

ETERNUS AX series オールフラッシュアレイ , ETERNUS HX series ハイブリッドアレイ

ONTAP SAN 構成のベストプラクティス

ONTAP 9

目次

1.	概要.....	7
1.1	対象	7
1.2	注意事項	7
2.	ベストプラクティスのまとめ	8
3.	SAN 機能.....	10
3.1	高速フェイルオーバー	10
3.2	パフォーマンス容量	11
3.3	低レイテンシの一貫したパフォーマンス	12
3.4	クラスタ間のコピーオフロード	12
3.5	NVMe over Fabrics(NVMe-oF) プロトコル、NVMe over Fibre Channel(NVMe/FC) サポート	12
3.6	SnapMirror Synchronous	13
3.7	Asymmetric Namespace Access (ANA) サポート	13
3.8	512 バイトのブロックサイズのサポート	13
3.9	NVMe で VMware Compare and Write をサポート	14
3.10	NVMe 無停止ボリューム移動	14
3.11	NVMe QoS サポート	14
4.	SAN/NAS 共通の機能.....	15
4.1	最低限のサービス品質	15
4.2	Balanced placement	15
5.	ONTAP 9.7 の新機能.....	16
5.1	All SAN Array (ASA)	16
6.	ONTAP 9.8 の新機能.....	17
6.1	仮想マシン ID (VMID)	17
6.2	NVMe-oF プロトコルの共存	18
6.3	より大きな LUN の最大サイズ	18
6.4	Persistent ports	18
7.	ONTAP 9.9.1 の新機能.....	20
7.1	System Manager	20
7.2	SnapMirror Business Continuity (SMBC)	20
7.3	ASA のクラスタの最大サイズは 1 組の HA ペアから 12 ノードに増加	21

7.4	単一 LUN のパフォーマンス向上	21
7.5	ネストされた igroup	22
7.6	NVMe/FC vVol のサポート	22
7.7	NVMe リモート I/O のサポート	23
7.8	ASA では NVMe/FC のサポートを追加	23
8.	ONTAP プロトコルと SAN プロトコル.....	24
8.1	ONTAP の概要	24
8.2	SAN のパフォーマンスを最適化するための考慮事項	24
8.2.1	ボリューム	24
8.3	LUN	26
8.3.1	FC インオーダー配信	27
8.3.2	結論	27
8.4	ボリューム構成	28
8.5	ホスト接続	28
8.6	パス選択	29
8.7	パス選択の変更	30
8.8	FC と NPIV	32
8.9	パス管理と Selective LUN Mapping	34
8.10	Selective LUN Mapping	35
8.11	ポートセット	36
8.12	管理インターフェース	36
8.13	DataMotion for LUNs	38
8.13.1	LUN 移動と LUN コピーの比較	38
8.13.2	ストレージ効率に関する考慮事項	39
8.13.3	データ保護に関する考慮事項	39
8.13.4	拡張性とスループットに関する考慮事項	39
8.13.5	データ管理とワークフローに関する考慮事項	40
8.13.6	DataMotion と Selective LUN Mapping：パスの検出と破棄	40
8.14	パス管理のベストプラクティス	41
9.	拡張性に優れた SAN の価値ある提案と機能	42
9.1	統合ターゲットおよび管理単位としての SVM	42
9.2	ノードおよびクラスタレベルの拡張性	42
9.3	クラスタ間の LUN と LIF の移動	42
9.4	Foreign LUN Import (FLI)	43

目次

図 3.1	使用済みパフォーマンス容量と最適点を示すパフォーマンス容量	11
図 3.2	FC フレーム対 NVMe/FC フレーム	12
図 6.1	VMID	17
図 6.2	Persistent ports	19
図 6.3	Persistent ports ゾーニングの例	19
図 7.1	SMBC のトポロジ	20
図 7.2	12 ノード SAN	21
図 7.3	リモート I/O サポートなしの NVMe-oF	23
図 7.4	リモート I/O サポートありの NVMe-oF	23
図 8.1	作業をより多くのボリュームに分散する効果	25
図 8.2	複数の LUN に作業を分散する効果	27
図 8.3	ONTAP でのパスの概要	29
図 8.4	HA フェイルオーバー中のパス	30
図 8.5	ポートまたはスイッチの障害時のパス	31
図 8.6	ボリュームまたは LUN の移動中のパス	31
図 8.7	ボリュームまたは LUN の移動後のパス	32
図 8.8	System Manager の FC アダプタ	34
図 8.9	System Manager のネットワークインターフェース	34
図 8.10	SVM 作成時の管理 LIF の作成	37
図 8.11	既存の SVM に対する管理 LIF の作成	37
図 8.12	管理 LIF の詳細	38

表目次

表 3.1	高速フェイルオーバーのテイクオーバーとギブバックタイミングのガイドライン	11
表 9.1	ONTAP の拡張性.....	42

はじめに

本書では、ETERNUS AX/HX series の SAN の概要とベストプラクティスについて説明します。

第2版
2025年3月

登録商標

本製品に関連する他社商標については、以下のサイトを参照してください。
<https://www.fujitsu.com/jp/products/computing/storage/trademark/>
本書では、本文中の™、®などの記号は省略しています。

本書の読み方

対象読者

本書は、ETERNUS AX/HX の設定、運用管理を行うシステム管理者、または保守を行うフィールドエンジニアを対象としています。必要に応じてお読みください。

関連マニュアル

ETERNUS AX/HX に関連する最新の情報は、以下のサイトで公開されています。
<https://www.fujitsu.com/jp/products/computing/storage/manual/>

本書の表記について

■ 本文中の記号

本文中では、以下の記号を使用しています。

注意

お使いになるときに注意していただきたいことを記述しています。必ずお読みください。

備考

本文を補足する内容や、参考情報を記述しています。

1. 概要

本書では、SAN 接続ホストの観点から、ETERNUS AX/HX series のクラスタ化された SAN の概要を説明します。また、ETERNUS AX series の SAN 構成を使用してパフォーマンスを最適化する方法を示します。さらに、ONTAP の高可用性とデータ移行機能を活用するためのベストプラクティスについても説明します。

1.1 対象

本書は、ONTAP 9.7 以降を実行するストレージソリューションで、iSCSI や FC などのソリューションを設計するシステムおよびストレージのアーキテクトを対象としています。読者は次の知識があることを前提としています。

- ハードウェアおよびソフトウェアソリューションに関する一般的な知識がある
- FC や iSCSI などのブロックアクセスプロトコルに精通している

1.2 注意事項

この文書は ONTAP 管理の一般的な入門書ではありません。概要については、「ETERNUS AX/HX Series システムアドミニストレーションリファレンス」と「ETERNUS AX/HX Series SAN アドミニストレーションガイド」を参照してください。SAN プロトコルを使用する ONTAP クラスタの SAN 関連の制限事項については、「ETERNUS AX/HX Series SAN 構成ガイド」を参照してください。これらの文書は[マニュアルサイト](#)に掲載されています。

2. ベストプラクティスのまとめ

各ベストプラクティスの詳細については、次のリンクを参照してください。

- SVM の作成と同時に FCP または iSCSI サービスを作成します。
[「8.2 SAN のパフォーマンスを最適化するための考慮事項」](#)
- 管理作業を軽減し、整合性グループとして機能し、ストレージ効率を最大化するために、関連性があり、同様のパフォーマンス要件と管理要件を持つ LUN を同じボリュームに結合します。
[「8.2 SAN のパフォーマンスを最適化するための考慮事項」](#)
- ボリューム数を増やしてパフォーマンスを最適化し、LUN 数を増やします。ほとんどの場合、理想的なレイアウトは、約 8 個のボリュームと 8 ～ 16 個の LUN です。
[「8.2 SAN のパフォーマンスを最適化するための考慮事項」](#)
- LUN を含むボリュームは、FCP または iSCSI を使用してデータを提供するためにネームスペースに結合する必要はありません。
[「8.4 ボリューム構成」](#)
- SVM は、iSCSI またはファイバーチャネルを使用してデータを提供する予定のすべてのストレージコントローラ上に、Ethernet ネットワークまたはファイバーチャネルファブリックごとに 1 つの LIF を持つ必要があります。
[「8.6 パス選択」](#)
- クラスタ化された ONTAP ストレージに接続されたファイバーチャネルファブリックでは、N_Port ID virtualization (NPIV) が有効になっている必要があります。
[「8.8 FC と NPIV」](#)
- ファイバチャネルファブリックゾーニング構成では、ターゲットとして NPIV 仮想 World Wide Port Name (WWPN) のみを使用します。ターゲットポートの物理 WWPN は使用しないでください。
[「8.8 FC と NPIV」](#)
- Selective LUN mapping とは、ほとんどの LUN にストレージコントローラとその高可用性 (HA) パートナーに対応するパスがそれぞれ 4 本 (直接接続 2 本、間接接続 2 本) ある構成です。デフォルトでは、LUN が同クラスタ内の別の HA ペアに移動するたびに LUN マッピングを変更します。
[「8.12 管理インターフェース」](#)
- データ移動操作を容易にするため、または追加の I/O リソースを活用するために、必要に応じてより多くのパスを作成します。ただし、ホスト OS がサポートできるパスの最大数を超えないようにします。
[「8.13.6 DataMotion と Selective LUN Mapping：パスの検出と破棄」](#)
- LUN マッピングを変更する場合は、ホスト上の標準的な手順に従って新しいパスを検出し、削除されたパスを破棄します。
[「8.13.6 DataMotion と Selective LUN Mapping：パスの検出と破棄」](#)
- FCP または iSCSI でデータを提供する SVM には、SVM 管理インターフェースが必要です。
[「8.12 管理インターフェース」](#)
- 重複排除または圧縮されている LUN 上で LUN move を使用する場合は、デスティネーションボリュームでもこれらのポリシーが有効になっていることを確認してください。
[「8.13.2 ストレージ効率に関する考慮事項」](#)
- LUN move を使用して、Snapshot コピーの管理に使用されるソフトウェアによって保護されているボリュームから LUN を移動する場合、そのソフトウェアは、ローカル以外の Snapshot コピーを持つ LUN が存在する可能性を認識する必要があります。
[「8.13.3 データ保護に関する考慮事項」](#)
- LUN のレプリケーションをより詳細に制御するには、LUN move の一時停止およびスロットル機能を使用します。
[「8.13.4 拡張性とスループットに関する考慮事項」](#)
- LUN move を使用して、既存のデータ移動およびレプリケーションワークフローを短縮します。
[「8.13.5 データ管理とワークフローに関する考慮事項」](#)
- マニュアルサイトに掲載の「ETERNUS AX/HX Series SAN 構成ガイド」で指定されているように、SAN データを提供するクラスタのクラスタサイズ制限を超えないようにしてください。

- Active IQ Unified Manager を使用して、クラスタ内の全ノードについて、パフォーマンスがその容量以下であることを確認します。vol move を使用して、パフォーマンス容量が過剰にプロビジョニングされたノード上にあるワークロードをリバランスします。
[「3.2 パフォーマンス容量」](#)
- ファブリック内のすべてのファイバチャネルスイッチについて、In Order Delivery の保証を設定します。
[「8.3.1 FC インオーダー配信」](#)

3. SAN 機能

ONTAP 9 の SAN 関連機能には、次のようなものがあります。

- クラスタのハードニング。
- クォーラム処理の改善 (いくつかの自動修復機能を含む)。
- First Failure Detection と信頼性、可用性、保守性 (RAS) の強化。
- パフォーマンスの向上を確認するための Foreign LUN Import (FLI)。インポートされた LUN 検証ワークフローのパフォーマンスが向上し、検証をより迅速に完了できるようになりました。
- FLI のハードニング。FLI では、テイクオーバーまたはギブバック (以前はインポートの再起動が必要だったイベント) の後でもインポートを存続させて再起動できるように、一部のハードニングが行われています。
- これらの拡張機能は、主に次の 2 つのカテゴリに分類されます。

これらの拡張機能は、主に次の 2 つのカテゴリに分類されます。

- **ハードニングと耐障害性**

これらは、ONTAP が障害に直面したときの回復力を高めるために ONTAP コードに加えられた拡張です。多くの場合、ONTAP は障害に遭遇し、人間の介入なしに自己修復することができます。

- **RAS**

これらの拡張機能は、障害をより簡単に分離、認識、診断できるように、メッセージの作成と適切なカウンタの収集を中心に分類されます。これにより、ストレージ管理者と当社サポート部門は、多くの場合外部から認識可能な影響が生じる前に、障害状態をより迅速に診断して解決できます。

ストレージアーキテクト、プロフェッショナルサービス、ストレージ管理者にとって重要で強い影響力のある追加機能には、次のようなものがあります。

- 高速フェイルオーバー
- パフォーマンス容量
- 低レイテンシの一貫したパフォーマンス
- クラスタ間のコピーオフロード
- NVMe over Fabrics(NVMe-oF) プロトコル、NVMe over Fibre Channel(NVMe/FC) サポート
- SnapMirror Synchronous
- Asymmetric Namespace Access (ANA) サポート
- 512 バイトのブロックサイズのサポート
- NVMe で VMware Compare and Write をサポート
- NVMe 無停止ボリューム移動
- NVMe QoS サポート

これらの機能については、以降のセクションで説明します。

3.1 高速フェイルオーバー

ONTAP 9 では引き続きコードの最適化と機能強化が行われており、ONTAP HA ペアがパートナーのワークロードのテイクオーバーとギブバックに要する時間が短縮されています。ETERNUS AX series の SAN 構成で運用する場合、計画的なテイクオーバーとギブバックは 2 ～ 10 秒以内に完了します。計画外のテイクオーバーとギブバックは、2 ～ 15 秒以内に完了します。これらの測定はいずれも、HA ペアの一方のノードが、HA ペアの他方のノードからフェイルオーバーまたはギブバックをするのにかかる時間です。テストでは、ほとんどのオペレーティングシステムスタックは、テイクオーバーまたはギブバックイベントが開始されてから 4 ～ 7 秒後に I/O を再開しました。

3. SAN 機能

3.2 パフォーマンス容量

表 3.1 に、ONTAP のプラットフォームタイプおよびバージョン別のフェイルオーバー時間を要約します。すべての番号は、HA ペアが[マニュアルサイト](#)に掲載の「ETERNUS AX/HX series ビジネス・クリティカルなワークロード向けのベスト・プラクティス」で定義される所定の SAN 構成にあることを前提としています。

表 3.1 高速フェイルオーバーのテイクオーバーとギブバックタイミングのガイドライン

プラットフォーム	計画的テイクオーバー	計画外のテイクオーバー
ETERNUS AX series の ONTAP 9	10 秒	15 秒
Flash Pool または SSD アグリゲートを使用した ETERNUS HX series	30 秒	60 秒

3.2 パフォーマンス容量

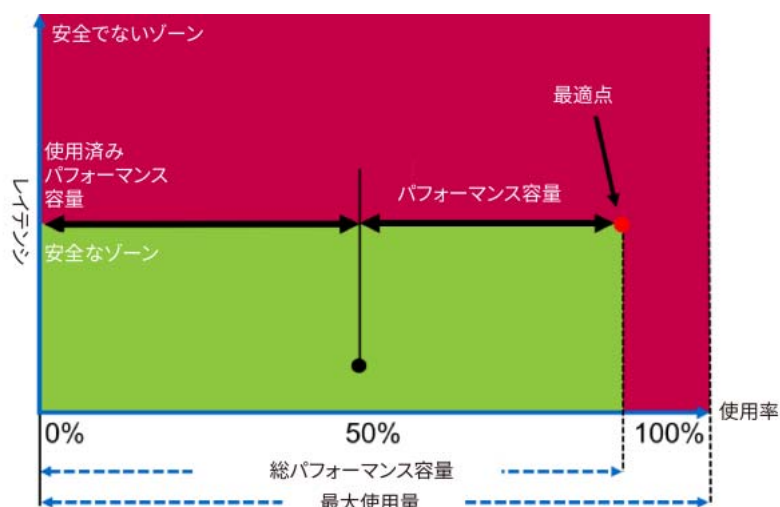
パフォーマンス容量は、コントローラによって収集されたカウンタマネージャ統計情報を使用する機能です。Active IQ Unified Manager は、この統計情報を使用してパフォーマンスカウンタを分析し、一貫性のある低レイテンシを維持しながら生成できる最大 IOPS を動的に最適化します。つまり、ストレージ管理者は、追加のワークロードを追加するために、コントローラまたは HA ペアに十分なパフォーマンスが残っているかどうかを推測する必要がなくなります。

ETERNUS AX series の SAN のビジネスクリティカルな構成では、Active IQ Unified Manager は、ONTAP で生成されたパフォーマンス容量計算を使用して、IOPS を最大化すると同時に、1 ミリ秒未満のレイテンシを維持します。

フェイルオーバーが存在する場合、レイテンシが多少高くなる可能性があります。明らかに、これは作業負荷計画で考慮する必要があります。

パフォーマンス容量データと Active IQ Unified Manager の視覚化を使用して、テイクオーバー時のパフォーマンスを許容可能なレイテンシ内に維持しながら IOPS を最適化できます。一貫した低レイテンシを維持しながらストレージパフォーマンスを最適化するための、パフォーマンス容量および Active IQ Unified Manager の使用の詳細については、[マニュアルサイト](#)に掲載の「ETERNUS AX/HX series ビジネス・クリティカルなワークロード向けのベスト・プラクティス」を参照してください。[図 3.1](#) は、パフォーマンス容量を示しています。安全なゾーンと安全でないゾーン (レイテンシによって定義されます) が表示され、現在使用されているパフォーマンス容量と最適なポイントも示されます。最適なポイントは、レイテンシターゲットを維持しながら IOPS を最大化するポイントです。グラフから、パフォーマンス容量とは、使用済みの容量を最適なポイントから差し引いたときに残っているパフォーマンス容量の量であることが容易にわかります。

図 3.1 使用済みパフォーマンス容量と最適点を示すパフォーマンス容量



3.3 低レイテンシの一貫したパフォーマンス

ONTAP 9 は、パフォーマンス容量と Active IQ Unified Manager を使用するとともに ETERNUS AX series の規定の SAN 構成ガイドラインに従うと、一貫性のある低レイテンシのパフォーマンスを実現し、保証することができます。

3.4 クラスタ間のコピーオフロード

クラスタ間コピーオフロードは、Windows が LUN 間のコピーブロックを無停止でオフロードできるようにする ONTAP 9 の機能です。これにより、LUN へのアクセスを維持しながら、CPU 負荷、ネットワーク帯域幅、移動時間を大幅に削減できます。LUN は移動中ずっとオンラインのままです。この機能は、Microsoft Windows Server のコピーオフロード機能 (ODX) の拡張機能です。SAN では、Windows LUN でのみサポートされます。

3.5 NVMe over Fabrics(NVMe-oF) プロトコル、NVMe over Fibre Channel(NVMe/FC) サポート

NVMe over Fabrics(NVMe-oF) プロトコル、NVMe over Fibre Channel(NVMe/FC) をサポートしています。NVMe/FC は、FC フレーム内に NVMe コマンドセットをカプセル化し、[図 3.2](#) に示すように、既存の SCSI-3 コマンド記述子ブロックを置き換えます。

図 3.2 FC フレーム対 NVMe/FC フレーム

- FCP – FCフレーム内にカプセル化されたSCSI-3コマンドセット



- FC-NVMe – FCフレーム内にカプセル化されたNVMeコマンドセット

新しい NVMe コマンドセットは次のとおりです。

- コマンドの合理化
- すべてのソフトウェアロックの削除
- コンテキストスイッチの削減
- マルチスレッドの増加：最大キュー深度が 64K の 64K キュー

これらの最適化によって、使用されているブロックプロトコルを置き換えるだけで、スループットを向上させ、ワークロードのレイテンシを削減する、はるかに効率的な高パフォーマンスプロトコルが作成されました。新しいプロトコルのメリットを得るためにアプリケーションを書き直す必要はありません。

3.6 SnapMirror Synchronous

SnapMirror Synchronous(SM-S) により、同期レプリケーションが可能です。MetroCluster は同期レプリケーションを提供しますが、すべてのデータをミラーリングする必要があります。この同期レプリケーション機能により、同期レプリケーションが必要なワークロードを選択できるようになりました。SM-S には次の属性と機能があります。

- FC および iSCSI 向けのボリューム単位の同期データレプリケーションです。
- 目標復旧ポイント (RPO) はゼロ、目標復旧時間 (RTO) は極めて低いです。ビジネス継続性ソリューションではありません。
- 追加の外部ハードウェア、ソフトウェア、またはネットワークは必要ありません。
- レプリケーションエラーが発生しても、プライマリボリュームへのアプリケーション I/O は中断されません。レプリケーションエラーが修正されると、自動的にリカバリおよび再同期が実行されます。
- strict モードでプライマリとセカンダリボリューム間に 100% の同期性を保証します。
- SM-S は、クラスタ全体のレプリケーションではなく、ボリュームレベルでレプリケーションとデータ保護を詳細に管理したいお客様に適しています (クラスタ全体のレプリケーションの場合は MetroCluster を使用してください)。
- SM-S は、必要な SnapMirror ライセンス以外の追加ライセンスです。

3.7 Asymmetric Namespace Access (ANA) サポート

NVMe/FC ターゲットの一部として ANA を導入しています。非対称論理ユニットアクセス (ALUA) と同様に、ANA ではイニシエーター側とターゲット側の両方の実装を使用して、各 OS スタックで使用されるストレージ高可用性マルチパスソフトウェアでホスト側のマルチパス実装が動作するための、すべてのパスおよびパス状態情報を提供できるようにします。ANA が機能するためには、ターゲットとイニシエータの両方が ANA を実装し、サポートする必要があります。どちらかの側が利用または実装されていない場合、ANA は機能できず、NVMe/FC はストレージの高可用性をサポートしていない状態に戻ります。このような状況では、アプリケーションは冗長性のために高可用性をサポートする必要があります。

NVMe/FC は ANA プロトコルに依存して、パスとターゲットの両方のフェイルオーバーに必要なマルチパスとパス管理を提供します。ANA プロトコルは、ホストがあるパスから別のパスへのパスとフェイルオーバーを管理できるように、NVMe サブシステムがパスおよびサブシステムエラーをホストにどのように通信するかを定義します。ANA は NVMe/FC において、ALUA が FCP プロトコルと iSCSI プロトコルの両方で果たす役割と同じ役割を果たします。マルチパス I/O (MPIO) または Device Mapper Multipathing (DM- マルチパス) などのホスト OS パス管理を備えた ANA は、NVMe/FC 向けのパス管理およびフェイルオーバー機能を提供します。

3.8 512 バイトのブロックサイズのサポート

ONTAP がネイティブにサポートしている 4K のブロックサイズに加えて、512 バイトのブロックサイズオプションが NVMe ネームスペースも含め、すべてでサポートされています。この機能は、VMware データストアや、512 バイトブロックを使用する Oracle Automatic Storage Management(ASM) ディスクグループなど、既存の 512 バイト構成との統合を効率化します。4096 バイト (4K バイト) ブロックがデフォルトのままです。ただし、新しいブロックサイズ引数と、512 バイトと 4096 バイトの両方の値が、`vserver nvme namespace create` コマンドとそれに対応する API の両方で使用できます。

3.9 NVMe で VMware Compare and Write をサポート

VMware Storage vMotion などの機能をサポートするために、ONTAP では VMware Compare and Write/ Atomic Test and Set(CAW/ATS) 融合オペレーションをサポートしています。ATS プリミティブは、CAW を使用して NVMe 比較コマンドと NVMe 書き込みコマンドを融合し、最初に比較操作を実行します。比較操作が成功すると、書き込みは完了します。失敗した場合、書き込みは中止されます。

3.10 NVMe 無停止ボリューム移動

NVMe-oF では、ボリューム移動のサポートが追加されています。これにより、ストレージ管理者は業務を停止することなく、NVMe ネームスペースを含むボリュームをアグリゲート間で移動できます。

3.11 NVMe QoS サポート

NVMe QoS サポートでは、イベント管理サービス (EMS) イベントである `nvmf.qos.mismatched.policy` が利用可能です。EMS は、ポリシーの不一致がある場合にユーザーに警告します。このイベントはソフト強制であり、`vserver nvme subsystem map add` 操作中にのみ検証されます。このイベントには、次の制限が適用されます。

- サブシステム内のすべてのネームスペースは、同じ QoS ポリシーを持つボリュームに存在する必要があります。
- サブシステム内のすべてのネームスペースは、正常に機能するために、同じ QoS スロットルポリシーを持つボリュームに存在する必要があります。

4. SAN/NAS 共通の機能

4.1 最低限のサービス品質

QoS の最大値により、管理者はワークロードに割り当てることができる IOP の最大値またはデータレートを割り当てることができます。これにより、ワークロードが急増したときに隣接するワークロードが不足するのを防ぐことができます。QoS の最小値は逆の方法で動作します。最小値がワークロードに関連づけられると、そのワークロードからの I/O が他の I/O よりも優先され、パフォーマンス要件を満たすことができます。最小値と最大値を組み合わせることで、共有環境のワークロードを効果的に形成し、優先順位を付けることができます。QoS の最小値は、ETERNUS AX series プラットフォームの SAN オブジェクト (ボリュームと LUN) 上で使用できます。

4.2 Balanced placement

Balanced placement は、既存の簡素化されたアプリケーションプロビジョニングを強化します。2 つのメトリックを使用して、各クラスタノードで使用可能な領域の量とすべてのノードに残っているパフォーマンス容量を分析し、使用可能なリソースに基づいて最適化されるワークロード配置決定を行うことで、新しくプロビジョニングされるワークロードの最適な場所を決定し、初期ワークロード配置を最適化します。

使用されているアプローチの概要を次に示します。

クラスタの各アグリゲートは、現在利用可能な (空きの) IOPS (パフォーマンス容量の残り / ヘッドルーム) を判断するために評価されます。これは、QoS ヘッドルームの計算で得たアグリゲートの最適な最大 IOPS から始めて、(再び QoS ヘッドルームで得た) アグリゲートの現在の IOPS または (アグリゲートに存在する全ボリュームのモデル化 IOPS の合計である) モデル化された IOPS のいずれかの最大値を引くことによって行われます。モデル化された IOPS は、(ボリューム上の QoS ポリシーによって設定される) 最大 IOPS の 10%、または最適なストレージサービス (QoS ポリシーが設定されていない場合) に基づいたボリュームサイズの IOPS/TB の 10% です。

アグリゲートは、ドライブタイプ (低コストを好むこと)、空き IOPS、空き容量の順にリスト内でソートされます。リストが検索され、IOPS と容量の要件を満たす最初のアグリゲートが選択されます。

ボリュームがプロビジョニングされた後、QoS の最大値は定義されたサービスレベルに基づいて設定されます。

また、Balanced placement には以下の拡張機能も導入されています。

- 適応型 QoS のサポート。
- AppDM で作成されたアプリケーションオブジェクトでサポートされるボリューム数が 10 から 16 に増加。
- Balanced placement による、既存アプリケーションコンポーネントのストレージサービスレベル要件の変更と、その要件をインプレースで満たすことができない場合における要求の拒否。

使用可能なリソースに基づいてサービスレベルのコミット要求を満たすことができない場合、変更はブロックされ、ストレージ管理者はリソースを追加するか、リソースの配置を手動で調整するか、要求された変更を破棄する必要があります。

5. ONTAP 9.7 の新機能

5.1 All SAN Array (ASA)

ASA は、対称アクティブ / アクティブの SAN 専用 HA ペアです。つまり、ASA は、両方のコントローラを経由して基礎となる LUN に接続するパスをアドバタイズして使用します。このアーキテクチャ変更の効果は、ホストが HA ペアのいずれかのコントローラでホストされている任意の LUN へのアクティブパスを常に持つことを保証することです。これは、さもなければ停止を伴う計画的または計画外のフェイルオーバーやギブバックをほぼ瞬時に行うため、重要です。これにより、2 つのノードのうちの 1 つをスタンバイモードで孤立させる必要があるなどの妥協がなく、最新のモジュラーアレイアーキテクチャにフレームアレイ機能がもたらされます。詳細については、[マニュアルサイト](#)の「ETERNUS AX/HX series ビジネス・クリティカルなワークロード向けのベスト・プラクティス」を参照してください。

6. ONTAP 9.8 の新機能

ONTAP 9.8 はいくつかの新機能を追加しており、その一部は統合プラットフォームと ASA プラットフォームの両方で使用できます。その他の機能は、最初は SAN のみの ASA で導入されています。これらの機能は、ONTAP の今後のリリースで統合 ONTAP クラスタに追加される可能性が高いと予想されます。ONTAP の新しい SAN 機能は次のとおりです。

- 仮想マシン ID (VMID)- 仮想マシン (VM) のテレメトリ機能の強化
- NVMe-oF プロトコルの共存

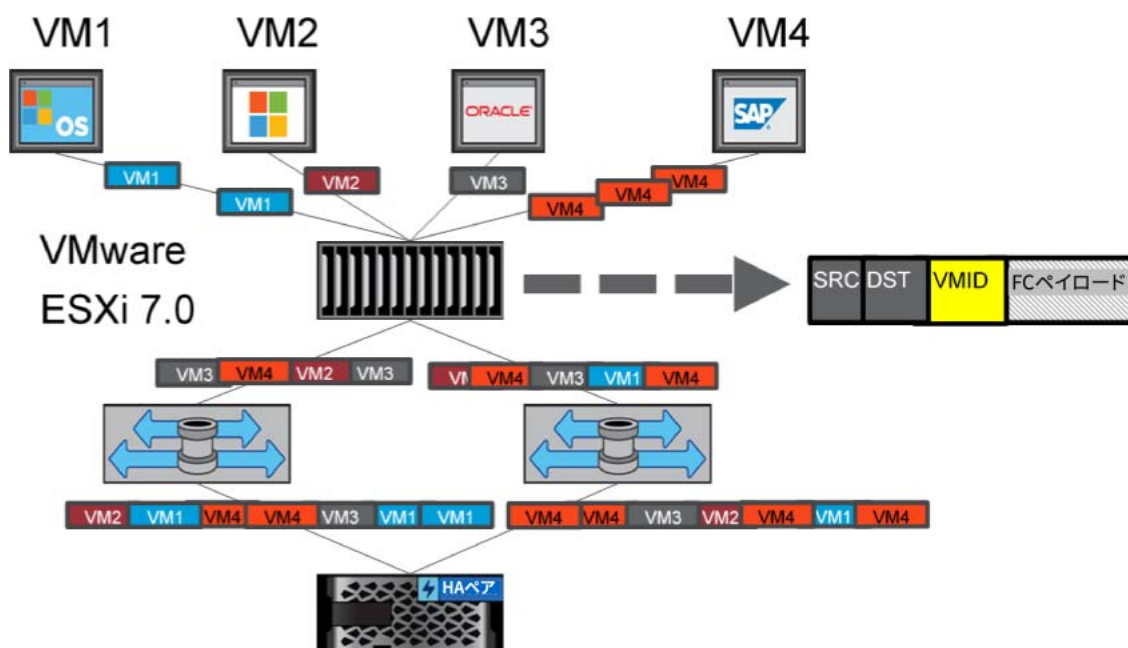
ASA で初めて提供されたその他の新機能は次のとおりです。

- 最大 LUN サイズの増加
- Persistent ports

6.1 仮想マシン ID (VMID)

VMware は、ホストしている各 VM のグローバル一意識別子を生成します。これらの情報はヘッダフィールドに書き込まれ、各 FC フレームは特定の VM に関連づけることができるようにタグ付けされます。これにより、管理者は共有 FC LUN によってバックアップされたデータストアを使用して、各 VM からの I/O を識別して追跡できます。VMID 機能の前は、可能な最高レベルの細分性は、I/O が関連づけられているデータストアを識別する機能でした。VMID を使用すると、管理者はデータストアを個別に共有する各 VM の I/O 特性を識別および追跡できます。この機能により、エンドツーエンドの QoS が可能になり、トラフィックパターン、ワークロード特性、大幅に強化されたトラブルシューティング、およびより詳細な VM トラフィックの分析とレポートについてより良い理解が可能になります。この機能は、Brocade スイッチ (Gen 6 以降) で最初にサポートされます。

図 6.1 VMID



FC ファブリックを介して共有データストア内の各 VM からの I/O を追跡するには、次の手順を実行します。

- 1 ハイパーバイザは、各 VM にグローバルに一意の ID を割り当てます。
- 2 VMID は、VM からの各フレームにタグ付けされます。
- 3 スイッチとストレージノードは、各フレームと VMID を伝達し、反映します。

6.2 NVMe-oF プロトコルの共存

ONTAP 9.8 は、NVMe/FC を自 SVM において他のブロックおよびファイルプロトコルから分離するという要件を削除しました。この削除は当初、NVMe/FC のリリースに必要な QA 回帰テストを減らして、NVMe/FC のリリースを高速化するために行われました。プロトコルを分離することで、エンジニアリング QA チームは、FCP、iSCSI、NFS、SMB、S3 などの他のプロトコルに対する NVMe-oF の影響をテストする必要がなくなるため、NVMe-oF プロトコルをリリースしました。プロトコル共存に必要な回帰試験が QA 試験計画に加えられ、さらに ONTAP 9.8 と ONTAP のその後の全リリースに対して実施されました。

6.3 より大きな LUN の最大サイズ

ONTAP 9.8 は、LUN の最大サイズを 16TB から 128TB に拡張しました。この増加に伴い、ボリュームのサイズも 100TB から 300TB に増加しています。これらのより大きな LUN の最も一般的な用途は、ハイパーバイザーデータストアをバックアップする LUN です。

6.4 Persistent ports

Persistent ports はテイクオーバーの影響を軽減します。これを行うには、HA パートナーの対応する物理ポートにシャドウ LIF を作成します。ノードが引き継がれると、対応するパートナーノード上のシャドウ LIF が WWPN を含む元の LIF の ID を引き継ぎます。

Persistent ports が動作するのは、以前のパスステータスを down (faulty) に変更する前に、ホスト MPIO スタックに、シャドウ LIF を active optimized (AO) としてアダプタイズできるためです。

ホスト MPIO スタックは、I/O の中断が最小限に抑えられるように、I/O を次の AO パス (かつてのシャドウ LIF) にシフトします。ターゲットの状態 (安定した状態またはテイクオーバー中) に関係なく、ホストから見てターゲットへのパスの数は常に同じになります。

Persistent ports は、ASA 上の ONTAP 9.8 で導入されています。ノードを ONTAP 9.8 にアップグレードすると、この機能はデフォルトで有効になります。

Persistent ports のベストプラクティスでは、FCP ポートの特性が HA ペア内で同一である必要があります。

- FCP ポート数
- FCP ポート名
- FCP ポート速度
- FCP LIF WWPN ベースのゾーニング
- アクティブな LIF とシャドウ LIF の両方がイニシエータと同じゾーンに存在する必要があります

これらのベストプラクティスのいずれかに違反すると、次の本文を含む EMS メッセージが生成されます。

```
EMS : scsiblade.lif.persistent.ports.fcp.init.error
```

Persistent ports 機能は FC では使用できますが、iSCSI では使用できません。WWPN は基本的にシャドウ LIF でもスプーフィングされるため、ゾーンのメンバーシップを WWPN で識別する必要があります。

図 6.2 に、Persistent ports を示します。

図 6.2 Persistent ports

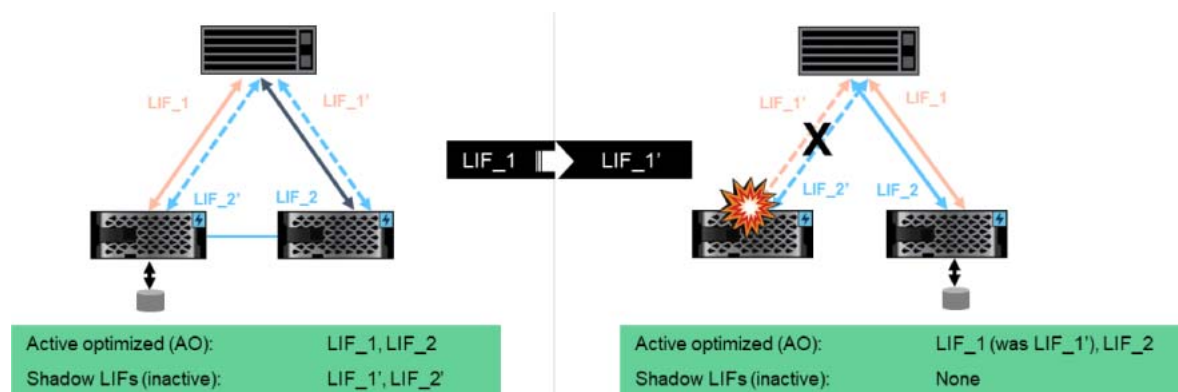
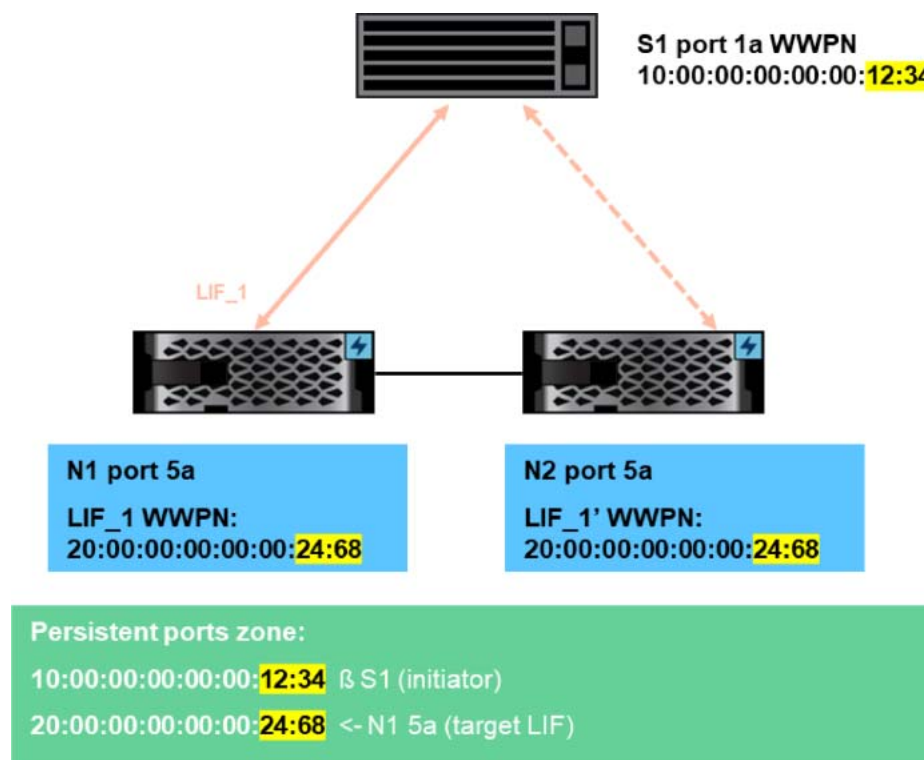


図 6.3 に、Persistent ports ゾーニングの例を示します。

図 6.3 Persistent ports ゾーニングの例



7. ONTAP 9.9.1 の新機能

ONTAP 9.9.1 では、ONTAP SAN に次のような多くの機能拡張が追加されています。

- System Manager
- SnapMirror Business Continuity (SMBC)
- ASA のクラスタの最大サイズは 1 組の HA ペアから 12 ノードに増加
- 単一 LUN のパフォーマンス向上
- ネストされた igroup
- NVMe/FC vVol のサポート
- NVMe リモート I/O のサポート
- ASA は NVMe/FC のサポートを追加

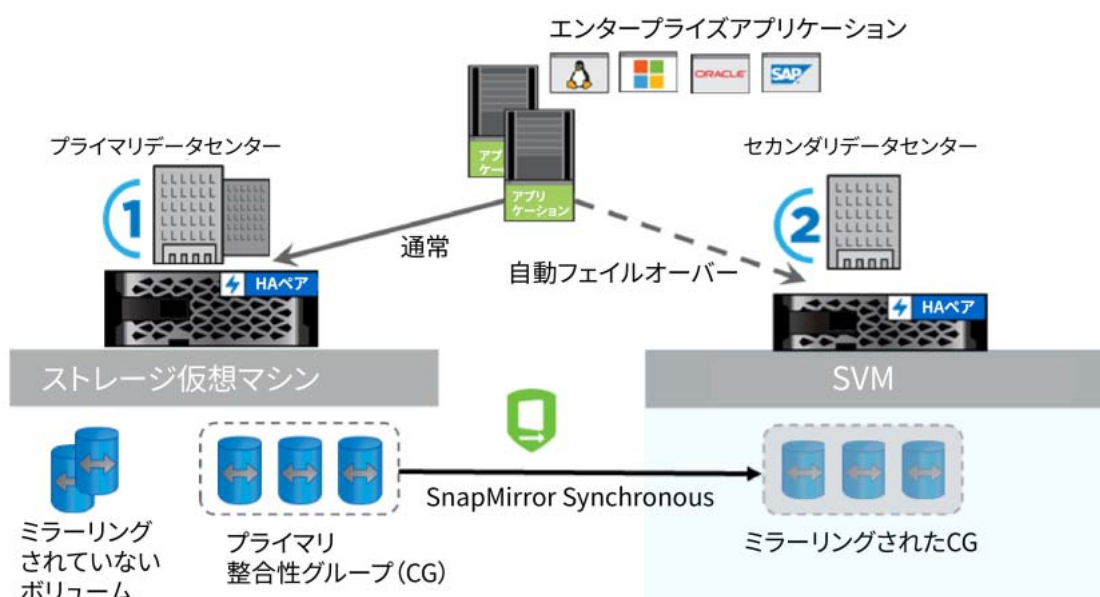
7.1 System Manager

System Manager は、実行されたすべてのアクションの Rest API 呼び出しを表示できるようになりました。この機能は、ワークフローのスクリプト作成や構文チェックを行う管理者を支援します。一般的に実行される操作が強化され、簡素化されます。たとえば、新しい LUN を 1 つの igroup にマッピングするように求められた場合、可能性のある igroup がすべて表示されます。または、1 つの igroup で WWPN を指定するように求められた場合、表示されている既存の WWPN がすべて表示され、選択できます。これにより、時間を節約し、入力ミスを減らすことができます。

7.2 SnapMirror Business Continuity (SMBC)

ONTAP 9.9.1 は、アプリケーションの整合性グループを使用してアプリケーションを同期的にレプリケートする SnapMirror Synchronous を使用する SMBC を導入して、2 つのサイト間ですべてのアプリケーションオブジェクトを管理およびレプリケートします。SMBC では、同期的に複製された 2 つのサイト間でフェイルオーバーを自動的に実行します。これにより、停止時間が短縮され、自動フェイルオーバーの管理と両方のミラーの保守に関連する管理コストが大幅に削減されます。[図 7.1](#) に SMBC のトポロジを示します。

図 7.1 SMBC のトポロジ



7.3 ASA のクラスタの最大サイズは 1 組の HA ペアから 12 ノードに増加

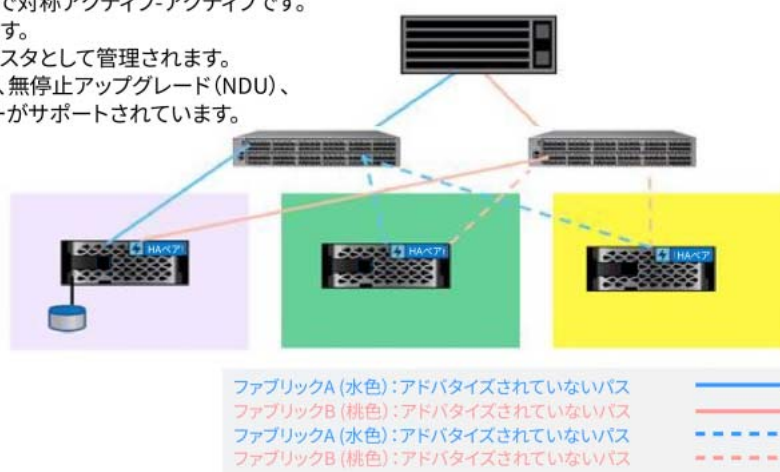
単一 HA ペア以上の構成では、クラスタ化されて 1 つの管理プランを共有する ASA HA ペアの認証連携が、基本的に ASA クラスタであることを理解することが重要です。このレイアウトでは、通常の無停止操作 (NDO) と ONTAP クラスタのその他の機能をすべて使用できます。ただし、これらのノードでサポートされている NAS プロトコルや機能、およびその他の ASA 固有の特性はありません。対称アクティブ - アクティブアクセスを提供することが ASA を定義する主な機能であるため、ASA をより大規模なクラスタに統合する場合にアクセスとパス設定がどのように機能するかを理解することが重要です。1 つの HA ペア内の ASA は両方のノードを経由するすべてのパスを AO または優先パスとしてアドバタイズするため、ホスト MPIO スタックはすべての AO パスを使用します。単一の HA ペアよりも大きい ASA クラスタの場合、各 HA ペアは、そのペアがホストする LUN へのすべてのパスを AO としてアドバタイズします。ただし、ホストの HA ペアの一部ではない他のコントローラを経由するパスは、最適化されていないアクティブパス (ANO パスまたは非優先パス) としてアドバタイズされます。これらのパスは、ホスト MPIO スタックでは使用されません。さらに、デフォルトでは、ONTAP には Selective LUN Map (SLM) と呼ばれる機能があり、ホストの HA ペア経由でのみパスをアドバタイズするため、SLM で追加パスをアドバタイズしない限り、ホストは最適化されていないパスや優先度の低いパスを認識しません。

SLM の詳細については、[「8.10 Selective LUN Mapping」 \(P.35\)](#) を参照してください。詳細については、[マニュアルサイト](#)に掲載の「ETERNUS AX/HX series ビジネス・クリティカルなワークロード向けのベスト・プラクティス」を参照してください。

図 7.2 12 ノード SAN

ONTAP 9.9.1で導入された12ノードのオールフラッシュSANアレイ (ASA)

- ASAのHAペアは、HAペア内で対称アクティブ-アクティブです。
- HAペア間のパスは非対称です。
- すべてのノードは単一のクラスタとして管理されます。
- すべての無停止操作 (NDO)、無停止アップグレード (NDU)、vol/LUN移動、およびコピーがサポートされています。



7.4 単一 LUN のパフォーマンス向上

ONTAP 9.9.1 では、単一 LUN のパフォーマンスが大幅に向上しています。これらの改善は主に、より多くの LUN I/O 操作を並列化して、より多くの処理を同時に実行できるようにすることによって実現されています。パフォーマンス向上の度合いは、メモリ /CPU コントローラが大きいほど大きくなる傾向があります。単一の LUN のパフォーマンスに対するメリットの大部分は、次の領域で最も大きくなると予想されます。

- LUN によってデータストアがバックアップされる仮想化。

多くの場合、ほとんどの LUN は論理ボリュームマネージャやその他のアプリケーションのアグリゲーションを使用して分離されるため、単一の LUN のパフォーマンス向上は多くのお客様には影響しません。一般的な使用パターンである多数の LUN が使用されている場合、単一の LUN のパフォーマンスは全体のパフォーマンスに影響しません。

7.5 ネストされた igroup

ONTAP 9.9.1 では、LUN マスキングを簡素化するために、ネストされた igroup が追加されています。ネストされた igroup では、既存の igroup を新しい igroup に追加できます。この機能により、どのイニシエータがどの LUN にアクセスできるかをより簡単に定義できます。ユーザーは、ストレージ管理者にとって意味のある igroup 名やエイリアスをより創造的に使用できるようになります。その他、次のような igroup の機能強化が行われました。

- igroup とイニシエータにコメントが割り当てられている可能性があります。
- LUN の igroup のイニシエータには (イニシエータを表示するための新しい CLI 専用コマンド) が表示されます。

次の例では、コメントを追加し、新しいコメントでイニシエータを示しています。

```
tme-a700s-clus:> lun igroup initiator modify -initiator 10:00:00:10:9b:34:9f:34 -comment "This is a comment about 10:00:00:10:9b:34:9f:34"
```

```
tme-a700s-clus:> lun igroup initiator show
```

Initiator	Comment	Vserver
svm0	10:00:00:10:9b:34:9f:34	This is a comment about 10:00:00:10:9b:34:9f:34
svm0	10:00:00:10:9b:34:9f:35	-
svm0	10:00:00:90:fa:d1:ea:f7	-
svm0	10:00:00:90:fa:d1:ea:f8	-

4 entries were displayed.

igroup には、最大 3 レベルのネストを含めることができます。

- 親要素の igroup
- 子要素の igroup
- 孫要素の igroup

ネストされた igroup に関するその他の注意事項は次のとおりです。

- オペレーティングシステムとプロトコルは一致している必要があり、変更されていない可能性があります。
- `allow_delete_while_mapped = True` は、`lun:igroup_nested_delete`, `igroup_delete` にマッピングされている場合、igroup のネストを解除して削除します。
- 孫要素をネスト解除したり削除したりすることは、その親のいずれかがマッピングされている場合できません。
- 孫要素がマップされている場合は、子要素をネスト解除または削除できます。
- ネストされた関係内の igroup を削除 (`igroup delete`) することができ、削除されていない igroup は維持されます (マップされていない場合)。

7.6 NVMe/FC vVol のサポート

ONTAP 9.9.1 では、NVMe/FC vVol のサポートが追加されました。この拡張により、VMware 管理では、vSphere が vVol/vCenter 内でそのオプションをサポートした後に、vVol を使用してストレージを管理および自動化できます。

7.7 NVMe リモート I/O のサポート

ONTAP 9.9.1 では、NVMe-oF にリモート I/O サポートが追加されています。これにより、NVMe-oF パス設定が、アクティブ - 非アクティブモデルから、他のすべての ONTAP ブロックプロトコルが使用する AO/ANO モデルに変更されます。

図 7.3 は、リモート I/O サポートのない NVMe-oF を示しています。

図 7.3 リモート I/O サポートなしの NVMe-oF

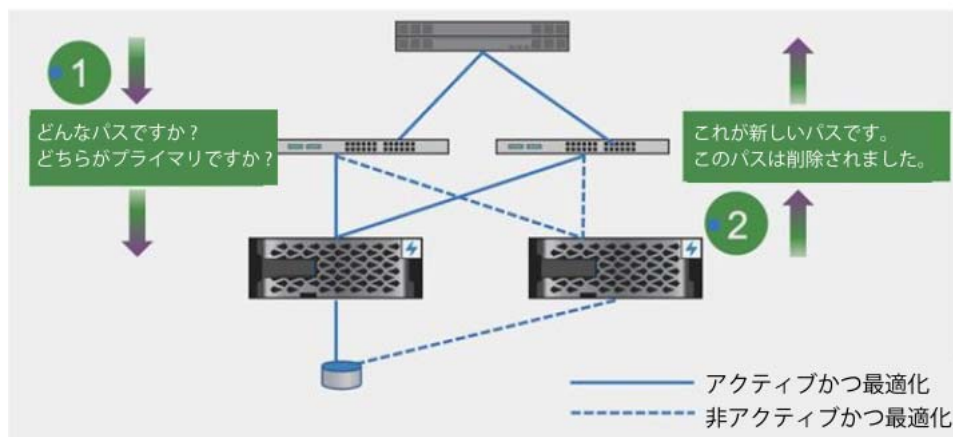


図 7.4 リモート I/O サポートありの NVMe-oF

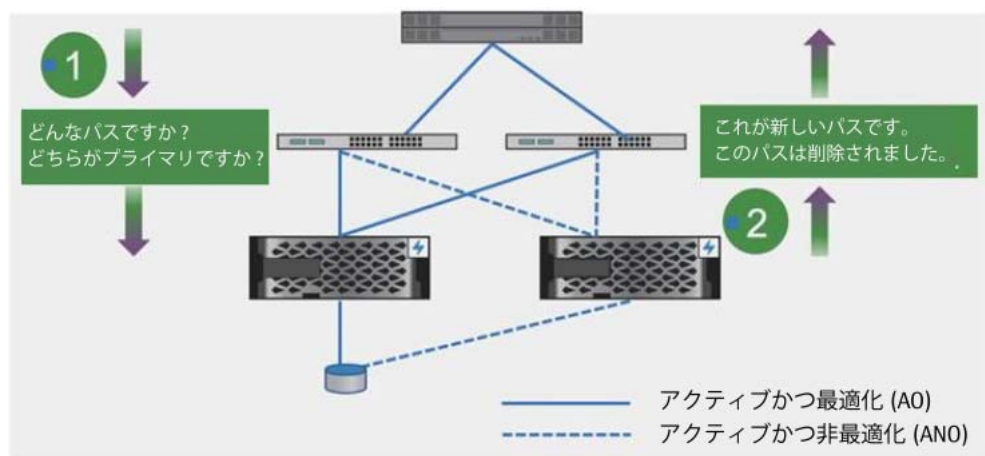


図 7.3 と図 7.4 を比較しても、大きな違いはないように思われます。NVMe-oF でリモート I/O がサポートされている場合、すべてのパスがアクティブになります。つまり、これらのパスのいずれかに送信された I/O が確認され、応答または応答されます。以前は、リモート I/O がないと、非アクティブなパスは使用できませんでした。

7.8 ASA では NVMe/FC のサポートを追加

ONTAP 9.9.1 では、NVMe/FC が追加のブロックプロトコルとして追加されています。FC や iSCSI とは異なり、ASA 上の NVMe/FC は引き続き非対称です (AO/ANO)。これは、NVMe-oF がリモートパスとローカルパスでどのように動作するかが異なるためです。

8. ONTAP プロトコルと SAN プロトコル

8.1 ONTAP の概要

ONTAP を実行するストレージコントローラは、ノードと呼ばれます。これらのノードは、クラスタ化されたシステムに集約されます。クラスタ内のノードは相互に継続的に通信し、クラスタの活動を調整します。また、専用クラスタネットワークへの冗長パスを使用することで、ノード間でデータを透過的に移動します。

クラスタの基本単位はノードですが、ノードは HA ペアの一部としてクラスタに追加されます。HA ペアは、HA インターコネクト (専用のクラスタネットワークとは別) を介して相互に通信し、HA ペアのディスクへの冗長接続を維持することにより、高可用性を実現します。シェルフには HA ペアのいずれかのメンバーに属するディスクを含めることができますが、HA ペア間でディスクを共有することはできません。

クラスタはノード単位ではなくクラスタ全体で管理され、データは 1 つ以上の SVM から提供されます。各 SVM は、物理アグリゲートからプロビジョニングされるボリューム (および LUN) の形式でストレージを所有するように構成され、LIF は、物理 Ethernet ネットワークまたは FC ターゲットポートのいずれかに割り当てられます。LUN は SVM のボリューム内に作成され、ホストにマッピングされてストレージ領域を提供します。SVM はノードに依存せず、クラスタベースです。ボリュームやネットワークポートなどの物理リソースをクラスタ内の任意の場所で使用できます。

8.2 SAN のパフォーマンスを最適化するための考慮事項

ONTAP は、できるだけ多くの作業を同時に処理するために、複数のプロセッサコアを使用するように最適化されています。多くの ONTAP 操作は、使用可能な複数のプロセッサコアに分散できます。ただし、複数のプロセッサコアに分割できない一部の操作では、達成可能な最大パフォーマンスが低下する可能性があります。これらの非分散スレッドは、ワークロードが複数のオブジェクトに適切に分割されている場合に達成できる最大パフォーマンスを制限する可能性があります。より多くのオブジェクトを使用することで、特定のワークロードに対してプロビジョニングされるボリュームと LUN の数を増やすことで、使用可能なすべてのコアを使用してパフォーマンスを最適化できます。

8.2.1 ボリューム

特定のアプリケーションにプロビジョニングするボリュームの数を検討する場合、まず SAN のコンテキストでボリュームをどのように使用するかを検討することが重要です。考慮するストレージベンダーによっては、ボリュームという言葉にさまざまな意味を持たせることができます。ONTAP のコンテキストでボリュームについて説明する場合、ONTAP SAN のコンテキストでどのボリュームが使用されるかを理解することが重要です。ONTAP では、ボリュームは次の機能を提供します。

- ホストする LUN の管理コンテナ。これは、ボリュームが複数の LUN を保持している場合に便利です。これにより、ボリューム内でホストされる複数の LUN の管理が容易になります。
- Snapshot コピーはボリュームレベルで作成されるため、ボリューム内のすべてのブロックがキャプチャされるため、ONTAP ボリュームを整合性グループとして使用できます。つまり、同じボリュームでホストされている複数の LUN がすべて同時に Snapshot コピーを作成します。これにより、LUN グループ全体のワークロードの整合性が大幅に低下します。
- Storage Efficiency 機能は、主にボリュームレベルで実行されます。データセットが大きいほど、共通ブロックが見つかる可能性が高くなることを意味します。つまり、一般的にボリュームが大きいほど、ストレージ効率が向上します。

これまで、サイズの大きいボリュームの数が少ないほど管理が容易で、整合性グループとして使用でき、ストレージ効率が向上することを説明してきましたが、パフォーマンスを最適化するためにプロビジョニングするボリュームの数についても考慮する必要があります。

前述したように、一部の処理スレッドは複数のプロセッサコアに分散できません。このように作業を複数のコアに分散できない場合は、ワークロードを複数のボリュームに分散することで管理できます。これにより、1つのコアで実行する必要がある分散不可能な作業の量を分割できます。各ボリュームには独自の非分散スレッドがあるため、作業を分散するボリュームが多いほど、これらの非分散スレッドの多くを複数のプロセッサコアで同時に処理できるため、スループットが向上します。

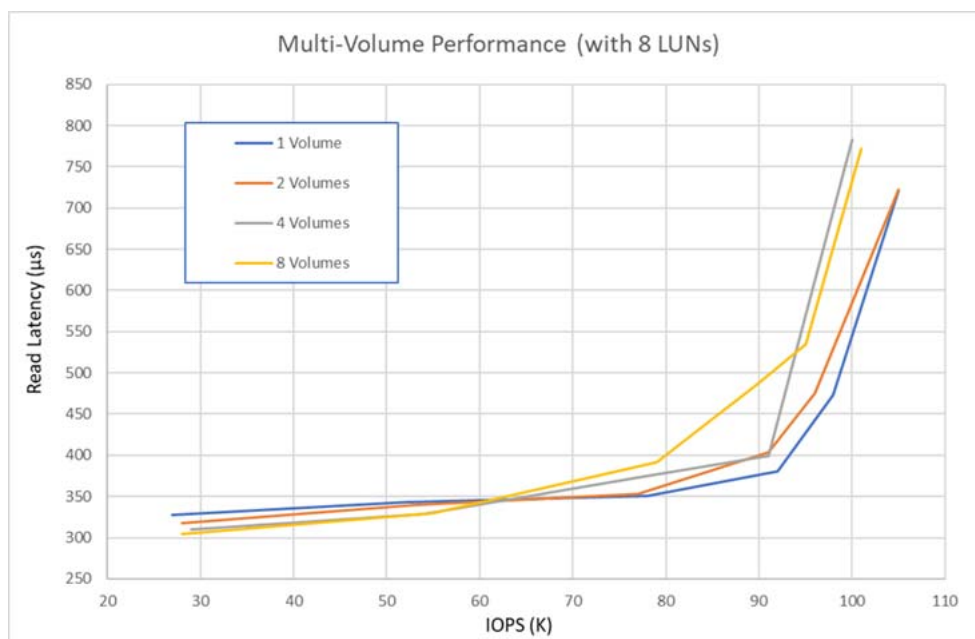
つまり、同じボリューム内の多数の LUN をグループ化する場合と、ボリューム数を増やす場合とでは、バランスをとる必要があります。一般的に、多くのテストと長年にわたるカスタマーエクスペリエンスから、ボリューム数をある程度まで増やすことでパフォーマンスの最適化がほとんどの場合で可能になるガイドラインを提供してきました。その後、ボリュームを増やしてもパフォーマンスは向上しませんが、複雑さが増し、ストレージ効率が低下することが判明しました。

4 ～ 16 個のボリュームを使用することを推奨しており、理想的なボリューム数は 8 ～ 16 個です。

備考

図 8.1 と図 8.2 は、ボリュームと LUN を追加した結果の一例です。特定のワークロードを保証またはベンチマークしたり、パフォーマンスの推定値を提供したりすることを目的としていません。

図 8.1 作業をより多くのボリュームに分散する効果



ベストプラクティス

- 多くの場合、パフォーマンスを最大化するために 8 ～ 16 個のボリュームを使用することを推奨しています。この推奨事項は、特定のコントローラ上で他にボリュームが存在していないことが前提です。他のボリュームが存在する場合、通常は最大 8 個の LUN を追加してパフォーマンスを最大にすることを検討できます。
- 相互に関連し、同様のパフォーマンスと管理要件を持つ LUN は、1 つのボリュームでホストできます。同じボリュームを使用することで、次のメリットを実現できます。
 - 共通の管理コンテナを使用することで、管理の複雑さを軽減します。
 - Snapshot コピーとデータ保護、または Snapshot コピーを基盤とするレプリケーションソリューションは、ボリュームレベルで管理されます。特定のアプリケーション (または特定のホスト) がホストしているすべての LUN が共通のボリュームによってホストされている場合、そのボリュームは整合性グループとして効果的に機能できます。Snapshot コピーのコンテキストはボリュームであり、そのボリュームがホストするすべてのアイテムがすべての Snapshot コピーによってキャプチャされます。
 - Storage Efficiency では、ボリュームを編成コンテナとして使用し、すべての Storage Efficiency オブジェクトとメタデータはボリュームレベルで保存されます。したがって、ボリュームに含まれる LUN の数が多いほど、効率性 (ブロックの重複排除、圧縮、コンパクションなど) が向上します。
- 同じボリューム内の複数の LUN を組み合わせても効果がないことがあります。しかし、関連性があり、同様のパフォーマンス要件を持ち、共通の整合性グループに含めることでメリットが得られる LUN の組み合わせの場合は、グループ化する必要があります。

8.3 LUN

LUN (iSCSI または FC など) には、複数のプロセッサコアに分散できないスレッドがあります。したがって、パフォーマンスを第一に考慮する場合は、サイズを大きくして LUN の数を減らすのではなく、サイズの小さい LUN の数を増やすことを推奨しています。ボリュームと同様に、ワークロードのパフォーマンスは、複数の LUN に作業を分散させることで最適化できます。これにより、より多くのプロセッサコアを使用してワークロード I/O を同時に処理できます。

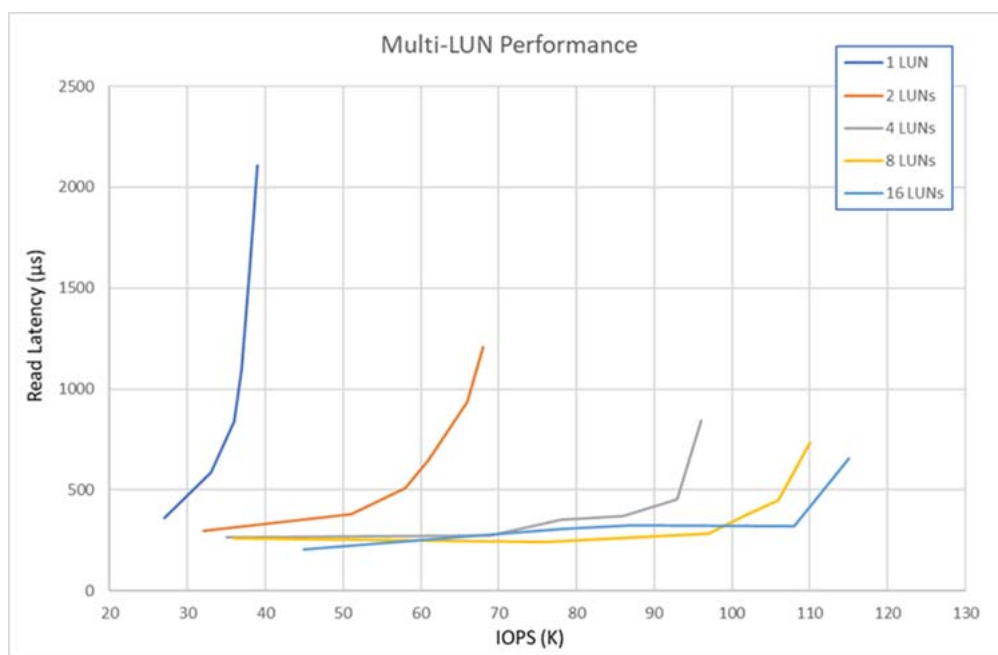
アプリケーションが使用する LUN の数を増やすために使用できるアプローチには、次のようなものがあります。

- 論理ボリュームマネージャ (LVM) は、複数の LUN を 1 つのボリュームに結合し、ホストの OS またはアプリケーションに提供します。LVM は通常、Linux および UNIX OS で使用されます。Oracle ASM では、複数の LUN を 1 つのストレージオブジェクトとして集約することもできます。
- 基本的に OS またはアプリケーションが提供されている LUN を管理している場合、複数の LUN の制御情報を OS またはアプリケーションに伝達することが有効なケースがあります。

ベストプラクティス

サイズを大きくして LUN の数を減らすよりも、サイズの小さい LUN の数を増やすことを推奨しています。8 ～ 16 個の LUN が理想的です。

図 8.2 複数の LUN に作業を分散する効果



8.3.1 FC インオーダー配信

FC スイッチは、インオーダー配信 (IOD) を保証するように構成する必要があります。この手順は ONTAP 操作には必要ありませんが、フレームの順序が正しくない場合やドロップフレームが発生した場合に、ONTAP ではやりとりが中断されるため、ベストプラクティスです。そのため、ONTAP は、イニシエータが SCSI タイムアウト閾値に達したときにフレームを再送信するようにイニシエータ (ホスト) に通知する必要があります。このプロセスには 60 秒かかる場合があります。ONTAP はこの状況を切り抜け、回復しますが、SCSI タイムアウトと再送信時間が原因でレイテンシの悪化が発生します。

ファブリック内のすべての FC スイッチで IOD が設定されている場合、ONTAP は順序が正しくないフレームを受信しないため、フレームの再送信を待つような長いホスト SCSI のタイムアウトには耐えられません。

8.3.2 結論

ボリュームを追加しても、パフォーマンス上のメリットはほとんどありません。どんな変化も誤差の範囲にすぎません。図 8.1 と図 8.2 は、8 個の LUN を持つ 1 つのボリュームが、良好なレイテンシで容易な 100K のランダム IOPS を提供していることを示しています。ここから推測できるのは、特定のアプリケーションで使われる LUN の数を増やすことで、アプリケーションのパフォーマンスを向上できるということです。ボリュームの数を増やすことでパフォーマンスが向上しますが、LUN の数を増やすことに比べてパフォーマンスの向上ははるかに小さくなります。どちらの場合も、ボリューム数を増やすことによるリターンは小さくなる傾向があります。さらに、ボリューム数を増やすと、スケールに対するリターンが減少する可能性があります。

備考

まれに、ワークロードを複数のボリュームに分散させて性能を向上させることができますが、これは主に、1 つのアプリケーションがコントローラのすべての性能を消費している場合に当てはまります。たとえば、500K の IOPS を送信する必要がある大規模なデータベースが 1 つあり、レイテンシをマイクロ秒単位で最小化したい場合は、複数のボリュームが必要です。このような状況では、使用中のボリューム数だけでなく、構成のすべての側面を考慮する必要があります。

上に示したデータに基づいて、次の推論を行うことができます。

- 作業を複数の LUN に分散させると、パフォーマンスが大幅に向上します。
- 1 つの LUN で約 35K IOPS を達成できます。2 つの LUN を使用すると、ほぼ 2 倍の約 70K IOPS になります。
- LUN が 8 個に達すると、メリットは減少し始めます。4 個の LUN を使用することも可能ですが、8 個の方が若干適しています。
- 8 個の LUN を持つ 1 つのボリュームは、容易な 100K のランダム IOPS を適切なレイテンシで提供します。これは、一般的なデータベースの 99% が必要とする I/O よりも多くなります。

備考

図 8.1 および図 8.2 は、前述の結論における数値が特定の試験から得られたものであることを示しています。結論は妥当であり、ベストプラクティスの推奨事項を示しています。ただし、概念と推奨事項を説明するために、具体的な数値を掲載しています。特定のボリューム、LUN、アプリケーションがパフォーマンスに関して達成できることを保証またはガイドラインとして捉えるべきではありません。

8.4 ボリューム構成

クラスタでボリュームをプロビジョニングする場合、重複排除、スペースリザベーション、ストレージ効率に関する考慮事項の多くは同じです。主な違いの 1 つは、ONTAP ストレージクラスタ上のボリュームが、個々のノードではなく SVM に向けられることです。また、それに付随して、NFS または CIFS プロトコルを使用してファイルシステムをエクスポートする目的で、SVM 全体のグローバルネームスペースにボリュームをマッピングできます。ただし、グローバルネームスペース内に特定のボリュームがあるかどうかは、FC または iSCSI などを使用して提供されるデータには影響しません。

ベストプラクティス

LUN を含むボリュームは、ブロックプロトコルを使用してデータを提供するためにグローバルネームスペースに連結する必要はありません。igroup から LUN へのマッピングのみが必要です。

8.5 ホスト接続

ブロックプロトコルを使用して ONTAP ストレージクラスタが提供するデータにアクセスするホストは、SCSI プロトコルに対する ALUA 拡張機能を使用して、どのパスが特定の LUN に対して直接で、どのパスが間接であるかを判断する必要があります。ALUA 標準では、直接パスをアクティブ / 最適化パス、間接パスをアクティブ / 非最適化パスと呼びます。すべての ALUA 情報は、データに使用されるものと同じ iSCSI 接続または FC 接続などを使用して、帯域内で要求および配信されます。

特定のパスのステータスは、ホストが特定の LUN について検出した各パスにパスステータス照会を送信することによって検出可能になります。このパスステータスの照会は、ストレージシステムが SCSI 要求の結果とともに追加データを送信し、パスのステータスが更新されたこと、およびパスの優先順位を再検出する必要があることをホストに通知したときにトリガーされます。

ALUA はよく知られた広く展開された標準であり、ONTAP によって提供されるブロックデータへのアクセスのための要件である。ONTAP ブロックアクセスプロトコルで動作することがテストされ、認定されているすべてのオペレーティングシステムが ALUA をサポートします。

8.6 パス選択

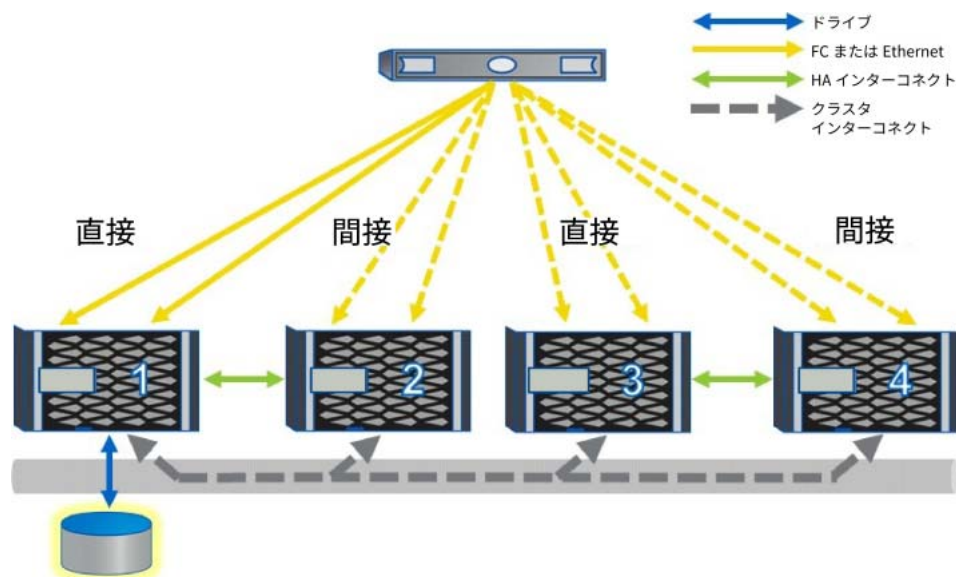
SVM が所有するすべての LIF は LUN に対する書き込み要求と読み取り要求を受け入れますが、その LUN をバックアップするディスクを実際に所有しているクラスタノードは常に 1 つだけです。これにより、LUN への使用可能なパスが、直接パスと間接パスの 2 つのタイプに効果的に分割されます。

LUN の直接パスは、SVM の LIF とアクセス対象の LUN が同じノード上に存在するパスです。物理ターゲットポートからディスクに移動するために、クラスタネットワークを横断する必要はありません。

図 8.3 は、直接パス上の 1 というラベルが付いたノードが所有する LUN 上のデータにアクセスするホストを示しています。LUN はストレージアグリゲート上にあるため、このノードへのパスはすべて直接パスです。SAN プロトコルのセットアップでは、LUN への複数の直接パスを持つことが一般的です。冗長性およびデータアクセスの耐障害性を目的として、第 2 のパスは、通常、別個の Ethernet ネットワークまたは FC ファブリック上にあり、スループット目的でネットワークまたはファブリックごとに追加のパスを使用できます。

ALUA を使用すると、ホストは間接パスに依存する前に、利用可能な任意の直接パスでトラフィックを転送できます。したがって、障害が発生していない状況で間接パスを使用することはほとんどありません。

図 8.3 ONTAP でのパスの概要



間接パスは、SVM の LIF とアクセス対象の LUN が異なるノード上に存在するデータパスです。データは、物理ターゲットポートからディスクに移動するためにクラスタネットワークを横断する必要があります。クラスタネットワークは高速で高可用性であるため、ラウンドトリップに大きなレイテンシを発生させることはありませんが、これは最大効率のデータパスではありません。適切に構成された SAN 環境では、ホストによる間接パスの使用は最小限に抑えられます。

すべてのホストは、クラスタ内の物理リソースを使用する SVM とのみ通信するため、実際には、クラスタへのすべての接続は、LUN にアクセスするホスト上で実行されている MPIO ソフトウェアによって管理され、その結果、通常の操作では直接パスのみが使用されます。

ベストプラクティス

すべての SVM には、各クラスタノードおよび各 FC ファブリックまたは Ethernet ネットワーク上で LIF を割り当てる必要があります。たとえば、4 ノードのクラスタが 3a および 4a の FC ターゲットポートを使用して A および B という 2 つの独立した FC ファブリックに接続されている場合、FC を使用してデータを提供する SVM には、node1:3a、node1:4a、node2:3a、node2:4a、node3:3a、node3:4a、node4:3a、および node4:4a の 8 つの LIF が必要です。4 つ以上のノードを持つクラスタでは、管理を容易にし、オペレーティングシステムのパス数の制限を考慮して、特定の LUN へのアクセスに使用するパスの数を制限する必要があります。詳細は、「[8.9 パス管理と Selective LUN Mapping](#)」(P.34) を参照してください。

NFS や CIFS などの NAS プロトコルで ONTAP ストレージクラスタを使用することに慣れている管理者の場合、これらのプロトコルを提供する LIF と、ブロック iSCSI または FC を提供する LIF には区別があります。NAS LIF はノード間で自由に移動できます。また、NAS LIF は、HA またはポートフェイルオーバー中に移動するノードおよびポートを指定するフェイルオーバーグループに属することもできます。これに対し、SAN LIF は、SCSI イニシエータと SCSI ターゲットの間で同時に確立される複数のパスのエンドポイントを表し、ホストの MPIO ソフトウェアが実際に I/O を受信するパスを管理します。その結果、NAS LIF とは異なり、SAN LIF はフェイルオーバーしません。SAN のフェイルオーバーメカニズムでは、複数のパスをプロビジョニングし、ホストでマルチパス (MPIO) ソフトウェアを使用して、ホストに提供される複数のパスを管理します。

このように動作が異なるため、iSCSI プロトコルを使用してデータを提供する Ethernet LIF は、NAS プロトコルを使用してデータを提供することもできません。

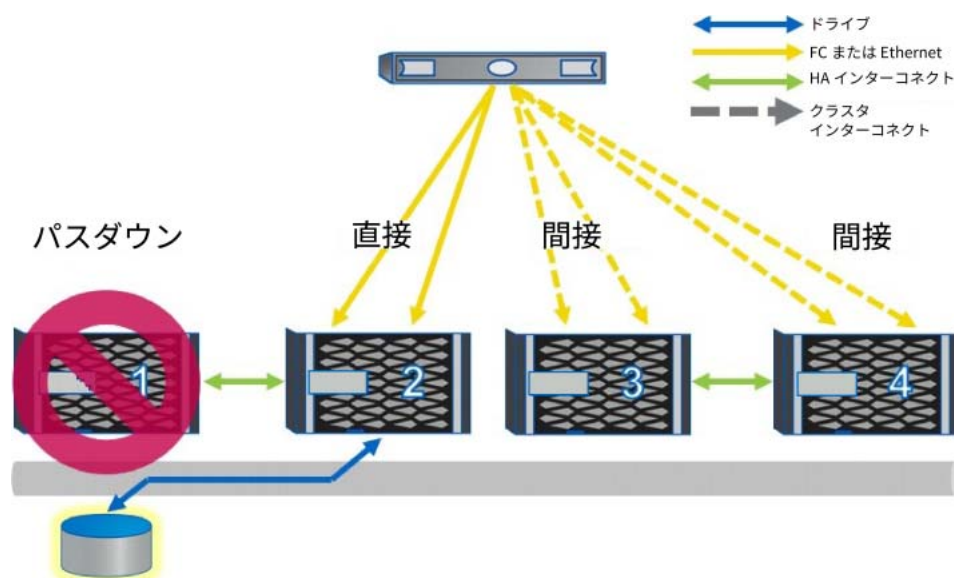
8.7 パス選択の変更

ホストがクラスタ上のデータにアクセスするために選択したパスを変更する可能性のあるイベントは3つあります。

■ HA フェイルオーバー

HA フェイルオーバーイベントでは、ダウンノード上の LIF はオフラインとして表示され、ダウンノードを引き継いだ HA パートナー上の LIF は直接パスになります。この状態は、ALUA パスの照会によって自動的に変化し、管理上の変更は必要ありません。[図 8.4](#) に HA フェイルオーバー中のパスを示します。

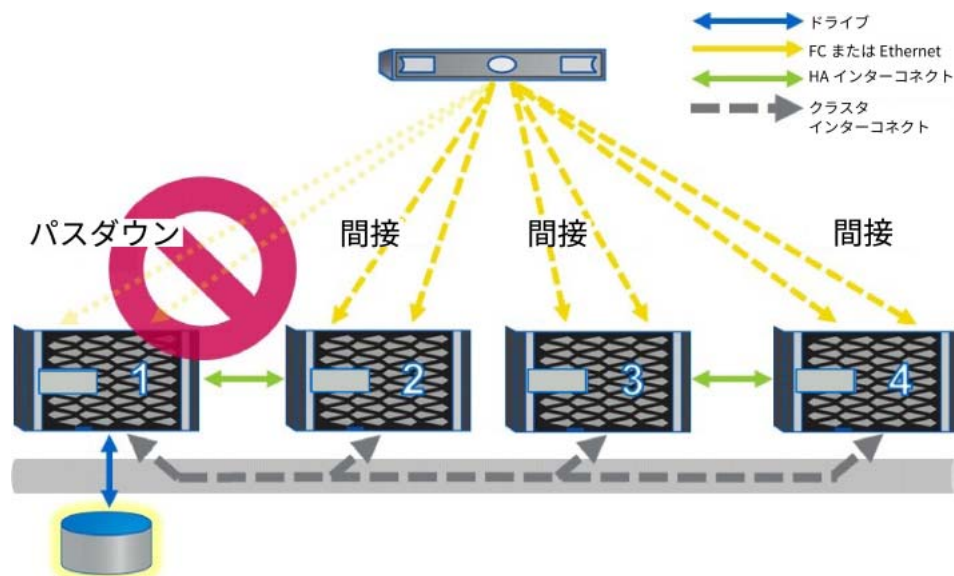
図 8.4 HA フェイルオーバー中のパス



■ ポートまたはスイッチの障害

ポートまたはスイッチの障害時には、これ以上直接パスは使用できません。パスの優先順位は変わりません。ホスト上で実行されている MPIO ソフトウェアは、直接パスが再び使用可能になるまで、代替間接パスを選択します。ALUA パスの状態は変化しません。[図 8.5](#) に、ポートまたはスイッチの障害時のパスを示します。

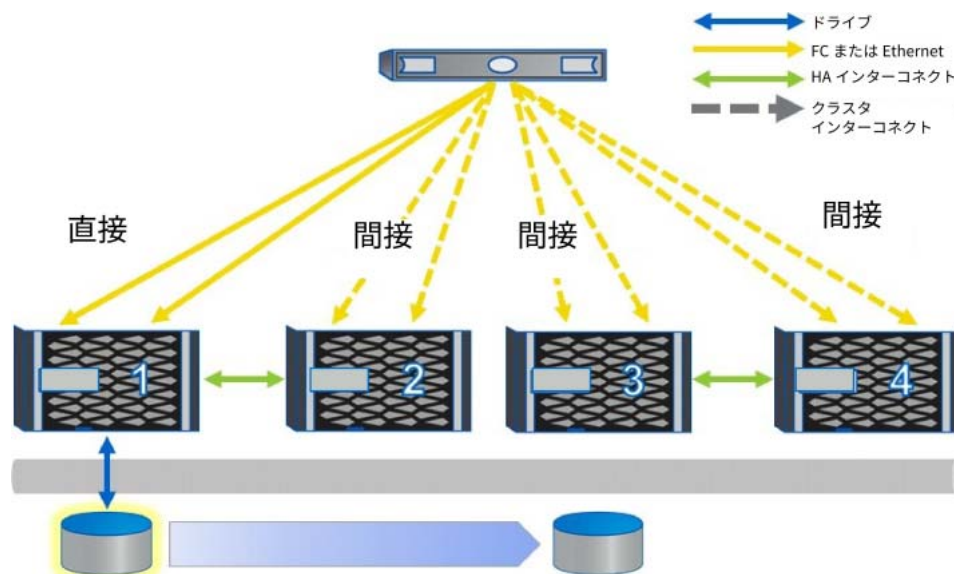
図 8.5 ポートまたはスイッチの障害時のパス



■ ボリュームまたは LUN の移動

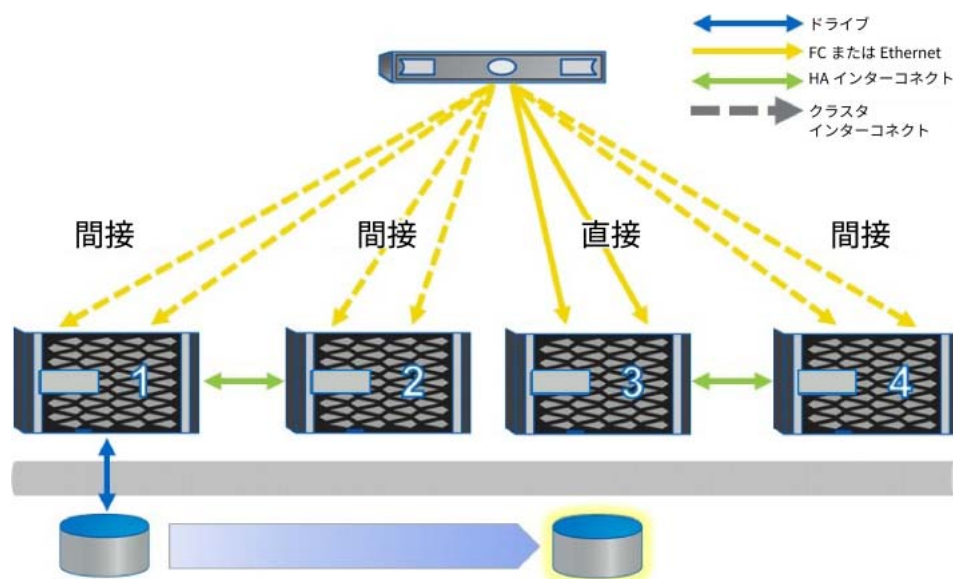
ボリュームは、volume move 機能を使用してノード間で透過的に移動され、LUN は lun move によってノード間で透過的に移動されます。[図 8.6](#) は、ボリュームまたは LUN の移動中のパスを示しています。

図 8.6 ボリュームまたは LUN の移動中のパス



ボリューム移動では、カットオーバーが発生し、ボリュームの新しいノードが読み取りおよび書き込み要求の処理を開始すると、新しいノードに直接パスがあり、古いノードに間接パスがあるように、パスステータスが更新されます。すべてのパスは常に使用可能です。図 8.7 に、ボリュームまたは LUN の移動後のパスを示します。

図 8.7 ボリュームまたは LUN の移動後のパス



LUN 移動の場合、すべてのデータがターゲットに転送される前にカットオーバーが発生し、読み取り要求はクラスターネットワークを介してソースノードに渡されて処理されます。LUN 移動機能の動作の詳細については、[「8.13 DataMotion for LUNs」 \(P.38\)](#) を参照してください。

8.8 FC と NPIV

ONTAP ノードでは、HA ペアの物理 FC ターゲットアダプタのアドレスに基づいて単一のワールドワイドノード名 (WWNN) と関連する WWPN を使用するのではなく、NPIV を使用して、すべての論理インターフェースが独自の WWPN を使用して FC ファブリックにログインできるようにします。これにより、同じ FC ファブリックスイッチに接続されているホストは、どの物理ノードがどの LIF を所有しているかに関係なく、同じ SCSI ターゲットと通信できます。仮想ポートは SCSI ターゲットサービスを提供し、データを送受信します。

ベストプラクティス

FC LIF が正しく動作するためには NPIV 対応の FC スイッチが必要です。FC LIF を作成する前に、ONTAP システムに接続されているファブリックスイッチで NPIV が有効になっていることを確認してください。

Cisco NX-OS を使用している場合、NPIV のステータスは `show npiv status` コマンドを実行して確認できます。

```
N5K-A# show npiv status
NPIV is enabled
```

Brocade FabOS を使用している場合、portcfgshow コマンドは NPIV の性能とステータスを表示します。

BRCD_8K:admin> portcfgshow																
Ports of Slot 0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Speed	AN	AN	AN	AN	AN	AN	AN	AN	10	10	10	10	10	10	10	10
Fill Word	0	0	0	0	0	0	0	0	-	-	-	-	-	-	-	-
AL_PA Offset 13	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Trunk Port	ON	ON	ON	ON	ON	ON	ON	ON	-	-	-	-	-	-	-	-
Long Distance	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
VC Link Init	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Locked L_Port	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Locked G_Port	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Disabled E_Port	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Locked E_Port	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
ISL R_RDY Mode	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
RSCN Suppressed	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Persistent Disable	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
LOS TOV enable	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
NPIV capability	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON	ON
NPIV PP Limit	126	126	126	126	126	126	126	126	126	126	126	126	126	126	126	126
QOS E_Port	AE	AE	AE	AE	AE	AE	AE	AE	..	..	..	..	..	..	..	..
EX Port	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Mirror Port	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Rate Limit	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Fport Buffers	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Port Auto Disable	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
CSTL mode	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..

ストレージ管理コンソールからは、接続されているスイッチの NPIV ステータスについて直接確認することはできません。ただし、ローカル FC トポロジを調べることで、ファブリックスイッチポートで NPIV が有効になっているかどうかを確認できます。次の例では、ポート 2/1 に複数の WWPN が接続されており、その一部が仮想ポートであるため、NPIV を有効にする必要があります。

cluster::> node run -node node01 fcp topology show						
Switch Name: N5K-A						
Switch Vendor: Cisco Systems, Inc.						
Switch Release: 5.0(2)N2(1a)						
Switch Domain: 200						
Switch WWN: 20:66:00:0d:ec:b4:94:01						
Port	Port WWPN	State	Type	Attached WWPN	Port ID	
2/1	20:01:00:0d:ec:b4:94:3f	Online	F-Port	50:0a:09:83:8d:4d:bf:f1	0xc80033	
				20:1c:00:a0:98:16:54:8c	0xc80052*	
				20:0e:00:a0:98:16:54:8c	0xc80034*	
				20:10:00:a0:98:16:54:8c	0xc8003f	
2/3	20:03:00:0d:ec:b4:94:3f	Online	F-Port	50:0a:09:83:8d:3d:c0:1c	0xc8002c	

ベストプラクティス

物理 WWPN (50:0a:09:8x から始まる) は SCSI ターゲットサービスを提供しないため、ファブリックにログインしているように見えますが、FC ファブリックのどのゾーン構成にも含めないでください。

これらの WWPN を表示するには、fcp adapter show -fields fc-wwpn コマンドを使用するか、System Manager の [ネットワーク] - [FC ポート] で表示される [FC/FCoE Adapters] ペインを使用します ([図 8.8](#) を参照)。

代わりに、network interface show コマンドの出力結果と、System Manager の [ネットワーク] -

[イーサネットポート] ペイン ([図 8.9](#) を参照) に表示される仮想 WWPN (20: で始まる WWPN) のみを使用してください。

図 8.8 System Manager の FC アダプタ

ノード	0c	0d	0e	0f
H02200-01	10 Gbps	10 Gbps	10 Gbps	10 Gbps
有効化 / 無効化	☒	☒	☒	☒
WWPN	50:0a:09:83:80:83:c9:7d	50:0a:09:84:80:83:c9:7d	50:0a:09:81:80:83:c9:7d	50:0a:09:82:80:83:c9:7d
ネットワークインターフェイス	0	0	0	0
データリンク速度	0	0	10 Gbps	0
ポートアドレス	0	0	10000	0
プロトコル	FC	FC	FC	FC
スループット (MB/s)	0 MB/s	0 MB/s	0 MB/s	0 MB/s
最近の変更	18:35:59			
H02200-02	10 Gbps	10 Gbps	10 Gbps	10 Gbps
有効化 / 無効化	☒	☒	☒	☒
WWPN	50:0a:09:83:80:83:c9:13	50:0a:09:84:80:83:c9:13	50:0a:09:81:80:83:c9:13	50:0a:09:82:80:83:c9:13
ネットワークインターフェイス	0	0	0	0
データリンク速度	0	0	10 Gbps	0

図 8.9 System Manager のネットワークインターフェース

ノード	eth0	eth1	eth2	eth3	eth4
H02200-02	1 Gbps	10 Gbps	10 Gbps	1 Gbps	1 Gbps
有効化 / 無効化	☒	☒	☒	☒	☒
MTU	1500	9000	9000	1500	1500
ネットワークインターフェイス	1	1	1	2	0
ブロードキャストアドレス	Default	Cluster	Cluster	Default	Default
IPアドレス	Default	Cluster	Cluster	Default	Default
タイプ	物理	物理	物理	物理	物理
スループット (MB/s)	0 MB/s	0.01 MB/s	0.01 MB/s	0 MB/s	0 MB/s
最近の変更	18:35:57				
H02200-01	1 Gbps	10 Gbps	10 Gbps	1 Gbps	1 Gbps
有効化 / 無効化	☒	☒	☒	☒	☒
MTU	1500	9000	9000	1500	1500
ネットワークインターフェイス	2	1	1	3	0

8.9 パス管理と Selective LUN Mapping

3 つ以上のノードを持つクラスタは、従来よりも多くのパスを持つ可能性があります。複数のファブリックに接続されたクラスタや、ファブリックごとに複数のストレージコントローラが接続されたクラスタでは、使用可能なパスの数を迅速に増やすことができます。

これにより、ストレージ管理者に次の潜在的な問題が発生します。

- ターゲットポートの数が多いと冗長性には効果的ですが、運用上の負担が大きくなる可能性があります。FC 環境では、より大規模で複雑なゾーンとゾーンセット、追跡が必要なクラスタ SVM に属するより大規模な WWPN のテーブル、iSCSI 環境の場合は、LUN を必要とするホストすべてに対する多数のセッション確立のいずれかが必要になります。
- 多くのオペレーティングシステムでは、アクセス可能なパスの数に上限があります。特に、多数のパスと多数の LUN を持つホストでは、LUN の列挙やパスステータスの問題が発生する可能性があります。

- 要求の厳しい高スループットのワークロードの中には、トラフィックを重要度の低いトラフィックから分離して競合を軽減することでメリットが得られるものもありますが、ALUA パステータスには、ある直接パスを別の直接パスよりも優先するメカニズムがありません。
- ONTAP ストレージ OS には、確立されたパスの総数 (イニシエーター - ターゲット接続 (ITN) として知られる) に対するテスト済みの上限があります。ストレージコントローラの制限の詳細については、[マニュアルサイト](#)に掲載の「ETERNUS AX/HX Series SAN 構成ガイド」を参照してください。

提供されるパスの総数を制限することを検討してください。ただし、HA フェイルオーバーまたはパスに障害が発生した場合に、データへの直接パスと可用性 / 冗長性の両方を確保するには、少なくとも、データがアクセスされているボリュームを含むノードとその HA パートナーの両方にパスが存在する必要があります。

ストレージ OS 機能を使用して LUN が提供するパスを制限する方法には、FC ゾーニングまたは iSCSI セッション管理のみを使用してパスを制限するのではなく、デフォルトで有効に設定されている Selective LUN mapping とポートセットの 2 つがあります。

8.10 Selective LUN Mapping

SLM は、ONTAP クラスタにすでに存在する LUN マッピングテーブルに追加されたもので、LUN パス、igroup、LUN ID のすべての論理リンクで構成されます。LUN は複数の igroup (特にホストベースのクラスタリングシナリオでは) にマッピングされる可能性があるため、そして、igroup には複数の LUN がマッピングされる可能性があるため、すべての LUN マッピングの詳細を得るためにこの表が必要です。

ONTAP を実行しているストレージクラスタでは、これらのプロパティに加えて、すべてのマッピングに、次に示すように、一覧表示されたストレージコントローラから同じマッピングで指定された igroup への LUN の存在を示すレポートノードの一覧も含まれます。

```
san-cluster::> lun mapping show -fields reporting-nodes
vserver      path                igroup lun-id reporting-nodes
-----
SAN_Default_SVM /vol/host1/lun0    linux1 0 node-01,node-02
SAN_Default_SVM /vol/host2/lun0    linux2 0 node-01,node-02
SAN_Default_SVM /vol/host2/lun1    linux2 1 node-03,node-04
```

備考

- デフォルトでは、作成された LUN マッピングにはデフォルトの Selective LUN mapping ポリシーが適用されます。このポリシーでは、LUN が存在するボリュームと HA パートナーを含むノードから LUN が提供されます。
- 以前のバージョンの ONTAP ストレージ OS で作成された LUN マッピングには、ストレージクラスタ内のすべてのノードから LUN を提供するという以前のデフォルト動作を反映するために、レポートノードのリストにワイルドカードエントリがあります。

ただし、LUN マッピングには、HA ペアでグループ化されている限り、クラスタ内の他の任意のノードまたはすべてのノードを含めることができます。また、空白またはワイルドカードにすることもできます。この場合、LUN はクラスタ内のすべてのノードに存在するものとしてレポートされます。このように、ストレージ管理者は、パスを提供するストレージコントローラを非常に詳細に選択できます。

8.11 ポートセット

ポートセットを使用すると、管理者はインターフェースグループをマスクして、そのグループにマッピングされている LUN を使用可能なターゲットポートの総数のサブセットで 사용할 수 있도록 할 수 있습니다. 이 기능은, 클러스터화된 ONTAP 에서 사용할 수 있습니다.

備考

現在ポートセット의 멤버である LIF は、ポートセットから削除されるまで 변경できません. 変更後に 포트 세트에 추가할 수 있지만, 포트 세트에 충분한 수의 LIF 을 남겨두고, 데이터 패스의 호스트 요건을 충족하도록 주의할 필요가 있습니다.

HA フェイルオーバー 또는 무중단 운영 이벤트의 경우에, 데이터에 대한 직접 패스와 사용성 /冗長性의 양쪽을 보장하기 위해 필요한 패스는, 데이터가 액세스되고 있는 볼륨과 그 HA 파트너를 포함하는 노드에 대한 패스입니다.

8.12 管理インターフェース

ブロックプロトコルを使用して 데이터를 제공하는 SVM に属する LIF は管理目的にも 사용되지 않으며, ONTAP 스토리지 클러스터 상의 관리의 논리 단위가 SVM であるため, 모든 SVM は, 블록 프로토콜을 사용하여 데이터를 제공하는 인터페이스에 추가하여, 관리 인터페이스를 가져야 합니다.

ベストプラクティス

ブロック 데이터를 제공하는 SVM 的管理 인터페이스에는, 다음의 프로パティが必要です.

- LIF 类型が data
- 데이터 프로토콜을 할당하지 않음 (-data-protocols none)
- 관리 액세스를 허용하는 방화벽 정책 (-firewall-policy mgmt)
- Snapshot 复制의 생성 또는 관리 등의 데이터 관리의 목적으로 LIF への 액세스가 필요한 호스트에 대해, LIF 을 액세스 가능한 상태로 유지하는 페일 오버 그룹 및 정책 (페일 오버 그룹의 세부 사항은, [마뉴얼 사이트](#)에 게시된 「ETERNUS AX/HX Series 네트워크 관리 리퍼런스」의 「LIF 的 페일 오버 그룹 및 정책의 설정」을 참조하십시오.)

さらに, SVM 레벨의 관리アカウント을 사용할 수 있도록 할 필요가 있습니다. SVM 的 생성時に 생성되는 디폴트의アカウント은 vsadminアカウント이지만, 처음에 security login password 명령을 사용하여 패스워드를 할당하고, 다음에 security login unlock 명령을 사용하여 잠금 해제할 필요가 있습니다.

8. ONTAP プロトコルと SAN プロトコル

8.12 管理インターフェース

System Manager を使用してクラスタを管理する場合、SVM の作成中に SVM 管理 LIF を作成することも、通常の LIF 作成中に SVM 管理 LIF を指定することもできます。[図 8.10](#)、[図 8.11](#)、および[図 8.12](#)を参照してください。

図 8.10 SVM 作成時の管理 LIF の作成

Storage VM の追加

STORAGE VM 名
svm2

アクセスプロトコル
SMB/CIFS, NFS ISCSI **FC**

☒ 自動化 FC

FC ポートの設定

ノード	0c	0d	0e	0f
HQ2200-01	☐	☐	☒	☐
HQ2200-02	☐	☐	☒	☐

Storage VM 管理
☒ 監視表アカウントを管理する

図 8.11 既存の SVM に対する管理 LIF の作成

ネットワークインターフェイスの追加

インターフェイス名
if_svm2_045

プロトコル
データ ☒ クラスタ管理 ☐ Storage VM 管理

プロトコル
☐ SMB/CIFS と NFS ☒ ISCSI ☐ FC

Storage VM
svm2

IP アドレス
192.168.20.100

サブネットマスク
255.255.255.0

ゲートウェイ
オプションのゲートウェイを追加

保存 キャンセル

管理 LIF の詳細は、[図 8.12](#) のようになります。

図 8.12 管理 LIF の詳細

名前	ス...	Storage VM	IPspace	アドレス	現在のノード	現在のホ...	プロトコル	タイプ	スループット
if_Default_357	✓		Default	10.26.238.183	H02200-02	eth		クラスター...	0
if_Default_372	✓		Default	10.26.238.184	H02200-01	eth		クラスター...	0
H02200-01_mgmt1	✓		Default	10.26.238.232	H02200-01	eth1		クラスター...	0
H02200-01_dia1	✓		Cluster	10.26.238.116.139	H02200-01	eth		クラスター...	0.01
H02200-01_dia2	✓		Cluster	10.26.238.177.142	H02200-01	eth		クラスター...	0.01
cluster_mgmt	✓		Default	10.26.238.234	H02200-01	eth1		クラスター...	0
if_svm1_118	✓	svm1	Default	10.26.238.2	H02200-01	eth	SSH, CIFS...	データ	0
if_svm2_438_1	✓	svm2		20.04.481.204ac2044Aa	H02200-01	be	FC	データ	0
if_svm2_438_2	✓	svm2		20.04.481.204ac2044Aa	H02200-02	be	FC	データ	0

8.13 DataMotion for LUNs

LUN の移動とコピーは、ボリュームごとではなく、LUN ごとにボリューム、アグリゲート、ストレージコントローラ、HA ペア間で実行できます。LUN ごとに実行する場合は、`lun move` コマンドと `lun copy` コマンドを使用するか、API 呼び出しを使用します。

この方法を使用して移動またはコピーされた LUN は、ほぼ瞬時に使用可能になります。デスティネーション LUN が作成され、そのメタデータと属性が取得されると、LUN は「プロモート」し、ホストからの I/O 要求を受信できるようになります。一方、ソースからのデータは、クラスタインターコネクトを介してバックグラウンドでコピーされます。デスティネーションにまだ到着していないデータに対して読み取り要求が来ると、読み取り要求を満たす前にデスティネーションからソースに対して通信するようにトリガーされます。書き込み要求が来ると、対象のデータがデスティネーションに直接書き込まれます。

8.13.1 LUN 移動と LUN コピーの比較

DataMotion for LUNs を使用して LUN を移動する場合と、LUN をコピーする場合とでは、いくつかの違いがあります。

LUN は、同じ SVM 内のボリューム間でコピーすることも、他の SVM(クラスタ管理者によって実行された場合) 内のボリュームにコピーすることもできます。コピー先の SVM の構成が同じであるとは想定できないため、LUN の移動は、同じ SVM 内のボリューム間でのみ可能です。まったく別の FC WWNN または iSCSI ターゲット名を持ちます。LUN コピーのデスティネーションには、それに対応する 1 つ以上の LUN マッピングがないため、SVM 間コピーでは問題ありません。

LUN 移動のソースはアクティブなファイルシステムに存在する必要があります。LUN コピーのソースは Snapshot コピー内に置くことができます。Snapshot コピーは不変であり、そこからデータを移動することはできません。

デフォルトでは、LUN コピーは早い段階で昇格されますが、LUN 移動は遅い段階で昇格されます。

- 早い段階で昇格すると、LUN は I/O を受信できますが、Snapshot コピーは作成できません。
- 遅い段階で昇格すると、LUN は I/O を受信でき、Snapshot コピーを作成できます。

8.13.2 ストレージ効率に関する考慮事項

DataMotion for LUNs を使用して移動またはコピーされた LUN は、圧縮または重複排除された状態でターゲットに書き込まれません。

ベストプラクティス

デスティネーションボリュームに重複排除データまたは圧縮データが含まれていない場合、重複排除または圧縮をオンにすると、次にストレージ効率を向上させる際に処理するブロックのリストに書き込まれる LUN のブロックが追加されるため、ブロックスキャンで検出する必要はありません。

データは、ボリューム内で重複排除またはクローン作成を使用してのみ共有が可能になります。コピーされた LUN 内のすべてのデータはソースボリューム内のデータの複製であり、ソースボリューム上の Snapshot コピーでロックされている LUN に属するすべてのデータは、その Snapshot コピーが期限切れになるか削除されるまで、LUN が移動されていてもディスク上に残ります。

インライン圧縮がオンになっているボリュームでは、DataMotion for LUNs によって到着した LUN は圧縮されません。

8.13.3 データ保護に関する考慮事項

データ保護に関する考慮事項は、主に、移動されたがコピーされていない LUN に適用されます。これは、コピーとは、ソースデータがソースボリュームにまだ存在していることを意味するためです。

移動された LUN には、ソースボリューム内にまだ存在する可能性のある、関連づけられた Snapshot コピーは含まれません。ソースボリュームの Snapshot コピー内の LUN データも移動する必要がある場合は、LUN コピーを使用して Snapshot コピーから LUN をコピーできます。重複排除後、ボリュームに移動された LUN とで、使用可能な重複ブロックを共有します。

移動された LUN についてさらに考慮する必要があるのは、その LUN がソースボリュームに関連づけられているデータ保護関係に必ずしも参加していないということです。そのため、新しい SnapMirror 関係の作成などのフォローアップアクションが必要になる場合があります。デスティネーションがすでにこのような関係に含まれている場合は、データレプリケーションのデスティネーションによって消費されているスペースが増えることによって発生するアクションを実行する必要があります。

ベストプラクティス

LUN の移動を ONTAP ストレージクラスタの外部にあるソフトウェアと組み合わせて使用し、LUN を含む Snapshot コピーを管理する場合は、ソフトウェアが DataMotion for LUNs の機能を認識し、(例えば) ボリューム内にもう存在していないかもしれない LUN をそのボリューム内の Snapshot コピーからリストアできることを確認してください。これが不可能な場合は、LUN の移動がデータ保護ワークフローに影響を与える可能性があります。

8.13.4 拡張性とスループットに関する考慮事項

LUN の移動またはコピー操作は、`-max-throughput` 引数を使用して操作ごとに抑制できます。操作の抑制は、オペレーションの開始時に適用することも、`lun copy modify` コマンドまたは `lun move modify` コマンドを使用して既存のオペレーションに適用することもできます。

同時に操作できる移動またはコピー操作の最大数は 50 です。以降の操作はキューに入れられます。この制限は、移動またはコピー操作のデスティネーションに適用されます。

ベストプラクティス

LUN のコピー / 移動操作は、バックグラウンドでデータのコピーが開始された後、いつでも一時停止および再開できます。移動またはコピーを一時停止しても、バックグラウンドでデータが移動されることはありませんが、まだ到着していないデータに対する要求がソース LUN に転送されて処理されてしまいます。

8.13.5 データ管理とワークフローに関する考慮事項

DataMotion for LUNs を使用する場合、その他の ONTAP 機能との相互作用を考慮する必要があります。

- LUN の移動またはコピー元として使用されている LUN は、処理中は削除できません。
- LUN の移動またはコピー元として使用されている LUN は、処理中に SnapRestore を使用して置き換えることはできません。

LUN の移動またはコピー元として使用されている LUN が、ボリューム移動操作によって移動されているボリューム内にある場合、移動ボリュームのカットオーバー期間中に LUN の移動またはコピーが一時停止します。

ベストプラクティス

一部の既存のワークフローでは、DataMotion for LUNs を利用して、必要な手順の数を減らすことができます。

- これまでは、LUN を含むボリュームを複製するには、ボリューム全体のクローンを作成する必要がありました。これで、ボリュームが別の SVM に存在している場合でも、空のボリュームまたはすでに占有されているボリュームに、別のボリュームの Snapshot コピーからの LUN コピーを格納できます。実質的には、以前はボリューム内で使用できたサブボリューム LUN のクローン作成機能を、他のボリュームに拡張できるようになりました。
- 従来は、既存のレイアウトと LUN とボリュームの比率を変更するには、ボリュームのクローンを作成してから不要な LUN を削除するか、ボリューム管理を使用してホスト側のコピーを使用して新しい LUN に古い LUN のデータを格納する必要がありました。LUN を少数のボリュームに統合することでストレージ効率を向上できる場合や、パフォーマンスやストレージ階層化のニーズを満たすために、他の LUN を含むボリューム内の 1 つの LUN を再配置する必要がある場合は、LUN が稼働中でも業務を停止することなくボリューム間で LUN を移動できます。

8.13.6 DataMotion と Selective LUN Mapping : パスの検出と破棄

ストレージクラスタ上の LUN マッピングを変更して新しいパスを作成するか、既存のパスを削除する場合、その LUN に接続されているホストは SCSI バスの再スキャンを実行する必要があります。したがって、HA ペア間で LUN を移動する場合、手順は次のようになります。

- 1 LUN マッピングを変更し、`lun mapping add-reporting-nodes` コマンドを使用して新しいレポートノードを追加します。
- 2 LUN にアクセスしているホストで SCSI バスの再スキャンを実行し、新しいパスを検出します。
- 3 LUN を無停止で移動します。ALUA はホストに対してパスステータスの変更を通知し、ホストは新しい直接パスに I/O を移動し始めます。
- 4 LUN マッピングを変更し、`lun mapping remove-reporting-nodes` コマンドを使用して古いレポートノードを削除します。
- 5 LUN にアクセスしているホストで SCSI バスの再スキャンを実行し、古いパスを破棄します。

再スキャン中に複数の LUN で、新しいパスの検出や、古いパスの削除ができます。

注意

LUN の移動が完了し、ホストの修復手順 (SCSI バスの再スキャンなど) が完了するまで、レポートノードを削除しないでください。新しいレポートノードの追加前、LUN の移動完了前、そしてホストの修復手順がすべて完了する前にレポートノードを削除すると、移動した LUN にアクセスできなくなる可能性があります。

8.14 パス管理のベストプラクティス

ストレージ管理レベルで使用可能なパスの数を制限するには、ONTAP 機能を使用する必要があります。

ベストプラクティス

- 接続されている FC ファブリックまたは Ethernet ネットワークごとに 1 つのターゲット LIF を持つストレージコントローラの場合、LUN マッピングで提供されるパスのデフォルト数は、アクセス対象のボリュームと LUN を含むストレージコントローラからの 2 本の直接パスと、HA パートナーからの 2 本の間接パスの合計 4 本です。
- デフォルトでは、Selective LUN mapping により、LUN とその HA パートナーを所有するストレージコントローラへの LUN パスが制限されますが、一時的または永続的に追加のノードがマッピングの一部になる場合があります。
- 接続されている FC ファブリックまたは Ethernet ネットワークごとに複数のターゲット LIF を持つクラスタでは、追加のパスを使用して LUN ごとにより多くの帯域幅またはキュー深度を提供したり、ポートセットを使用して特定の LIF に対して igroup ごとにトラフィックをチャネル化したりできます。
- デフォルトの LUN マッピングで提供されるよりも多くのパスを必要とする LUN の場合、ほとんどの場合 8 つのパスで十分であり、すべてのホスト SAN 実装でサポートされるパス数です。さらに多くのパスを必要とする LUN については、「ETERNUS AX/HX Series SAN 構成ガイド」に、サポートされるホスト OS ごとにテストされたパスの最大数が掲載されています。
- クラスタ内の HA ペア間で LUN を移動する vol move や lun move などの LUN 移動イベントには、移動イベントが開始される前に、デスティネーションストレージコントローラを使用して LUN が存在することを確認する手順を含める必要があります。lun mapping add-reporting-nodes コマンドを使用して、新しいパスを追加できます。移動が完了したら、lun mapping remove-reporting-nodes コマンドを使用して、直接パスではなくなった、元のパスを削除します。
- LUN に対して提供されるパスを変更することは、新しいパスを検出して古いパスを破棄するために、ホスト SCSI バスの再スキャンを実行する必要があることも意味します。パスの変更に関するホストの観点からのベストプラクティス、および現在パスが存在しない HA ペアへの移動に対応するために LUN マッピングを変更する必要がある場合の手順については、[「8.13.6 DataMotion と Selective LUN Mapping：パスの検出と破棄」\(P.40\)](#) を参照してください。
- LUN マッピングを変更するには、LUN にアクセスしているホストを変更する必要があるため、ホストでの管理手順が望ましくない場合や、HA ペア間での LUN 移動が頻繁に行われる場合は、LUN マッピングでノードのリストを拡張することを検討してください。

9. 拡張性に優れた SAN の価値ある提案と機能

このセクションでは、主な設計目標について説明します。これらの目標には、データ移動のための無停止運用、パフォーマンスの最適化、キャパシティプランニング、システムレベルのハードウェア交換を可能にする、大規模な統合アーキテクチャの提供が含まれていました。これは現在利用可能な主要機能の完全なリストではありませんが、拡張性の高い SAN 機能と ONTAP がストレージ市場の他の機能とどの程度異なっているかを示しています。

9.1 統合ターゲットおよび管理単位としての SVM

ストレージクラスタは、SVM ベースの単一 WWNN をクラスタのすべてのメンバーに拡張して、すべてのノードが同じターゲットを提示し、複数のターゲットが同じ物理ハードウェア上に共存できるようにしています。同じ概念がストレージ管理にも適用されます。すべてのデータは、SVM に関連づけられたボリュームと、SVM の一部として構成された iSCSI または FC ターゲットから提供されるため、ストレージを一度に 1 ノード管理するという時間のかかるプロセスではなく、SVM ごとにクラスタが管理されます。

このように SVM レベルでの管理に重点を置くことで、ストレージ管理の安全なマルチテナンシーモデルを実装することが可能になります。

9.2 ノードおよびクラスタレベルの拡張性

ONTAP はノードレベルとクラスタレベルで拡張性を提供していますが、SAN プロトコルが ONTAP に導入されてからはその両方で拡張性が向上しました。SAN 構成の制限に関する最新の詳細については、[マニュアルサイト](#)に掲載の「ETERNUS AX/HX Series SAN 構成ガイド」を参照してください。概要については、[表 9.1](#) を参照してください。

表 9.1 ONTAP の拡張性

ONTAP のバージョン	9.7 以降
クラスタあたりのノード数	12
ノードあたりの LUN 数	8,192
クラスタあたりの LUN 数	98,304
iSCSI セッション数 / ノード	8,192
FC LUN 数 / ノード	8,192

9.3 クラスタ間の LUN と LIF の移動

ONTAP の以前のバージョンでは、クラスタ内の任意のノードから他の任意のノードへ、無停止でボリュームを移動できました。ボリュームベースではなく LUN ベースでボリュームとストレージコントローラ間で LUN のコピーおよび移動も可能になりました。LUN コピーを使用すると、LUN が即座に使用可能になることにより、クローン作成操作を短縮できます。

通常の操作では、LIF またはボリュームをクラスタノード間で移動する必要はありませんが、状況によっては、ボリュームまたは LIF のいずれかを業務の運用中にノード間で移行することが望ましい場合があります。

LUN およびボリュームをあるノードから別のノードに移行するには、ターゲットノードがホストのダイレクトパスを提供できなくてはなりません ([「8.6 パス選択」\(P.29\)](#) を参照してください)。

あるノードおよびポートから別のノードおよびポートへの LIF の移行は、LIF を削除して再作成するのではなく変更することで、管理上の負担を軽減しつつ行えます。その IP アドレスまたは WWPN は同じままなので、ファブリックゾーニングやホストの変更は必要ありません。SAN LIF は、対象となる LIF (ポートではありません) がオフラインの場合にのみ変更できます。SAN LIF は、`network interface modify -status-admin down` コマンドを使用して管理上オフラインに設定できます。

ベストプラクティス

クラスタのメンバーシップを変更する際は、クラスタのサイズ制限を超えないでください。ブロックプロトコル使用時のクラスタサイズの制限については、[マニュアルサイト](#)に掲載の「ETERNUS AX/HX Series SAN 構成ガイド」を参照してください。

9.4 Foreign LUN Import (FLI)

FC を使用してサードパーティ製アレイおよび ETERNUS AB/HB series のストレージコントローラから LUN をインポートできます。

この機能は ONTAP ソフトウェアに含まれており、使用するためのライセンスやその他の機器を必要としません。インポート処理中に一部のストレージコントローラの FC または UTA2 ポートをイニシエータモードに設定するだけで済みます。UTA2 ポートを使用している場合、FLI がサポートするトランスポートは FC だけなので、これらのポートを FCP パーソナリティに設定する必要があります。

FLI を実行するストレージコントローラは、FC ターゲットから提供された LUN を調べて、同じサイズとジオメトリの LUN をストレージ内の既存のボリュームに作成し、全ソース LUN のデータのコピーをブロック単位で作成します。このコピーには、適切なブロックのアライメントを維持するために必要なオフセットが設定されます。ONTAP で作成された LUN はプロトコルに依存しないため、FC を使用してインポートされた LUN は、ネイティブ LUN と同様に iSCSI を使用してホストに認識される場合があります。

このインポート手順は、オンラインモードまたはオフラインモードで実行できます。オンラインの FLI インポートとは、ソース LUN とターゲット LUN の間にインポート関係を作成し、ホストが新しい場所にストレージをマウントする必要がある場合にのみ、LUN がオフラインになることを意味します。その LUN への I/O は通常どおり続行できます。インポートが完了して関係が壊れるまで、ONTAP はソースとデスティネーションの両方に受信データを多重化します。オフラインの FLI インポート中は、インポートが完了し、インポート関係が壊れるまで、ソース LUN とデスティネーション LUN の両方がホストからアクセスできません。

ETERNUS AX series では、ETERNUS AF/DX series をソースとして使用してインポートを実行できます。このバージョンの FLI 手順の詳細については[マニュアルサイト](#)に掲載の「ETERNUS AX/HX Series Foreign LUN Import を使用した SAN の移行」を参照してください。

ETERNUS AX series オールフラッシュアレイ , ETERNUS HX series ハイブリッドアレイ
ONTAP SAN 構成のベストプラクティス

P3AG-6332-02Z0

発行年月 2025 年 3 月

発行責任 エフサステクノロジーズ株式会社

- 本書の内容は、改善のため事前連絡なしに変更することがあります。
- 本書の内容は、細心の注意を払って制作致しましたが、本書中の誤字、情報の抜け、本書情報の使用に起因する運用結果に関しましては、責任を負いかねますので予めご了承ください。
- 本書に記載されたデータの使用に起因する第三者の特許権およびその他の権利の侵害については、当社はその責を負いません。
- 無断転載を禁じます。