니다.

2.2. 성능 사용 사례 확인

성공적인 Ceph 배포의 가장 중요한 단계 중 하나는 클러스터의 사용 사례와 워크로드에 적합한 가격대 성능 프로필을 식별하는 것입니다. 사용 사례에 적합한 하드웨어를 선택하는 것이 중요합니다. 예를 들어 클라우드 스토리지 애플리케이션에서 IOPS 최적화 하드웨어를 선택하면 하드웨어 비용이 불필요하게 증가합니다. 반면, IOPS 집약적인 워크로드에서 용량에 최적화된 하드웨어를 선택하는 경우 성능 저하에 대해 불만을 제기하는 사용자가 불만을 줄 수 있습니다.

Ceph의 주요 사용 사례는 다음과 같습니다.

- **최적화된 IOPS:** IOPS 최적화된 배포는 MYSQL 또는 MariaDB 인스턴스를 OpenStack의 가상 시스템으로 실행하는 등 클라우드 컴퓨팅 작업에 적합합니다. IOPS 최적화된 배포를 위해서는 15k RPM SAS 드라이브와 같은 고성능 스토리지와 빈번한 쓰기 작업을 처리하기 위해 별도의 SSD 저널이 필요합니다. 일부 높은 IOPS 시나리오에서는 모든 플래시 스토리지를 사용하여 IOPS 및 총 처리량을 개선합니다.
- **처리량 최적화:** 처리량 최적화 배포는 그래픽, 오디오 및 비디오 컨텐츠와 같이 상당한 양의 데이터를 처리하는 데 적합합니다. 처리량 최적화된 배포를 위해서는 전체 처리량 특성을 갖춘 네트워킹 하드웨어, 컨트롤러 및 하드 디스크 드라이브가 필요합니다. 쓰기 성능이 요구되는 경우 SSD 저널은 쓰기 성능을 크게 개선합니다.
- **최적화된 용량:** 용량에 최적화된 배포는 상당한 양의 데이터를 최대한 저렴하게 저장하는 데 적합합니다. 용량 최적화된 배포는 일반적으로 매력적인 가격대를 위해 성능을 교환합니다. 예를 들어 용량 최적화된 배포에서는 저널링에 SSD를 사용하는 대신 더 느리고 값비싼 SATA 드라이브를 사용하고 저널을 공동 배치하는 경우가 많습니다.

이 문서에서는 이러한 사용 사례에 적합한 Red Hat 테스트 하드웨어의 예를 설명합니다.

2.3. 스토리지 밀도 고려

하드웨어 계획에는 하드웨어 장애 발생 시 고가용성을 유지하기 위해 Ceph 데몬과 Ceph를 여러 호스트에 사용하는 기타 프로세스를 배포해야 합니다. 하드웨어 장애 발생 시 클러스터를 리밸런싱해야 하는 필요성과 함께 스토리지 밀도의 균형을 조정합니다. 일반적인 하드웨어 선택 실수는 백필(backfill) 및 복구 작업 중에 네트워킹에 과부하를 줄 수 있는 소규모 클러스터에서 매우 높은 스토리지 밀도를 사용하는 것입니다.

2.4. 동일한 하드웨어 구성

풀을 생성하고 풀 내의 OSD 하드웨어가 동일하도록 CRUSH 계층 구조를 정의합니다.

- 동일한 컨트롤러.
- 동일한 드라이브 크기.
- 동일한 RPM.
- 같은 검색 시간.
- 동일한 I/O.
- 동일한 네트워크 처리량.
- 동일한 저널 구성.

풀 내에서 동일한 하드웨어를 사용하면 일관된 성능 프로필을 제공하여 프로비저닝을 단순화하고 문제 해결을 간소화합니다.

2.5. RED HAT CEPH STORAGE의 네트워크 고려 사항

클라우드 스토리지 솔루션의 중요한 측면은 네트워크 대기 시간 및 기타 요인으로 인해 스토리지 클러스터가 IOPS에서 실행될 수 있다는 것입니다. 또한 스토리지 클러스터에 스토리지 용량이 부족하기 전에 대역폭 제약으로 인해 처리량이 부족할 수 있습니다. 즉, 네트워크 하드웨어 구성이 가격 대비 성능 요구 사항을 충족하기 위해 선택한 워크로드를 지원해야 합니다.

스토리지 관리자는 스토리지 클러스터를 최대한 빨리 복구하는 것을 선호합니다. 스토리지 클러스터 네트워크의 대역폭 요구 사항을 신중하게 고려하고 네트워크 링크를 덮어 쓰기하고 클라이언트-클러스터 간 트래픽에서 클러스터 내부 트래픽을 분리합니다. 또한 SSD(Solid State Disks), 플래시, NVMe 및 기타 고성능 스토리지 장치의 사용을 고려할 때 네트워크 성능이 점차 중요한 것으로 간주합니다.

Ceph는 공용 네트워크 및 스토리지 클러스터 네트워크를 지원합니다. 공용 네트워크는 Ceph 모니터와의 통신 및 클라이언트 트래픽을 처리합니다. 스토리지 클러스터 네트워크는 Ceph OSD 하트비트, 복제, 백필, 복구 트래픽을 처리합니다. **최소한** 하나의 10GB 이더넷 링크를 스토리지 하드웨어에 사용해야 하며 연결 및 처리량을 위해 10GB 이더넷 링크를 추가할 수 있습니다.

중요

Red Hat은 복제된 풀에서 여러 기본으로 **osd_pool_default_size** 를 사용하여 공용 네트워크의 배수가 되도록 대역폭을 스토리지 클러스터 네트워크에 할당할 것을 권장합니다. 또한 별도의 네트워크 카드에서 공용 및 스토리지 클러스터 네트워크를 실행하는 것이 좋습니다.

중요

프로덕션 시 Red Hat Ceph Storage 배포에는 10GB 이더넷을 사용하는 것이 좋습니다. 1GB 이더넷 네트워크는 프로덕션 스토리지 클러스터에 적합하지 않습니다.

드라이브 오류가 발생하는 경우 1GB 이더넷 네트워크에서 1TB의 데이터를 복제하는 데 3시간이 걸리며 3TB의 데이터를 9시간이 소요됩니다. 3TB를 사용하는 것이 일반적인 드라이브 구성입니다. 반면 10GB 이더넷 네트워크를 사용하면 복제 시간은 20분과 1시간입니다. Ceph OSD가 실패하면 풀 내의 다른 Ceph OSD에 포함된 데이터를 복제하여 스토리지 클러스터를 복구합니다.

랙과 같은 대규모 도메인에 장애가 발생하면 스토리지 클러스터가 훨씬 더 많은 대역폭을 사용한다는 것을 의미합니다. 대규모 스토리지 구현에 공통적인 여러 랙으로 구성된 스토리지 클러스터를 구축할 때 최적의 성능을 위해 "패트 트리" 설계의 스위치 간 네트워크 대역폭을 최대한 활용하는 것이 좋습니다. 일반적인 10GB 이더넷 스위치에는 48개의 10GB 포트와 40GB 포트 4개가 있습니다. 최대 처리량을 위해 스파

인에서 40GB 포트를 사용합니다. 또는 다른 랙 및 스파인 라우터에 연결하기 위해 사용하지 않는 10GB 포트를 40GB 이상의 포트로 집계하는 것이 좋습니다. 또한 네트워크 인터페이스를 결합하는 데 LACP 모드 4를 사용하는 것이 좋습니다. 또한, 특히 백엔드 또는 클러스터 네트워크에서 최대 전송 단위(MTU)가 9000인 점보 프레임을 사용합니다.

Red Hat Ceph Storage 클러스터를 설치하고 테스트하기 전에 네트워크 처리량을 확인합니다. Ceph에서 대부분의 성능 관련 문제는 일반적으로 네트워킹 문제로 시작합니다. kinked 또는 bent cat-6 케이블과 같은 간단한 네트워크 문제로 인해 대역폭이 저하될 수 있습니다. 전면 네트워크에 최소 10GB 이더넷을 사용합니다. 대규모 클러스터의 경우 백엔드 또는 클러스터 네트워크에 40GB 이더넷을 사용하는 것이 좋습니다.

중요

네트워크 최적화를 위해 Red Hat은 대역폭당 CPU를 개선하고 차단되지 않는 네트워크 스위치 백 플레인에 점보 프레임을 사용하는 것이 좋습니다. Red Hat Ceph Storage는 공용 네트워크와 클러스터 네트워크 모두에 대해 통신 경로의 모든 네트워킹 장치에 걸쳐 동일한 MTU 값이 필요합니다. 프로덕션에서 Red Hat Ceph Storage 클러스터를 사용하기 전에 환경의 모든 호스트 및 네트워킹 장치에서 MTU 값이 동일한지 확인합니다.

2.6. RAID 솔루션 사용 방지

Ceph는 코드 오브젝트를 복제하거나 삭제할 수 있습니다. RAID는 이 기능을 블록 수준에서 복제하고 사용 가능한 용량을 줄입니다. 따라서 RAID는 불필요한 비용이 듭니다. 또한 성능 저하된 RAID는 성능에 **부정적인** 영향을 미칩니다.

중요

Red Hat은 각 하드 드라이브를 나중 쓰기 캐싱이 활성화된 단일 볼륨으로 RAID 컨트롤러와 별도로 내보내는 것이 좋습니다.

이를 위해서는 배터리 지원 또는 스토리지 컨트롤러에서 비발성 플래시 메모리 장치가 필요합니다. 전원 실패로 인해 컨트롤러의 메모리를 유실할 수 있는 경우 대부분의 컨트롤러에서 나중 쓰기 캐싱을 비활성화하므로 배터리가 작동하는지 확인하는 것이 중요합니다. 시간이 지남에 따라 성능이 저하되므로 필요한 경우 정기적으로 점검하고 교체합니다. 자세한 내용은 스토리지 컨트롤러 공급업체 설명서를 참조하십시오. 일반적으로 스토리지 컨트롤러 벤더는 다운타임 없이 스토리지 컨트롤러 구성을 모니터링하고 조정할 수 있는 스토리지 관리 유틸리티를 제공합니다.

모든 SSD(Solid State Drives)를 사용하거나 컨트롤러당 드라이브 수가 많은 구성을 위해 Ceph와 함께 독립적인 드라이브 모드에서 JBDS(JBch of Drives)를 사용하는 것이 지원됩니다. 예를 들어 하나의 컨트롤러에 60개의 드라이브가 연결되어 있습니다. 이 시나리오에서는 나중 쓰기 캐싱이 I/O 경합의 소스가 될 수 있습니다. JBOD는 나중 쓰기 캐싱을 비활성화하므로 이 시나리오에서 이상적입니다. JBOD 모드를 사용할 경우의 한 가지 장점은 드라이브를 추가 또는 교체한 다음 물리적으로 연결된 직후 운영 체제에 드라이브를 노출하는 것입니다.

2.7. 하드웨어 선택 시 일반적인 실수 요약

- Ceph와 함께 사용할 수 있도록 강력한 레거시 하드웨어의 목적 변경.
- 동일한 풀에서 다른 하드웨어 사용.
- 10Gbps 이상의 네트워크 대신 1Gbps 네트워크 사용.
- 공용 및 클러스터 네트워크 설정을 무시합니다.

- JBOD 대신 RAID 사용.
- 성능 또는 처리량에 관계 없이 가격 기준으로 드라이브를 선택합니다.
- OSD 데이터의 저널링은 SSD 저널에 대한 사용 사례를 호출할 때 구동됩니다.
- 처리량 특성이 충분하지 않은 디스크 컨트롤러 보유.

이 문서에는 여러 워크로드에 대한 Red Hat 테스트 구성의 예제를 사용하여 심각한 하드웨어 선택 실수가 발생하지 않습니다.

2.8. 추가 리소스

- Red Hat 고객 포털에서 지원되는 구성 문서.

3장. 워크로드 성능 도메인 최적화

Ceph 스토리지의 주요 이점 중 하나는 Ceph 성능 도메인을 사용하여 동일한 클러스터 내에서 다양한 유형의 워크로드를 지원할 수 있다는 것입니다. 각 성능 도메인과 매우 다른 하드웨어 구성을 연결할 수 있습니다. Ceph 시스템 관리자는 적절한 성능 도메인에 스토리지 풀을 배포할 수 있으며 애플리케이션에 특정 성능 및 비용 프로필에 맞는 스토리지를 제공할 수 있습니다. 이러한 성능 도메인에 맞게 크기가 조정되고 최적화된 서버를 선택하는 것은 Red Hat Ceph Storage 클러스터를 설계하는 데 있어 필수적인 부분입니다.

다음 목록에서는 Red Hat이 스토리지 서버에서 최적의 Red Hat Ceph Storage 클러스터 구성을 식별하는 데 사용하는 기준을 제시합니다. 이러한 범주는 하드웨어 구매 및 구성 결정에 대한 일반적인 지침으로 제공되며, 고유한 워크로드 혼합을 충족하도록 조정할 수 있습니다. 선택한 실제 하드웨어 구성은 특정 워크로드 혼합 및 벤더 기능에 따라 달라집니다.

최적화된 IOPS

IOPS 최적화된 스토리지 클러스터에는 일반적으로 다음 속성이 있습니다.

- IOPS당 최저 비용.
- GB당 가장 높은 IOPS.
- 99번째 백분위 대기 시간 일관성.

일반적으로 IOPS 최적화된 스토리지 클러스터에는 다음이 사용됩니다.

- 일반적으로 블록 스토리지.
- SSD(반도체 드라이브)의 경우 하드 디스크 드라이브(HDD) 또는 2x 복제를 위한 3x 복제.
- OpenStack 클라우드의 MySQL.

처리량 최적화

처리량 최적화된 스토리지 클러스터에는 일반적으로 다음과 같은 속성이 있습니다.

- MBps당 최저 비용(처리량).
- TB당 최고 MBps.
- BTU당 가장 높은 MBps.
- 푸트당 최고 MBps.
- 97번째 백분위 대기 시간 일관성.

일반적으로 처리량 최적화된 스토리지 클러스터에는 다음이 사용됩니다.

- 블록 또는 개체 스토리지.
- 3배 복제.
- 비디오, 오디오 및 이미지를 위한 액티브 성능 스토리지.
- 스트리밍 미디어.

비용 및 용량 최적화

비용 및 용량 최적화된 스토리지 클러스터에는 일반적으로 다음과 같은 속성이 있습니다.

- TB당 최저 비용.
- TB당 최저 BTU.
- TB당 최저 발트 필요.

일반적으로 비용 및 용량에 최적화된 스토리지 클러스터는 다음과 같습니다.

- 일반적으로 오브젝트 스토리지.
- 사용 가능한 용량을 극대화하기 위한 일반적인 코딩 삭제
- 개체 아카이브.
- 비디오, 오디오 및 이미지 오브젝트 리포지토리.

성능 도메인 작동 방식

데이터를 읽고 쓰는 Ceph 클라이언트 인터페이스의 경우 Ceph 스토리지 클러스터는 클라이언트가 데이터를 저장하는 간단한 풀로 나타납니다. 그러나 스토리지 클러스터는 클라이언트 인터페이스에 완전히 투명한 방식으로 많은 복잡한 작업을 수행합니다. Ceph 클라이언트 및 Ceph 개체 스토리지 데몬(Ceph OSD 또는 단순히 OSD) 모두 개체 스토리지 및 검색에 대해 CRUSH(Replication) 알고리즘을 사용합니다. OSD는 클러스터 내의 스토리지 서버인 OSD 호스트에서 실행됩니다.

CRUSH 맵은 클러스터 리소스의 토폴로지를 설명하고 이 맵은 클라이언트 노드와 클러스터 내에 Ceph 모니터(MON) 노드에 있습니다. Ceph 클라이언트 및 Ceph OSD는 모두 CRUSH 맵과 CRUSH 알고리즘을 사용합니다. Ceph 클라이언트는 OSD와 직접 통신하여 중앙 집중식 오브젝트 조회 및 잠재적인 성능 병목 현상이 발생하지 않습니다. CRUSH 맵과 피어와의 통신을 인식하여 OSD는 복제, 백필링 및 복구를 처리할 수 있으므로 동적 오류 복구가 가능합니다.

Ceph는 CRUSH 맵을 사용하여 장애 도메인을 구현합니다. 또한 Ceph는 CRUSH 맵을 사용하여 성능 도메인을 구현합니다. 이 도메인은 단순히 기본 하드웨어의 성능 프로파일을 고려하여 고려합니다. CRUSH 맵은 Ceph에서 데이터를 저장하는 방법을 설명하고, 단순한 계층 구조(시크립 그래프) 및 규칙 세트로 구현됩니다. CRUSH 맵은 여러 계층 구조를 지원하여 한 가지 유형의 하드웨어 성능 프로필을 분리할 수 있습니다.

다음 예제에서는 성능 도메인을 설명합니다.

- 일반적으로 HDD(하드 디스크 드라이브)는 비용 및 용량 중심 워크로드에 적합합니다.
- 처리량에 민감한 워크로드는 일반적으로 SSD(반도체 드라이브)에서 Ceph 쓰기 저널과 함께 HDD를 사용합니다.
- MySQL 및 MariaDB와 같은 IOPS 집약적인 워크로드에서는 SSD를 사용하는 경우가 많습니다.

이러한 모든 성능 도메인은 Ceph 스토리지 클러스터에 공존할 수 있습니다.

4장. 서버 및 랙 솔루션

하드웨어 벤더는 최적화된 서버 수준 및 랙 수준 솔루션 SKU를 모두 제공하여 Ceph의 열정에 부응했습니다. Red Hat과의 공동 테스트를 통해 검증된 이 솔루션은 Ceph 배포에 대해 예측 가능한 가격 대비 성능 비율을 제공하며, 특정 워크로드에 맞게 Ceph 스토리지를 확장할 수 있는 편리한 모듈식 접근 방식을 제공합니다.

일반적인 랙 레벨 솔루션에는 다음이 포함됩니다.

- **네트워크 스위칭:** 중복 네트워크 스위칭은 클러스터를 상호 연결하고 클라이언트에 대한 액세스를 제공합니다.
- **Ceph MON 노드:** Ceph 모니터는 전체 클러스터 상태에 대한 데이터 저장소이며 클러스터 로그를 포함합니다. 프로덕션 환경에서는 클러스터 쿼럼에 최소 3개의 모니터 노드를 사용하는 것이 좋습니다.
- **Ceph OSD 호스트:** Ceph OSD 호스트는 개별 스토리지 장치별로 하나 이상의 OSD를 실행하는 클러스터의 스토리지 용량을 수용합니다. OSD 호스트는 워크로드 최적화와 설치된 데이터 장치에 따라 다르게 선택 및 구성됩니다. HDD, SSD 또는 NVMe SSD.
- **Red Hat Ceph Storage:** 많은 벤더가 서버 및 랙 레벨 솔루션 SKU와 함께 제공되는 Red Hat Ceph Storage용 용량 기반 서브스크립션을 제공합니다.

참고

Red Hat은 서버 및 랙 솔루션에 커밋하기 전에 *Red Hat Ceph Storage: 지원 구성* 문서를 검토할 것을 권장합니다. 추가 *지원이 필요한 경우 Red Hat* 지원팀에 문의하십시오.

IOPS 최적화된 솔루션

플래시 스토리지 사용이 증가함에 따라 조직은 Ceph 스토리지 클러스터에서 IOPS 집약적인 워크로드를 호스팅하여 사설 클라우드 스토리지로 고성능 퍼블릭 클라우드 솔루션을 에뮬레이션할 수 있습니다. 이러한 워크로드에는 일반적으로 MySQL, MariaDB 또는 PostgreSQL 기반 애플리케이션의 구조화된 데이터가 포함됩니다.

일반적인 서버에는 다음 요소가 포함됩니다.

- **CPU:** NVMe SSD당 10개의 코어로, 2GHz CPU를 가정합니다.
- **RAM:** 16GB 기준선, OSD당 5GB 추가.
- **네트워킹:** 2개의 OSD당 GbE(Gigabit Ethernet) 10개.
- **OSD 미디어:** 고성능의 고성능 엔터프라이즈 NVMe SSD.
- **OSDs:** NVMe SSD당 두 개.
- **bluestore WAL/DB:** OSD와 공동 배치된 고성능의 고성능 엔터프라이즈 NVMe SSD.
- **컨트롤러:** 기본 PCIe 버스.

참고

NVMe SSD의 경우 **CPU** 의 경우 SSD OSD당 두 개의 코어를 사용합니다.

표 4.1. 클러스터 크기별 IOPS 최적화된 Ceph 워크로드용 솔루션 SKU.

벤더	작음(250TB)	중간(1PB)	대규모 (2PB 이상)
SuperMicro [a]	SYS-5038MR-OSD006P	해당 없음	해당 없음
[a] 자세한 내용은 Supermicro® Total Solution for Ceph 를 참조하십시오.			

처리량 최적화 솔루션

처리량에 최적화된 Ceph 솔루션은 일반적으로 반정형 또는 구조화되지 않은 데이터를 중심으로 합니다. 대규모 블록 순차 I/O는 일반적입니다.

일반적인 서버 요소는 다음과 같습니다.

- **CPU:** HDD당 0.5코어, 2GHz CPU를 가정합니다.
- **RAM:** 16GB 기준선, OSD당 5GB 추가.
- **네트워킹:** 클라이언트 및 클러스터 연결 네트워크에 대해 각각 12개의 OSD당 10GbE.
- **OSD 미디어:** 7,200 RPM 엔터프라이즈 HDD.
- **OSDs:** HDD당 하나씩.
- **bluestore WAL/DB:** OSD와 공동 배치된 고성능의 고성능 엔터프라이즈 NVMe SSD.
- **HBA(호스트 버스 어댑터):** 작은 디스크(JBOD)만으로도 가능합니다.

일부 벤더는 처리량에 최적화된 Ceph 워크로드를 위해 사전 구성된 서버 및 랙 수준 솔루션을 제공합니다. Red Hat은 Supermicro 및 Quanta Cloud Technologies(T)를 통해 서버의 광범위한 테스트 및 평가를 수행했습니다.

표 4.2. Ceph OSD, MON, TOR(최상급) 스위치용 랙 레벨 SKU.

벤더	작음(250TB)	중간(1PB)	대규모 (2PB 이상)
SuperMicro [a]	SRS-42E112-Ceph-03	SRS-42E136-Ceph-03	SRS-42E136-Ceph-03

표 4.3. 개별 OSD 서버

벤더	작음(250TB)	중간(1PB)	대규모 (2PB 이상)
SuperMicro [a]	SSG-6028R-OSD072P	SSG-6048-OSD216P	SSG-6048-OSD216P
QCT [a]	QxStor RCT-200	QxStor RCT-400	QxStor RCT-400
[a] 를 참조하십시오. 자세한 내용은 QxStor Red Hat Ceph Storage Edition 에서 참조하십시오.			

표 4.4. 처리량 최적화 Ceph OSD 워크로드의 경우 추가 서버 구성 가능.

벤더	작음(250TB)	중간(1PB)	대규모 (2PB 이상)
Dell	PowerEdge R730XD [a]	DSS 7000 [b], 노드	DSS 7000, jboss node
Cisco	UCS C240 M4	UCS C3260 [c]	UCS C3260 [d]
Lenovo	시스템 x3650 M5	시스템 x3650 M5	해당 없음

[a] 자세한 내용은 Dell PowerEdge R730xd 성능 및 크기 조정 가이드를 참조하십시오.

[b] 자세한 내용은 Dell EMC DSS 7000 성능 및 크기 조정 가이드를 참조하십시오.

[c] 자세한 내용은 Red Hat Ceph Storage 하드웨어 참조 아키텍처를 참조하십시오.

[d] 자세한 내용은 UCS C3260 을 참조하십시오.

비용 및 용량에 최적화된 솔루션

비용 및 용량 최적화된 솔루션은 일반적으로 높은 용량 또는 긴 아카이브 시나리오에 중점을 둡니다. 데이터는 반 구조화 또는 비정형일 수 있습니다. 워크로드에는 미디어 아카이브, 빅 데이터 분석 아카이브 및 머신 이미지 백업이 포함됩니다. 대규모 블록 순차 I/O는 일반적입니다.

일반적으로 솔루션에는 다음 요소가 포함됩니다.

- **CPU.** HDD당 0.5코어, 2GHz CPU를 가정합니다.
- **RAM.** 16GB 기준선, OSD당 5GB 추가.
- **네트워킹.** 12개의 OSD당 10GbE(클라이언트 및 클러스터 연결 네트워크용).
- **OSD 미디어.** 7,200 RPM 엔터프라이즈 HDD.
- **OSD.** HDD당 하나씩.
- **bluestore WAL/DB** Co- located on theHD.
- **HBA.** JBOD.

Supermicro 및 QCT는 비용 및 용량에 중점을 둔 Ceph 워크로드를 위한 사전 구성된 서버 및 랙 수준 솔루션 SKU를 제공합니다.

표 4.5. 비용 및 용량에 최적화된 워크로드에 맞게 사전 구성된 랙 레벨 SKU

벤더	작음(250TB)	중간(1PB)	대규모 (2PB 이상)
SuperMicro [a]	해당 없음	SRS-42E136-Ceph-03	SRS-42E172-Ceph-03

표 4.6. 비용 및 용량 최적화된 워크로드에 맞게 사전 구성된 서버 수준 SKU

벤더	작음(250TB)	중간(1PB)	대규모 (2PB 이상)
SuperMicro [a]	해당 없음	SSG-6048R-OSD216P [a]	SSD-6048R-OSD360P
QCT	해당 없음	QxStor RCC-400 [a]	QxStor RCC-400 [a]
[a] 자세한 내용은 Supermicro의 Ceph 총 솔루션을 참조하십시오.			

표 4.7. 비용 및 용량에 최적화된 워크로드에 대해 구성 가능한 추가 서버

벤더	작음(250TB)	중간(1PB)	대규모 (2PB 이상)
Dell	해당 없음	DSS 7000, jboss node	DSS 7000, jboss node
Cisco	해당 없음	UCS C3260	UCS C3260
Lenovo	해당 없음	시스템 x3650 M5	해당 없음

추가 리소스

- *삼성 NVMe SSD의 Red Hat Ceph Storage*
- *Red Hat Ceph Storage on the Infini All-¢ Storage System (Infinor All-Storage System)*
- *Red Hat Ceph Storage에 MySQL 데이터베이스 배포*
- *Intel® Data Center Blocks for Cloud - Red Hat OpenStack Platform with Red Hat Ceph Storage*
- *QCT 서버의 Red Hat Ceph Storage*
- *Intel 프로세서 및 SSD가 있는 서버의 Red Hat Ceph Storage*

5장. 컨테이너화된 CEPH의 최소 하드웨어 권장 사항

Ceph는 비독점 상용 하드웨어에서 실행할 수 있습니다. 소규모 프로덕션 클러스터 및 개발 클러스터는 보통 하드웨어를 사용하여 성능 최적화 없이 실행될 수 있습니다.

프로세스	기준	최소 권장 사항
ceph-osd-container	프로세서	OSD 컨테이너당 1x AMD64 또는 Intel 64 CPU CORE
	RAM	OSD 컨테이너당 최소 5GB의 RAM
	OS 디스크	호스트당 1x OS 디스크
	OSD 저장소	OSD 컨테이너당 1배 스토리지 드라이브. OS 디스크와 공유할 수 없습니다.
	block.db	선택 사항이지만 Red Hat은 데몬당 1x SSD 또는 NVMe 또는 Optane 파티션 또는 lvm을 권장합니다. 크기 조정은 **Blue** Store의 개체, 파일 및 혼합 워크로드를 위한 **block.data**, 블록 장치용 BlueStore, Openstack cinder 및 Openstack cinder 워크로드의 경우 block.data의 4%입니다.
	block.wal	데몬당 1x SSD 또는 NVMe 또는 Optane 파티션 또는 논리 볼륨(옵션). 크기가 작은 크기(예: 10GB)를 사용하고 **block.db** 장치보다 빠른 경우에만 사용합니다.
	네트워크	10GB 이더넷 NIC 2개
ceph-mon-container	프로세서	mon-container당 1x AMD64 또는 Intel 64 CPU CORE
	RAM	**mon-container**당 3GB
	디스크 공간	**mon-container** 당 10GB, 권장 50GB
	디스크 모니터링	필요한 경우 **Monitor rocksdb** 데이터용 1x SSD 디스크
	네트워크	1GB 이더넷 NIC 2개, 권장 10GB
ceph-mgr-container	프로세서	**mgr-container**당 1x AMD64 또는 Intel 64 CPU CORE
	RAM	**mgr-container**당 3GB
	네트워크	1GB 이더넷 NIC 2개, 권장 10GB
ceph-radosgw-container	프로세서	radosgw-container당 1x AMD64 또는 Intel 64 CPU CORE
	RAM	데몬당 1GB

프로세스	기준	최소 권장 사항
	디스크 공간	데몬당 5GB
	네트워크	1x 1GB 이더넷 NIC
ceph-mds-container	프로세서	mds-container당 1x AMD64 또는 Intel 64 CPU CORE
	RAM	**mds-container**당 3GB 이 숫자는 구성 가능한 MDS 캐시 크기에 따라 크게 달라집니다. RAM 요구 사항은 일반적으로 **mds_cache_memory_limit** 구성 설정에 설정된 양보다 두 배입니다. 또한 이는 전체 시스템 메모리가 아닌 데몬의 메모리입니다.
	디스크 공간	**mds-container** 당 2GB를 추가하고 가능한 디버그 로깅에 필요한 추가 공간을 고려하면 20GB가 좋습니다.
	네트워크	1GB 이더넷 NIC 2개, 권장 10GB 이 네트워크는 OSD 컨테이너와 동일한 네트워크입니다. OSD에 10GB 네트워크가 있는 경우 대기 시간과 관련했을 때 MDS에 동일한 네트워크를 사용해야 합니다.

6장. RED HAT CEPH STORAGE DASHBOARD에 권장되는 최소 하드웨어 요구 사항

Red Hat Ceph Storage 대시보드에는 최소 하드웨어 요구 사항이 있습니다.

최소 요구 사항

- 2.5GHz 이상의 4개 코어 프로세서
- 8GB RAM
- 50GB 하드 디스크 드라이브
- 1 기가비트 이더넷 네트워크 인터페이스

추가 리소스

- 자세한 내용은 관리 가이드의 *Ceph 스토리지 클러스터의 고급 모니터링* 을 참조하십시오.