# Red Hat Ceph Storage 5

## ハードウェアガイド

Red Hat Ceph Storage におけるハードウェア選択に関する推奨事項

Last Updated: 2022-05-21

# Red Hat Ceph Storage 5 ハードウェアガイド

Red Hat Ceph Storage におけるハードウェア選択に関する推奨事項

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.

## 法律上の通知



## 概要

本ガイドでは、Red Hat Ceph Storage で使用するハードウェアを選択する際の高度なガイダンスを提供します。 Red Hat では、コード、ドキュメント、Web プロパティーにおける配慮に欠ける用語の置き換えに取り組んでいます。まずは、マスター (master)、スレーブ (slave)、ブラックリスト (blacklist)、ホワイトリスト (whitelist) の 4 つの用語の置き換えから始めます。この取り組みは膨大な作業を要するため、今後の複数のリリースで段階的に用語の置き換えを実施して参ります。詳細は、弊社の CTO、Chris Wright のメッセージを参照してください。

# 目次

**第1章 エグゼクティブサマリー** ..... 3

**第2章 ハードウェアを選択する一般的な原則** ..... 5
2.1. 前提条件 5
2.2. パフォーマンスユースケースの特定 5
2.3. ストレージの密度を考慮する 5
2.4. 同じハードウェア構成 5
2.5. RED HAT CEPH STORAGE のネットワークに関する考察 6
2.6. RAID ソリューションの使用を回避 7
2.7. ハードウェアを選択する際のよくある間違いの概要 8
2.8. 関連情報 8

**第3章 ワークロードのパフォーマンスドメインの最適化** ..... 9

**第4章 サーバーおよびラックソリューション** ..... 11

**第5章 コンテナー化された CEPH のハードウェアの最小推奨事項** ..... 15

**第6章 RED HAT CEPH STORAGE DASHBOARD の推奨される最小ハードウェア要件** ..... 17

# 第1章 エグゼクティブサマリー

多くのハードウェアベンダーは、個別のワークロードプロファイル向けに設計された Ceph 最適化サーバーおよびラックレベルのソリューションの両方を提供するようになりました。ハードウェア選択プロセスを単純化し、組織のリスクを低減するために、Red Hat は複数のストレージサーバーベンダーと連携して、さまざまなクラスターサイズおよびワークロードプロファイルの特定のクラスターオプションをテストおよび評価しています。Red Hat の厳密な方法論は、パフォーマンステストと、幅広いクラスター機能とサイズの確立されたガイダンスを組み合わせたものです。適切なストレージサーバーとラックレベルのソリューションを使用することで、Red Hat Ceph Storage は、スループットに敏感でコストおよび容量を重視するワークロードから、新しい IOPS 集約型ワークロードまで、さまざまなワークロードに対応するストレージプールを提供できます。

Red Hat Ceph Storage は、エンタープライズデータの保存コストを大幅に削減し、組織が指数関数的なデータの成長を管理できるようにします。このソフトウェアは、パブリックまたはプライベートのクラウドデプロイメント向けの堅牢かつ最新のペタバイトスケールのストレージプラットフォームです。Red Hat Ceph Storage は、エンタープライズブロックおよびオブジェクトストレージ用の成熟したインターフェースを提供し、テナントに依存しない OpenStack® 環境によってアクティブなアーカイブ、リッチメディア、およびクラウドインフラストラクチャーのワークロードに最適なソリューションを提供します。[1].統合されたソフトウェア定義のスケールアウトストレージプラットフォームとして提供される Red Hat Ceph Storage は、以下のような機能を提供することで、企業がアプリケーションの革新性と可用性の向上に集中できるようにします。

- 数百ペタバイトへのスケーリング [2].
- クラスターに単一障害点はありません。
- 商用サーバーハードウェア上で実行することで、資本経費 (CapEx) を削減します。
- 自己管理および自己修復プロパティーで運用費 (OpEx) を削減します。

Red Hat Ceph Storage は、多様なニーズを満たすために、業界標準のハードウェア構成で動作することができます。クラスター設計プロセスを単純化および加速するために、Red Hat は、参加するハードウェアベンダーによるパフォーマンスおよび適合性のテストを実施しています。このテストにより、選択したハードウェアを負荷下で評価し、多様なワークロードに必要な性能とサイジングデータを生成することができ、Ceph ストレージクラスタのハードウェア選択を大幅に簡素化できます。本ガイドで説明しているように、複数のハードウェアベンダーが Red Hat Ceph Storage デプロイメントに最適化されたサーバーおよびラックレベルのソリューションを提供しており、IOPS、スループット、コスト、容量を最適化したソリューションが利用可能なオプションとして提供されています。

ソフトウェア定義ストレージは、要求の厳しいアプリケーションや段階的に増大するストレージのニーズを満たすスケールアウトソリューションを求める組織にとって、多くの利点を提供しています。複数のベンダーで実施される優れた方法論と広範囲のテストにより、Red Hat は、あらゆる環境の需要を満たすためにハードウェアを選択するプロセスを単純化します。重要なことは、本書に記載されているガイドラインやシステム例は、サンプルシステムにおける実稼働環境のワークロードの影響を定量化するための代用品ではないということです。

Red Hat Ceph Storage を実行するためのサーバーの設定に関する情報は、『Red Hat Ceph Storage Hardware Configuration Guide』に記載の方法論およびベストプラクティスを参照してください。Red Hat Ceph Storage テスト結果などの詳細情報は、一般的なハードウェアベンダーのパフォーマンスおよびサイジングに関するガイドを参照してください。

---

[1] 年 2 回の OpenStack ユーザー調査によると、Ceph は OpenStack をリードするストレージであり、その地位を確立しています。

[2] 詳細は、「Yahoo Cloud Object Store - Object Storage at Exabyte Scale」を参照してください。

# 第2章 ハードウェアを選択する一般的な原則

ストレージ管理者は、実稼働用の Red Hat Ceph Storage クラスターを実行するのに適切なハードウェアを選択する必要があります。Red Hat Ceph Storage のハードウェアを選択する際には、以下の一般原則を確認してください。この原則は、時間を節約し、よくある間違いを回避し、お金を節約し、より効果的な解決策を実現するのに役立ちます。

## 2.1. 前提条件

- Red Hat Ceph Storage の計画的な使用

## 2.2. パフォーマンスユースケースの特定

Ceph の導入を成功させるための最も重要なステップの 1 つは、クラスターのユースケースとワークロードに適した価格対性能プロファイルを特定することです。ユースケースに適したハードウェアを選択することが重要です。例えば、コールドストレージアプリケーション用に IOPS が最適化されたハードウェアを選択すると、ハードウェアのコストが必要以上に増加します。一方で、IOPS を多用するワークロードにおいて、より魅力的な価格帯のために容量を最適化したハードウェアを選択すると、パフォーマンスの低下に不満を持つユーザーが出てくる可能性が高くなります。

Ceph の主なユースケースは以下のとおりです。

- **最適化された IOPS:** IOPS が最適化されたデプロイメントは、MYSQL や MariaDB インスタンスを OpenStack 上の仮想マシンとして実行するなど、クラウドコンピューティングの操作に適しています。IOPS が最適化された導入では、15k RPM の SAS ドライブや、頻繁な書き込み操作を処理するための個別の SSD ジャーナルなど、より高性能なストレージが必要となります。一部の IOPS のシナリオでは、すべてのフラッシュストレージを使用して IOPS と総スループットが向上します。

- **最適化されたスループット:** スループットが最適化されたデプロイメントは、グラフィック、音声、ビデオコンテンツなどの大量のデータを提供するのに適しています。スループット最適化されたデプロイメントには、トータルスループット特性が許容されるネットワーキングハードウェア、コントローラー、ハードディスクドライブが必要です。書き込みパフォーマンスが必須である場合、SSD ジャーナルは書き込みパフォーマンスを大幅に向上します。

- **最適化された容量:** 容量が最適化されたデプロイメントは、大量のデータを可能な限り安価に保存するのに適しています。容量が最適化されたデプロイメントは通常、パフォーマンスがより魅力的な価格と引き換えになります。たとえば、容量を最適化したデプロイメントでは、ジャーナリングに SSD を使用するのではなく、より低速で安価な SATA ドライブを使用し、ジャーナルを同じ場所に配置することがよくあります。

本書は、これらのユースケースに適した Red Hat テスト済みハードウェアの例を提供します。

## 2.3. ストレージの密度を考慮する

ハードウェアのプランニングには、ハードウェア障害が発生した場合に高可用性を維持するために、Ceph デーモンや Ceph を使用する他のプロセスを多数のホストに分散させることが含まれていなければなりません。ハードウェア障害が発生した場合のクラスターのリバランスの必要性とストレージ密度のバランスを考慮してください。よくあるハードウェアの選択ミスは、小規模なクラスターで非常に高いストレージ密度を使用することで、バックフィルやリカバリー操作中にネットワークに負荷がかかりすぎる可能性があります。

## 2.4. 同じハードウェア構成

プールを作成し、プール内の OSD ハードウェアが同じになるように CRUSH 階層を定義します。

- 同じコントローラー。
- 同じドライブのサイズ。
- 同じ RPM。
- 同じシーク時間。
- 同じ I/O。
- 同じネットワークスループット。
- 同じジャーナル設定。

プール内で同じハードウェアを使用することで、一貫したパフォーマンスプロファイルが得られ、プロビジョニングが簡素化され、トラブルシューティングの効率が上がります。

**警告**

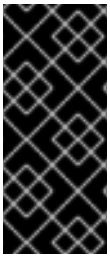
複数のストレージデバイスを使用する場合 (再起動時に場合によってはデバイスの順序が変わることがあります)。この問題のトラブルシューティングは、「How do I change the order of storage devices during boot in RHEL 7?」を参照してください。

## 2.5. RED HAT CEPH STORAGE のネットワークに関する考察

クラウドストレージソリューションの重要な点は、ネットワークのレイテンシーなどの要因により、ストレージクラスターが IOPS 不足になることです。また、ストレージクラスターがストレージ容量を使い果たす、はるか前に、帯域幅の制約が原因でスループットが不足することがあります。つまり、価格対性能の要求を満たすには、ネットワークのハードウェア構成が選択されたワークロードをサポートする必要があります。

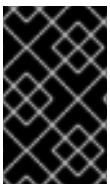
ストレージ管理者は、ストレージクラスターをできるだけ早く復旧することを望みます。ストレージクラスターネットワークの帯域幅要件を慎重に検討し、ネットワークリンクのオーバーサブスクリプションに注意してください。また、クライアント間のトラフィックからクラスター内のトラフィックを分離します。また、SSD (Solid State Disk) やフラッシュ、NVMe などの高性能なストレージデバイスの使用を検討する場合には、ネットワークパフォーマンスの重要性が増していることも考慮してください。

Ceph はパブリックネットワークとストレージクラスターネットワークをサポートしています。パブリックネットワークは、クライアントのトラフィックと Ceph Monitor との通信を処理します。ストレージクラスターネットワークは、Ceph OSD のハートビート、レプリケーション、バックフィル、リカバリーのトラフィックを処理します。ストレージハードウェアには、**最低でも** 10GB のイーサネットリンクを 1 つ使用し、接続性とスループット向けにさらに 10GB イーサネットリンクを追加できます。



**重要**

Red Hat では、レプリケートされたプールをもとに **osd_pool_default_size** を使用してパブリックネットワークの倍数となるように、ストレージクラスターネットワークに帯域幅を割り当てることを推奨しています。また、Red Hat はパブリックネットワークとストレージクラスターネットワークを別々のネットワークカードで実行することを推奨しています。



**重要**

Red Hat では、実稼働環境での Red Hat Ceph Storage のデプロイメントに 10GB のイーサネットを使用することを推奨しています。1GB のイーサネットネットワークは、実稼働環境のストレージクラスターには適していません。

ドライブに障害が発生した場合、1 GB イーサネットネットワーク全体で 1 TB のデータをレプリケートするには 3 時間かかります。3 TB には 9 時間かかります。3TB を使用するのが一般的なドライブ構成です。一方、10GB のイーサネットネットワークの場合、レプリケーションにかかる時間はそれぞれ 20 分、1 時間となります。Ceph OSD に障害が発生した場合には、ストレージクラスターは、含まれるデータをプール内の他の Ceph OSD にレプリケートして復元することに注意してください。

ラックなどの大規模なドメインに障害が発生した場合は、ストレージクラスターが帯域幅を大幅に消費することになります。複数のラックで構成されるストレージクラスター (大規模なストレージ実装では一般的) を構築する際には、最適なパフォーマンスを得るために、「ファットツリー」設計でスイッチ間のネットワーク帯域幅をできるだけ多く利用することを検討してください。一般的な 10 GB のイーサネットスイッチには、48 個の 10 GB ポートと 4 個の 40 GB のポートがあります。スループットを最大にするには、Spine (背骨) で 40 GB ポートを使用します。または、QSFP+ および SFP+ ケーブルを使用する未使用の 10 GB ポートを別のラックおよびスパインルーターに接続するために、さらに 40 GB のポートに集計することを検討します。また、LACP モード 4でネットワークインターフェースを結合することも検討してください。また、特にバックエンドやクラスターのネットワークでは、ジャンボフレーム、最大伝送単位 (MTU) 9000を使用してください。

Red Hat Ceph Storage クラスタをインストールしてテストする前に、ネットワークのスループットを確認します。Ceph のパフォーマンスに関する問題のほとんどは、ネットワークの問題から始まります。Cat-6 ケーブルのねじれや曲がりといった単純なネットワークの問題は、帯域幅の低下につながります。フロント側のネットワークには、最低でも10 GB のイーサネットを使用してください。大規模なクラスターの場合には、バックエンドやクラスターのネットワークに 40GB のイーサネットを使用することを検討してください。

**重要**

ネットワークの最適化には、CPU/帯域幅の比率を高めるためにジャンボフレームを使用し、非ブロックのネットワークスイッチのバックプレーンを使用することを Red Hat は推奨します。Red Hat Ceph Storageでは、パブリックネットワークとクラスターネットワークの両方で、通信パスにあるすべてのネットワークデバイスに同じ MTU 値がエンドツーエンドで必要となります。Red Hat Ceph Storage クラスターを実稼働環境で使用する前に、環境内のすべてのノードとネットワーク機器で MTU 値が同じであることを確認します。

## 2.6. RAID ソリューションの使用を回避

Ceph はコードオブジェクトの複製または消去が可能です。RAID は、この機能をブロックレベルで複製し、利用可能な容量を減らします。そのため、RAID は不要な費用です。さらに、劣化した RAID はパフォーマンスに **悪影響** を及ぼします。

**重要**

Red Hat では、各ハードドライブを RAID コントローラーから個別にエクスポートし、ライトバックキャッシングを有効にして 1 つのボリュームとして使用することを推奨します。

これには、ストレージコントローラー上にバッテリーが支援された、または不揮発性のフラッシュメモリーデバイスが必要です。停電が原因で、コントローラー上のメモリーが失われる可能性がある場合は、ほとんどのコントローラーがライトバックキャッシングを無効にするため、バッテリーが動作していることを確認することが重要です。電池は経年劣化するので、定期的に点検し、必要に応じて交換してください。詳細は、ストレージコントローラーベンダーのドキュメントを参照してください。通常、ストレージコントローラーベンダーは、ダウンタイムなしにストレージコントローラー設定を監視および調整するストレージ管理ユーティリティーを提供します。

Ceph で独立したドライブモードでの JBOD (Just a Bunch of Drives) の使用は、すべてのソリッドステートドライブ (SSD) を使用している場合や、コントローラーあたりのドライブ数が多い構成の場合にサポートされています。たとえば、1 つのコントローラーに接続されたドライブ数が 60 などの場合です。このシナリオでは、ライトバックキャッシングが I/O 競合のソースとなります。JBOD はライトバックキャッシュを無効にするため、このシナリオには適しています。JBOD モードを使用する利点の 1 つは、ドライブの追加や交換が簡単で、物理的に接続した後すぐにオペレーティングシステムにドライブを公開できることです。

## 2.7. ハードウェアを選択する際のよくある間違いの概要

- パワー不足のレガシーハードウェアを Ceph で使用するために再利用する。
- 同じプールで異種のハードウェアを使用する。
- 10Gbps 以上ではなく 1Gbps ネットワークを使用する。
- パブリックネットワークとクラスターネットワークの両方を設定することを怠っている。
- JBOD の代わりに RAID を使用する。
- パフォーマンスやスループットを考慮せずに、価格順でドライブを選択する。
- SSD ジャーナルのユースケース呼び出し時の OSD データドライブでジャーナリングを行う。
- スループットの特性が不十分なディスクコントローラーがある。

本書に記載されている Red Hat の異なるワークロード用のテスト済み構成の例を使用して、前述のハードウェアの選択ミスを回避してください。

## 2.8. 関連情報

- Red Hat カスタマーポータルでサポートされる構成に関するアーティクル 「Red Hat Ceph Storage でサポートされる構成」

# 第3章 ワークロードのパフォーマンスドメインの最適化

Ceph Storage の主な利点の 1 つとして、Ceph パフォーマンスドメインを使用して、同じクラスター内のさまざまなタイプのワークロードをサポートする機能があります。劇的に異なるハードウェア構成を各パフォーマンスドメインに関連付けることができます。Ceph システム管理者は、ストレージプールを適切なパフォーマンスドメインにデプロイし、特定のパフォーマンスおよびコストプロファイルに合わせたストレージでアプリケーションを提供できます。これらのパフォーマンスドメインに適切なサイズ設定と最適化されたサーバーを選択することは、Red Hat Ceph Storage クラスターを設計するのに不可欠な要素です。

以下の一覧は、ストレージサーバーで最適な Red Hat Ceph Storage クラスター設定の特定に Red Hat が使用する基準を示しています。これらのカテゴリーは、ハードウェアの購入および設定決定に関する一般的なガイドラインとして提供され、一意のワークロードの競合に対応するように調整できます。実際に選択されるハードウェア構成は、特定のワークロードミックスとベンダーの能力によって異なります。

### 最適化した IOPS

IOPS が最適化されたストレージクラスターには、通常、以下のプロパティーがあります。

- IOPS あたり最小コスト
- 1 GB あたりの最大 IOPS。
- 99 パーセンタイルのレイテンシーの一貫性。

IOPS に最適化されたストレージクラスタの一般的な用途は以下の通りです。

- 典型的なブロックストレージ。
- ハードドライブ (HDD) の 3x レプリケーションまたはソリッドステートドライブ (SSD) の 2x レプリケーション。
- OpenStack クラウド上の MySQL

### 最適化されたスループット

スループットが最適化されたストレージクラスターには、通常、以下のプロパティーがあります。

- MBps あたりの最小コスト (スループット)。
- TB あたり最も高い MBps。
- BTU あたりの最大 MBps
- Watt あたりの MBps の最大数。
- 97 パーセンタイルのレイテンシーの一貫性。

スループットを最適化したストレージクラスタの一般的な用途は以下の通りです。

- ブロックまたはオブジェクトストレージ。
- 3x レプリケーション。
- ビデオ、音声、およびイメージのアクティブなパフォーマンスストレージ。
- ストリーミングメディア。

### コストおよび容量の最適化

コストおよび容量が最適化されたストレージクラスターには、通常以下のプロパティーがあります。

- TB あたり最小コスト
- TB あたり最小の BTU 数。
- TB あたりに必要な最小 Watt。

通常は、コストおよび容量が最適化されたストレージクラスターに使用されます。

- 典型的なオブジェクトストレージ。
- 使用可能容量を最大化するためのイレイジャーコーディングの共通化
- オブジェクトアーカイブ。
- ビデオ、音声、およびイメージオブジェクトのリポジトリー。

### パフォーマンスドメインの仕組み

データの読み取りおよび書き込みを行う Ceph クライアントインターフェースに対して、Ceph Storage クラスターはクライアントがデータを格納する単純なプールとして表示されます。ただし、ストレージクラスターは、クライアントインターフェイスから完全に透過的な方法で多くの複雑な操作を実行します。Ceph クライアントおよび Ceph オブジェクトストレージデーモン (Ceph OSD または単に OSD) はいずれも、オブジェクトのストレージおよび取得にスケーラブルなハッシュ (CRUSH) アルゴリズムで制御されたレプリケーションを使用します。OSD は、OSD ホスト (クラスター内のストレージサーバー) で実行されます。

CRUSH マップはクラスターリソースのトポロジーを表し、マップは、クラスター内のクライアントノードと Ceph Monitor (MON) ノードの両方に存在します。Ceph クライアントおよび Ceph OSD はどちらも CRUSH マップと CRUSH アルゴリズムを使用します。Ceph クライアントは OSD と直接通信することで、オブジェクト検索の集中化とパフォーマンスのボトルネックとなる可能性を排除します。CRUSH マップとピアとの通信を認識することで、OSD は動的障害復旧のレプリケーション、バックフィル、およびリカバリーを処理できます。

Ceph は CRUSH マップを使用して障害ドメインを実装します。Ceph は CRUSH マップも使用してパフォーマンスドメインを実装します。パフォーマンスドメインは、基盤のハードウェアのパフォーマンスプロファイルを考慮してください。CRUSH マップは Ceph のデータの格納方法を記述し、これは単純な階層 (非周期グラフ) およびルールセットとして実装されます。CRUSH マップは複数の階層をサポートし、ハードウェアパフォーマンスプロファイルのタイプを別のタイプから分離できます。RHCS 2 以前では、パフォーマンスドメインは個別の CRUSH 階層に存在していました。RHCS 3 では、Ceph はデバイス「classes」でパフォーマンスドメインを実装します。

以下の例では、パフォーマンスドメインを説明します。

- ハードディスクドライブ (HDD) は、一般的にコストと容量を重視したワークロードに適しています。
- スループットを区別するワークロードは通常、ソリッドステートドライブ (SSD) の Ceph 書き込みジャーナルで HDD を使用します。
- MySQL や MariaDB のような IOPS を多用するワークロードでは、SSD を使用することが多いです。

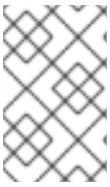
これらのパフォーマンスドメインはすべて、Ceph Storage クラスターに共存できます。

# 第4章 サーバーおよびラックソリューション

ハードウェアベンダーは、最適化されたサーバーレベルとラックレベルのソリューション SKU を提供することで、Ceph に対する熱意に応えてきました。Red Hat との共同テストで検証されたこれらのソリューションは、特定のワークロードに合わせて Ceph ストレージを拡張するための便利なモジュール式のアプローチにより、Ceph の導入において予測可能な価格対性能比を提供します。

一般的なラックレベルのソリューションには、以下が含まれます。

- **ネットワーク切り替え:** 冗長性のあるネットワークスイッチはクラスターを相互に接続し、クライアントへのアクセスを提供します。
- **Ceph MON ノード:** Ceph モニターはクラスター全体の健全性を確保するためのデータストアで、クラスターログが含まれます。実稼働環境でのクラスタークォーラムには、最低 3 台の監視ノードが強く推奨されます。
- **Ceph OSD ホスト:** Ceph OSD ホストはクラスタのストレージ容量を収容し、個々のストレージデバイスごとに 1 つ以上の OSD を実行します。OSD ホストは、ワークロードの最適化と、インストールされているデータデバイス (HDD、SSD、または NVMe SSD) の両方に応じて選択および設定されます。
- **Red Hat Ceph Storage:** サーバーおよびラックレベルのソリューション SKU の両方がバンドルされている Red Hat Ceph Storage の容量ベースのサブスクリプションを提供しています。



**注記**

Red Hat は、サーバーとラックソリューションにコミットする前に、**Red Hat Ceph Storage:Supported Configurations** のアーティクルを確認することを推奨します。その他のサポートは、**Red Hat サポート** にお問い合わせください。

### IOPS 最適化ソリューション

フラッシュストレージの利用が拡大する中、企業は Ceph ストレージクラスターで IOPS を多用するワークロードをホストし、プライベートクラウドストレージで高性能なパブリッククラウドソリューションをエミュレートするケースが増えています。これらのワークロードは通常、MySQL、MariaDB、または PostgreSQL ベースのアプリケーションからの構造化データを必要とします。

Ceph 書き込みジャーナルを同じ場所に配置した NVMe SSD は、通常、OSD をホストしています。一般的なサーバーには、以下の要素が含まれます。

- **CPU:** NVMe SSD ごとに 10 コア (2 GHz CPU を想定)。
- **RAM:** 16 GB ベースラインに加えて、OSD ごとに 5 GB
- **ネットワーク:** 2 OSD あたり 10 ギガビットイーサネット (GbE)
- **OSD メディア:** 高性能で高耐久のエンタープライズ NVMe SSD。
- **OSD:** NVMe SSD あたり 2 つ。
- **ジャーナルメディア:** 高性能で高耐久のエンタープライズ NVMe SSD (OSD と同じ場所に配置)
- **コントローラー:** ネイティブ PCIe バス。

**注記**

非 NVMe SSD の場合は、**CPU** 用に SSD OSD ごとに 2 つのコアを使用します。

**表4.1 IOPS が最適化された Ceph ワークロードのソリューション SKU (クラスターサイズ別)**

| ベンター | 小規模 (250TB) | 中規模 (1PB) | 大規模 (2PB 以上) |
|---|---|---|---|
| SuperMicro [a] | SYS-5038MR-OSD006P | 該当なし | 該当なし |

[a] 詳細は、「Supermicro® Total Solution for Ceph」を参照してください。

### スループットが最適化されたソリューション

スループットが最適化された Ceph ソリューションは通常、半構造化または非構造化データをベースとしています。大規模なブロックの連続 I/O は一般的です。OSD ホストのストレージメディアは通常 HDD で、SSD ベースのボリュームに書き込みジャーナルがあります。

一般的なサーバー要素には以下が含まれます。

- **CPU:** HDD ごとに 0.5 コア (2 GHz CPU を想定)
- **RAM:** 16 GB ベースラインに加えて、OSD ごとに 5 GB
- **ネットワーク:** クライアント向けネットワークおよびクラスター向けネットワーク用に、それぞれ 12 OSD ごとに 10 GbE
- **OSD メディア:** 7200 RPM のエンタープライズ HDD
- **OSD:** HDD ごとに 1 つ。
- **ジャーナルメディア:** 高耐久で高性能のエンタープライズシリアル接続 SCSI (SAS) または NVMe SSD。
- **OSD 対ジャーナルの比率:** 4-5:1 (SSD ジャーナルの場合)、または NVMe ジャーナルの場合は 12-18:2。
- **ホストバスアダプター(HBA):** 大量のディスク (JBOD)。

いくつかのベンダーは、スループットが最適化された Ceph ワークロードのための設定済みのサーバーおよびラックレベルのソリューションを提供しています。Red Hat は、Supermicro および Quanta Cloud Technologies (QCT) からサーバーのテストや評価を徹底して行っています。

**表4.2 Ceph OSD、MON、および TOR (top-of-rack) スイッチ向けのラックレベルの SKU。**

| ベンター | 小規模 (250TB) | 中規模 (1PB) | 大規模 (2PB 以上) |
|---|---|---|---|
| SuperMicro [a] | SRS-42E112-Ceph-03 | SRS-42E136-Ceph-03 | SRS-42E136-Ceph-03 |

**表4.3 個別の OSD サーバー**

| ベンター | 小規模 (250TB) | 中規模 (1PB) | 大規模 (2PB 以上) |
|---|---|---|---|
| SuperMicro [a] | SSG-6028R-OSD072P | SSG-6048-OSD216P | SSG-6048-OSD216P |
| QCT [a] | QxStor RCT-200 | QxStor RCT-400 | QxStor RCT-400 |

[a] 詳細は、「QCT: QxStor Red Hat Ceph Storage Edition」を参照してください。

**表4.4 スループットが最適化された Ceph OSD ワークロードの追加サーバー設定**

| ベンター | 小規模 (250TB) | 中規模 (1PB) | 大規模 (2PB 以上) |
|---|---|---|---|
| Dell | PowerEdge R730XD [a] | DSS 7000 [b], twin node | DSS 7000、ツインノード |
| Cisco | UCS C240 M4 | UCS C3260 [c] | UCS C3260 [d] |
| Lenovo | System x3650 M5 | System x3650 M5 | 該当なし |

[a] 詳細は、『Dell PowerEdge R730xd Performance and Sizing Guide for Red Hat Ceph Storage - A Dell Red Hat Technical White Paper』を参照してください。

[b] 詳細は、『Dell EMC DSS 7000 Performance & Sizing Guide for Red Hat Ceph Storage』を参照してください。

[c] 詳細は、「Red Hat Ceph Storage hardware reference architecture」を参照してください。

[d] 詳細は、「UCS C3260」を参照してください。

### コストおよび容量が最適化されたソリューション

コストと容量が最適化されたソリューションは、一般的に大容量化、またはより長いアーカイブシナリオに焦点を当てています。データは、半構造化または非構造化のいずれかになります。ワークロードには、メディアアーカイブ、ビッグデータアナリティクスアーカイブ、およびマシンイメージのバックアップが含まれます。大規模なブロックの連続 I/O は一般的です。より大きな費用対効果を得るために、OSD は通常、Ceph の書き込みジャーナルを HDD 上に併置してホストされています。

ソリューションには、通常、以下の要素が含まれます。

- **CPU:**HDD あたり 0.5 コア (2 GHz CPU を想定)
- **RAM:**16 GB のベースラインに加えて、OSD ごとに 5 GB。
- **ネットワーク:**12 OSD ごとに 10 Gb (それぞれクライアント向けおよびクラスター向けネットワーク用)
- **OSD メディア:**7200 RPM のエンタープライズ HDD。
- **OSD:**HDD ごとに 1 つ。
- **ジャーナルメディア:**HDD の同一場所に配置

- **HBA:**JBOD。

Supermicro および QCT は、コストと容量を重視した Ceph ワークロード向けに、構成済みのサーバーとラックレベルのソリューション SKU を提供しています。

**表4.5 構成済みコストと容量が最適化されたワークロードのためのラックレベル SKU**

| ベンター | 小規模 (250TB) | 中規模 (1PB) | 大規模 (2PB 以上) |
|---|---|---|---|
| SuperMicro [a] | 該当なし | SRS-42E136-Ceph-03 | SRS-42E172-Ceph-03 |

**表4.6 コストと容量が最適化されたワークロード用の構成済みのサーバーレベル SKU**

| ベンター | 小規模 (250TB) | 中規模 (1PB) | 大規模 (2PB 以上) |
|---|---|---|---|
| SuperMicro [a] | 該当なし | SSG-6048R-OSD216P [a] | SSD-6048R-OSD360P |
| QCT | 該当なし | QxStor RCC-400 [a] | QxStor RCC-400 [a] |
| [a] 詳細は、「Supermicro's Total Solution for Ceph」を参照してください。 | | | |

**表4.7 コストと容量が最適化されたワークロード用に構成可能な追加サーバー**

| ベンター | 小規模 (250TB) | 中規模 (1PB) | 大規模 (2PB 以上) |
|---|---|---|---|
| Dell | 該当なし | DSS 7000、ツインノード | DSS 7000、ツインノード |
| Cisco | 該当なし | UCS C3260 | UCS C3260 |
| Lenovo | 該当なし | System x3650 M5 | 該当なし |

**関連情報**

- Samsung NVMe SSD 上の Red Hat Ceph Storage
- Red Hat Ceph Storage on the InfiniFlash All-Flash Storage System from SanDisk
- Deploying MySQL Databases on Red Hat Ceph Storage
- Intel® Data Center Blocks for Cloud – Red Hat OpenStack Platform with Red Hat Ceph Storage
- Red Hat Ceph Storage on QCT Servers
- Red Hat Ceph Storage on Servers with Intel Processors and SSDs

# 第5章 コンテナー化された CEPH のハードウェアの最小推奨事項

Ceph は、プロプライエタリーでない商用ハードウェア上で稼働することができます。小規模な実稼働クラスターや開発クラスターは、適度なハードウェアで性能を最適化せずに動作させることができます。

| Process | 条件 | 最小推奨 |
|---|---|---|
| **ceph-osd-container** | プロセッサー | OSD コンテナーごとに 1x AMD64 または Intel 64 CPU CORE |
| | RAM | 1 OSD コンテナーごとに最小 5 GB の RAM |
| | OS ディスク | ホストごとに 1x OS ディスク |
| | OSD ストレージ | OSD コンテナーごとに 1x ストレージドライブ。OS ディスクと共有できません。 |
| | **block.db** | 任意ですが、Red Hat は、デーモンごとに SSD、NVMe または Optane パーティション、または lvm を 1 つ推奨します。サイズ設定は、オブジェクト、ファイルおよび混合ワークロード用に BlueStore に **block.data** の 4%、ブロックデバイス、Openstack cinder 、および Openstack cinder ワークロード用に BlueStore に **block.data** の 1% です。 |
| | **block.wal** | 任意ですが、デーモンごとに 1x SSD、NVMe または Optane パーティション、または論理ボリューム。サイズが小さい (10 GB など) を使用し、**block.db** デバイスよりも高速の場合にのみ使用します。 |
| | ネットワーク | 2x 10GB イーサネット NIC |
| **ceph-mon-container** | プロセッサー | mon-container ごとに 1x AMD64 または Intel 64 CPU CORE |
| | RAM | **mon-container** あたり 3 GB |
| | ディスク容量 | **mon-container** ごとに 10 GB (50 GB 推奨) |
| | 監視ディスク | 任意で、**Monitor rocksdb** データ用の 1x SSD ディスク |
| | ネットワーク | 2x 1GB イーサネット NIC、10 GB 推奨 |
| **ceph-mgr-container** | プロセッサー | **mgr-container**ごとに 1x AMD64 または Intel 64 CPU CORE |
| | RAM | **mgr-container** あたり 3 GB |
| | ネットワーク | 2x 1GB イーサネット NIC、10 GB 推奨 |

| Process | 条件 | 最小推奨 |
|---|---|---|
| **ceph-radosgw-container** | プロセッサー | radosgw-container ごとに 1x AMD64 または Intel 64 CPU CORE |
| | RAM | デーモンごとに 1 GB |
| | ディスク容量 | デーモンごとに 5 GB |
| | ネットワーク | 1x 1GB イーサネット NIC |
| **ceph-mds-container** | プロセッサー | mds-container ごとに 1x AMD64 または Intel 64 CPU CORE |
| | RAM | **mds-container** あたり 3 GB<br><br>この数は、設定可能な MDS キャッシュサイズに大きく依存します。通常、RAM 要件は、**mds_cache_memory_limit** 構成設定に設定された量の 2 倍です。また、これはデーモンのためのメモリーであり、全体的なシステムメモリではないことにも注意してください。 |
| | ディスク容量 | **mds-container** ごとに 2 GB、さらにデバッグロギングに必要な追加の領域を考慮すると、20GB から始めてみることが推奨されます。 |
| | ネットワーク | 2x 1GB イーサネット NIC、10 GB 推奨<br><br>これは、OSD コンテナーと同じネットワークであることに注意してください。OSD で 10GB のネットワークを使用している場合は、MDS でも同じものを使用することで、レイテンシーの面で MDS が不利にならないようにする必要があります。 |

# 第6章 RED HAT CEPH STORAGE DASHBOARD の推奨される最小ハードウェア要件

Red Hat Ceph Storage Dashboard のハードウェアの最低要件を以下に示します。

**最小要件**

- 4 コアプロセッサー (2.5 GHz 以上)
- 8 GB RAM
- 50 GB のハードドライブ
- 1 Gigabit イーサネットネットワークインターフェース

**関連情報**

- 詳細は、『Administration Guide』の「Monitoring a Ceph storage cluster with the Red Hat Ceph Storage Dashboard」を参照してください。