Dell PowerStore: Clustering and High Availability

2022 年 10 月

H18157.7

ホワイト ペーパー

要約

このホワイト ペーパーでは、PowerStore のクラスタリング テクノロジーの概要と、PowerStoreOS リリース向けプラットフォームの高可用性機能について説明します。

Dell Technologies

著作権

この資料に記載される情報は、現状有姿の条件で提供されています。Dell Inc.は、この資料に記載される情報に関する、どのような内容についても表明保証条項を設けず、特に、商品性や特定の目的に対する適応性に関する黙示の保証はいたしません。

本書に記載されているすべてのソフトウェアの使用、複写、および配布には、該当するソフトウェア ライセンスが必要です。

Copyright © 2020-2022 Dell Inc. その関連会社。All rights reserved.（不許複製・禁無断転載）。Dell Technologies、Dell、EMC、Dell EMC、ならびにこれらに関連する商標および Dell 又は EMC が提供する製品およびサービスにかかる商標は Intel、インテル、Intel ロゴ、Intel Inside ロゴ、Xeon は、米国およびその他の国における Intel Corporation の商標または登録商標です。Dell Inc.またはその関連会社の商標又は登録商標です。Published in the USA October 2022 H18157.7

掲載される情報は、発信現在で正確な情報であり、予告なく変更される場合があります。

目次

概要 4

はじめに 5

クラスタリング 8

高可用性 26

まとめ 44

参考資料 45

概要

概要

このホワイト ペーパーでは、Dell PowerStore で利用できる完全に冗長なハードウェアと高可用性機能について説明します。これらの機能は、ネットワークの障害や停電など、システム内および環境内のコンポーネント障害に持ちこたえるように設計されています。個々のコンポーネントで障害が発生しても、ストレージ システムはオンラインのままで、データを提供し続ける。また、複数の障害が別々のコンポーネント セットで発生した場合にも、動作し続けることができます。管理者は、障害についての警告を受け取った後、影響を及ぼすことなく障害が発生したコンポーネントを注文して交換できる。

対象読者

このドキュメントは、IT 管理者、ストレージ アーキテクト、パートナー、デル・テクノロジーズの社員を対象としています。対象読者には、PowerStore システムを使用して、Dell ネットワーク ストレージ環境の評価、取得、管理、運用、設計を行う個人も含まれます。

改訂履歴

日付	説明
（2020 年 4 月）	イニシャル リリース：PowerStoreOS 1.0.0
2020 年 9 月	高可用性に関するセクションの更新、その他のマイナー アップデート
2021 年 4 月	PowerStoreOS 2.0.0 のアップデート
2022 年 1 月	PowerStoreOS 2.1.0 のアップデート、テンプレートのアップデート
2022 年 4 月	PowerStoreOS 2.1.1 のアップデート
2022 年 7 月	PowerStoreOS 3.0 のアップデート
2022 年 8 月	「分散スペアリング」セクションの軽微な修正
2022 年 10 月	PowerStoreOS 3.2 のアップデート

フィードバックを歓迎いたします

デル・テクノロジーズおよび本ドキュメントの作成者は、本ドキュメントへのご意見をお待ちしております。デル・テクノロジーズ チームに E メールでお送りください。

著者： Jonathan Tang、Ryan Meyer

メモ：このトピックに関する他のドキュメントへのリンクについては、PowerStore 情報ハブを参照してください。

はじめに

ドキュメントの目的

データに常時アクセスできることは、新世代のビジネスに欠かすことのできない要素である。データにアクセスできなくなると、事業運営に支障をきたし、収益損失の原因になりかねない。そのため、IT 管理者は、データ センター内のすべてのコンポーネントに単一障害点がないことを確認する必要がある。このホワイトペーパーでは、PowerStore システムをクラスター化する方法について詳しく説明するとともに、PowerStore システムで利用できる高可用性機能について説明します。PowerStore は 99.9999%の可用性を実現するように設計されています（2020 年 4 月に策定されたデル・テクノロジーズの PowerStore 向け仕様に基づきます。実際のシステムの可用性は異なる場合があります）。

PowerStore 製品の概要

PowerStore は、ストレージの分野で新たなレベルの価値、柔軟性、シンプルさを実現する。コンテナベースのマイクロサービス アーキテクチャ、高度なストレージ テクノロジー、統合型の機械学習を利用して、データのパワーを解き放ちます。PowerStore は、パフォーマンス重視の設計を特徴とする汎用性の高いプラットフォームで、スケールアップおよびスケールアウト機能、常時データ削減、次世代メディアのサポートを提供します。

PowerStore は、パブリック クラウドのシンプルさをオンプレミス インフラストラクチャにもたらし、統合型の機械学習エンジンとシームレスなオートメーションによって運用を合理化する。環境の監視、分析、トラブルシューティングを容易にするための予測分析機能が用意されています。PowerStore は適応性が高く、アプライアンス上で特殊なワークロードを直接ホストし、中断なしでインフラストラクチャをモダナイズできる柔軟性をもたらします。また、柔軟な支払いソリューションを通じて投資保護も実現する。

用語

次の表に、本書で使用されているいくつかの用語の定義を示します。

表1. 用語

用語	定義
アプライアンス	ベース エンクロージャと、接続されている拡張エンクロージャを含むソリューション。
ベース エンクロージャ	両方のノード（ノード A とノード B）と 25 個の NVMe ドライブ スロットを収納するエンクロージャ。
クラスター	1 つまたは複数のアプライアンスを単一のグループにまとめたもの。管理インターフェイスを備えています。クラスターは、既存のクラスターにアプライアンスを追加することで、クラスターの許容量まで拡張できます。
拡張エンクロージャ	追加のストレージを提供するためにベース エンクロージャに接続できるエンクロージャ。
Fibre Channel (FC)プロトコル	Fibre Channel ネットワーク上で IP（Internet Protocol）および SCSI（Small Computer Systems Interface）コマンドの実行に使用されるプロトコル。
ファイル システム	SMB、NFS などのファイル共有プロトコルを介してアクセスできるストレージ リソース。

用語	定義
インターネット SCSI (iSCSI)	ネットワーク接続を経由してブロックレベル データ ストレージにアクセスするメカニズムを提供。
クラスター内管理(ICM) ネットワーク	PowerStore クラスターに含まれるアプライアンス間で継続的な管理接続を提供する内部管理ネットワーク。
クラスター内データ(ICD) ネットワーク	PowerStore クラスターに含まれるアプライアンス間で継続的なストレージ接続を提供する内部ネットワーク。
ネットワーク接続型ストレージ (NAS)サーバー	SMB、NFS、FTP、または SFTP プロトコルを使用して、ファイル システム共有およびエクスポート内でファイルのカタログ作成、整理、転送を行う、仮想化されたネットワーク接続型ストレージ サーバー。ファイル レベルのストレージ リソースを作成するには、マルチ テナントの基盤となる NAS サーバーを事前に作成しておく必要がある。NAS サーバーは、そのサーバーで稼働するファイル システムの構成パラメーターを管理します。
Network File System (NFS)	ネットワーク上の Linux/UNIX ホストからデータにアクセスできるようにするアクセス プロトコル。
Node	プロセッサーとメモリーを格納するアプライアンス内のコンポーネント。各アプライアンスは 2 つのノードで構成されます。
NVMe over Fibre Channel (NVMe/FC)	Fibre Channel ネットワーク経由で Non-Volatile Memory Express (NVMe)コマンドを実行するために使用されるプロトコル。
NVMe over TCP (NVMe/TCP)	Ethernet ネットワーク経由で Non-Volatile Memory Express (NVMe)コマンドを実行するために使用されるプロトコル。
PowerStore コマンド ライン インターフェイス(PSTCLI)	ユーザーが UI を使用する代わりにコマンドを入力して、ストレージ システム上のタスクを実行できるようにするインターフェイス。
PowerStore Manager	PowerStore システムの管理に使用される HTML5 ユーザー インターフェイス。
PowerStore T モデル	専用のハードウェアで実行されるコンテナベースのストレージ システム。このストレージ システムは、統合（ブロックおよびファイル）ワークロード、またはブロック最適化ワークロードをサポートします。
PowerStore X モデル	VMware ハイパーバイザーに導入された仮想マシン内で実行されるコンテナベースのストレージ システム。このストレージ システムでは、ブロック最適化ワークロードに加えて、ユーザーがアプリケーションをアレイに直接導入することもできます。
Representational State Transfer (REST) API	PowerStore クラスターのスクリプト化されたプログラムによる対話型管理制御を提供するリソース（オブジェクト）、操作、属性のセット。

用語	定義
SMB（サーバー メッセージ ブロック）	Microsoft Windows 環境で使用されるネットワーク ファイル共有プロトコル。CIFS とも呼ばれます。SMB は、ネットワーク上の Windows ホストからファイルやフォルダーにアクセスする際に使用する。
スナップショット	ストレージ リソースに格納されたデータのポイントインタイム ビュー。ユーザーは、スナップショットからのファイルのリカバリ、スナップショットからのストレージ リソースのリストア、スナップショットのクローンによるホストへのアクセス提供を行うことができる。
ストレージ ポリシーベースの管理（SPBM）	ポリシーを使用して VM のストレージ関連機能を制御し、そのライフ サイクル全体でコンプライアンスを確保。
ボリューム	iSCSI や Fibre Channel などのプロトコルを使用して共有できるブロック レベルのストレージ デバイス。
vSphere Virtual Volumes (VVols)	VM のデータを個々の仮想ボリュームに格納できるようにする VMware ストレージ フレームワーク。この機能により、VM 単位で、かつ SPBM に従ってデータ サービスを適用できます。Virtual Volumes は、この機能を有効にするために使用される個々のストレージ オブジェクトを指す場合もあります。
vSphere API for Array Integration (VAAI)	ストレージ関連タスクをストレージ システムにオフロードすることによって、ESXi ホストの使用率を向上させる VMware API。
vSphere API for Storage Awareness (VASA)	vSphere でストレージ システムの機能を判別できるようにするベンダー中立の VMware API。この機能を使用するには、通信のための VASA プロバイダーがストレージ システム上に必要です。

クラスタリング

概要

どの PowerStore アプライアンスも PowerStore クラスターに導入されています。クラスターには 1 つ以上、最大 4 つまでの PowerStore アプライアンスを設定できます。すべて PowerStore T モデル アプライアンスから構成されたクラスターか、すべて PowerStore X アプライアンスから構成されたクラスターを作成することができますが、同じクラスター内に両方のモデルのアプライアンスを混在させることはできません。マルチアプライアンス クラスターを導入する場合、このタスクを初期設定プロセスで実行することも、アプライアンスを既存のクラスターに追加することもできます。既存のクラスターからアプライアンスを削除することによって、PowerStore クラスターをスケールダウンできます。PowerStoreOS 3.2 では、PowerStore X アプライアンスのマルチアプライアンス クラスタリングがサポートされなくなりました。

PowerStore アプライアンスのクラスタリングには、次のように多くのメリットがあります。

- CPU、メモリー、ストレージ容量、フロントエンド接続性を向上させる簡単スケールアウト
- ストレージ リソースとコンピューティング リソースとで独立したスケーリング
- マルチアプライアンス クラスターの一元管理
- ホスト接続性の自動オーケストレーション
- 耐障害性とフォールト トレランスの向上

次の図に示す例を検討してください。ユーザーは、モデル番号が異なるアプライアンスをスケールアウトして、4 アプライアンス クラスターを作成することができます。スケールアウトのメリットに加えて、さまざまな数の拡張エンクロージャを使用して各アプライアンスをスケールアップできます。たとえば、図に示すアプライアンスには、それぞれ異なる数の拡張エンクロージャが接続されています。クラスター内のアプライアンスごとに異なるメディア タイプを設定できる。たとえば、PowerStore 1000T モデルには NVMe SCM ドライブを搭載する一方で、PowerStore 3000T、5000T、および 7000T モデルには NVMe SSD ドライブを搭載できる。このように柔軟なスケールアウトとスケールアップの導入により、お客様は、モデル番号、ドライブ数、ドライブ タイプに依存することなしにクラスターを拡張することができる。

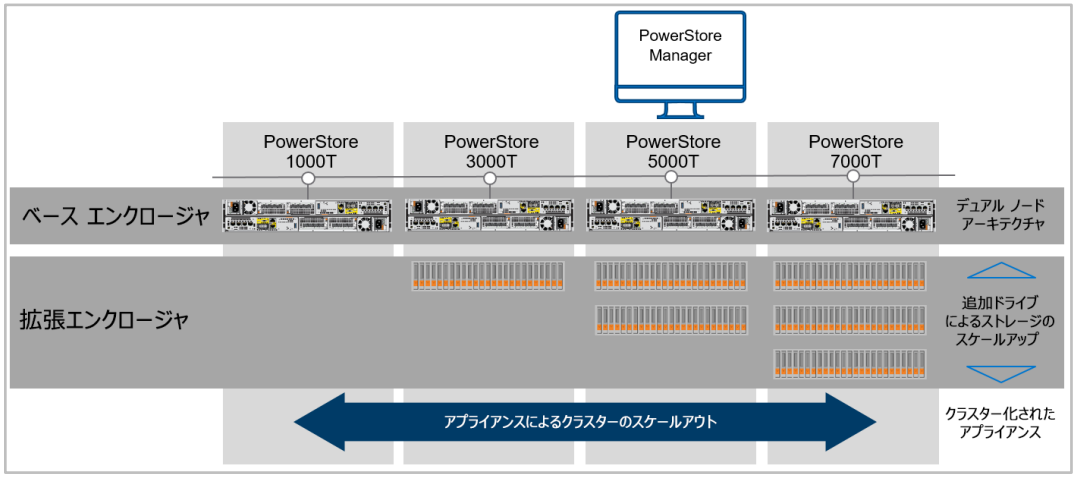

図1. マルチアプライアンス クラスター

要件

dell.com/powerstoredocs の『PowerStore クイック スタート ガイド』を参照して、PowerStore システムが正しくケーブル接続されていることを確認します。PowerStore クラスターの一部であるすべてのアプライアンスの各ノードは、ボンディング ポートを介して他のノードと通信する必要があります（次の図を参照）。ノードの相互通信を可能にするネットワークは、クラスター内管理(ICM)ネットワークおよびクラスター内データ(ICD)ネットワークと呼ばれる内部ネットワークです。マルチアプライアンスの PowerStore クラスターでは、ICM および ICD ネットワークは、自動生成された IPv6 アドレスを持つタグなし VLAN ネットワーク パケットを含んだトップオブラック スイッチ ネットワークを介して通信を行います。クラスター内のすべてのアプライアンスを、同じデータ センター内の同じラックまたは複数のラックに配置する必要がある。PowerStore アプライアンスが複数のスイッチに接続されている場合は、タグなしネットワーク（またはネイティブ VLAN）がスイッチ ポートで構成され、スイッチ間で共有されていることを確認する。同じキャンパス上の複数の異なる建物にクラスターを導入すること（いわゆるストレッチ クラスター）は、サポートされていない。PowerStoreOS 1.0.2 以降のシングルアプライアンス クラスターの場合、ICM ネットワークはトップオブラック スイッチではなく、アプライアンス内のバックプレーンを介して通信を行います。
PowerStoreOS 3.0 シングルアプライアンス クラスターでは、ICM ネットワークと ICD ネットワークの両方がアプライアンスのバックプレーンを介して通信を行います。これらのネットワークをトップオブラック スイッチから移動することで、シングルアプライアンスの PowerStore T システムを最小限のトップオブラック スイッチ構成で導入することができます。

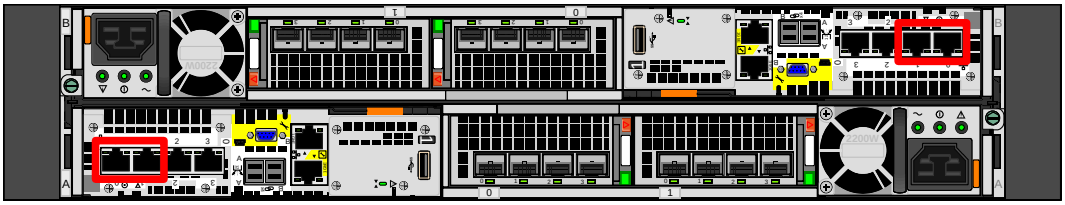

図2.　PowerStore アプライアンスの背面図

プライマリー アプライアンス

各クラスターには、初期構成時に選択される最初のアプライアンスであるプライマリー アプライアンスがある。PowerStore T モデル アプライアンスの場合、プライマリー アプライアンスは、初期構成プロセス中に統合モードまたはブロック最適化モードで構成できる。PowerStoreOS 2.0 以降では、ブロック最適化クラスターのオペレーティング システムによってプライマリー アプライアンスが自動的に選択される。

次の図は、統合モードが選択されたアプライアンスの例を示しています。ファイル サービスは、統合クラスターのプライマリー アプライアンスでのみ実行される。アプライアンスでファイル サービスが使用されている場合は、［**Storage Configuration**］フィールドで［**Unified**］モードが選択されていることを確認する。アプライアンスで iSCSI、FC、NVMe/TCP、NVMe/FC などのブロック ワークロードのみが使用されている場合は、［**ストレージ構成**］で［**ブロック最適化**］モードが選択されていることを確認します。

メモ：クラスターの構成後は、アプライアンスの構成タイプ（統合またはブロック最適化）を変更することはできません。

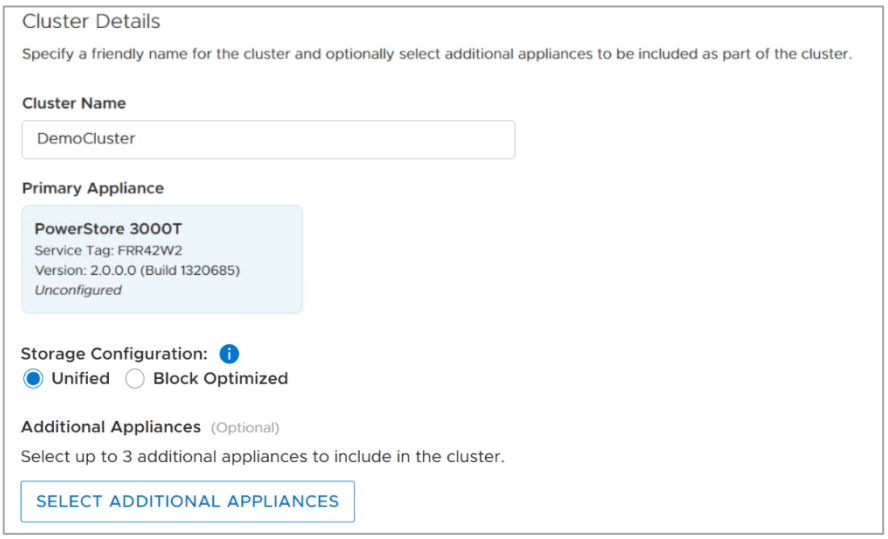

図3.　ユニファイド ストレージ構成

初期構成

初期構成ウィザードを完了するときに、PowerStore システムはネットワーク上の他のアプライアンスをスキャンする。ネットワーク上に複数のタイプの PowerStore T モデルまたは PowerStore X モデル アプライアンスがある場合は、最初に選択した PowerStore モデルのみがウィザードに表示される。たとえば、PowerStore T モデルをプライマリー アプライアンスとして選択すると、PowerStore T モデル アプライアンスのみが表示される。その他のアプライアンスをクラスターに追加するには、［**Select Additional Appliances**］ボタンをクリックする。クラスターの一部に含めることのできるすべてのアプライアンスが表示される。

次の図に、マルチアプライアンス クラスターを選択した場合の例を示します。プライマリー アプライアンスとして PowerStore T モデル アプライアンスが選択されているため、選択可能な PowerStore X モデル アプライアンスはない。検出プロセスの詳細については、『PowerStore：プラットフォームの概要』を参照してください。

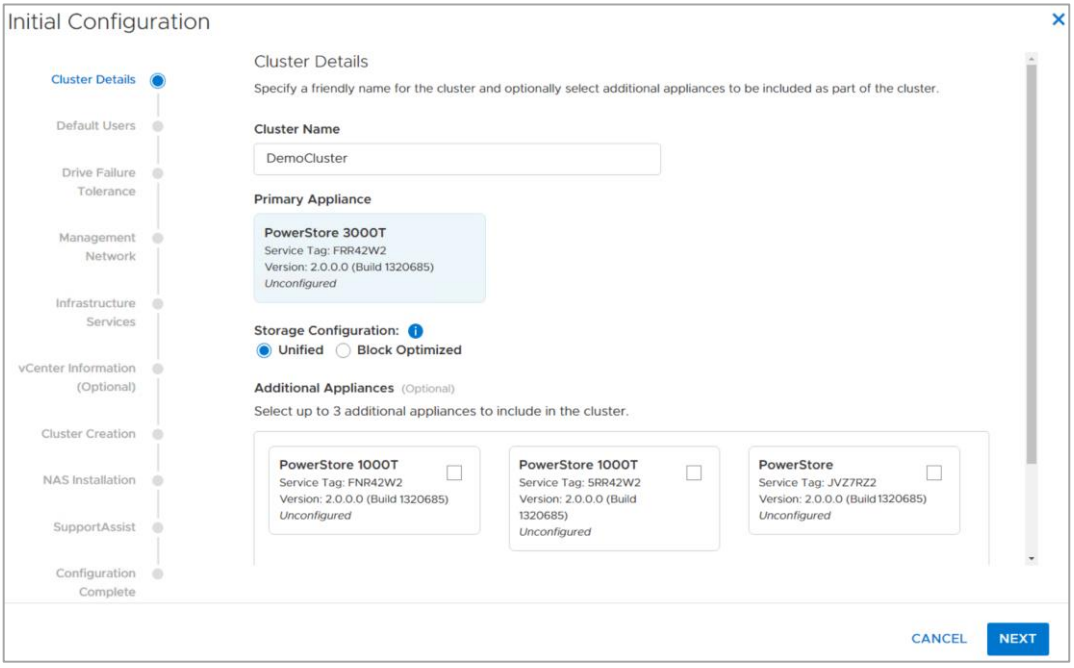

図4. マルチアプライアンスの選択

PowerStore T クラスター

以下は、初めてマルチアプライアンス PowerStore T クラスターを作成するときに考慮すべき一般的なガイドラインである。統合モードとブロック最適化モードの詳細については、本ホワイト ペーパーの以降のセクションを参照のこと。

- 最終的なクラスター サイズが 4 アプライアンスを超えてはならない。
- 正常かつ初期化されていないアプライアンスのみを選択できる。
- 各アプライアンスは、プライマリー アプライアンスから同期的に構成される。
- 統合モード（選択した場合）にすることができるのは、プライマリー アプライアンスのみである。
- クラスターに追加されたアプライアンスはすべて、ブロック最適化構成で自動的に再起動する。
- アプライアンスは、PowerStore システムのボンディング インターフェイス上でタグなし VLAN および IPv6 アドレスを使用して相互に通信できなければならない。
- すべてのアプライアンスのシステム ボンドが同じネイティブ VLAN 上になければならない。

統合クラスター

クラスターを初めてセットアップする場合は、そのクラスターでファイル サービスを使用するかどうかを決定する。まだ決まっていない場合は、将来的な柔軟性を最大限に高めるため、統合クラスターを導入することをお勧めします。「プライマリー アプライアンス」で説明したとおり、サービスは統合クラスターのプライマリー アプライアンスでのみ実行できます。クラスター内に他のアプライアンスが構成されている場合、それらのアプライアンスはブロック最適化構成で導入される。

次の図は、マルチアプライアンス クラスターを作成する場合の例を示しています。プライマリー アプライアンスは統合に設定されており、その他のアプライアンスはブロック最適化構成で再起動されます。

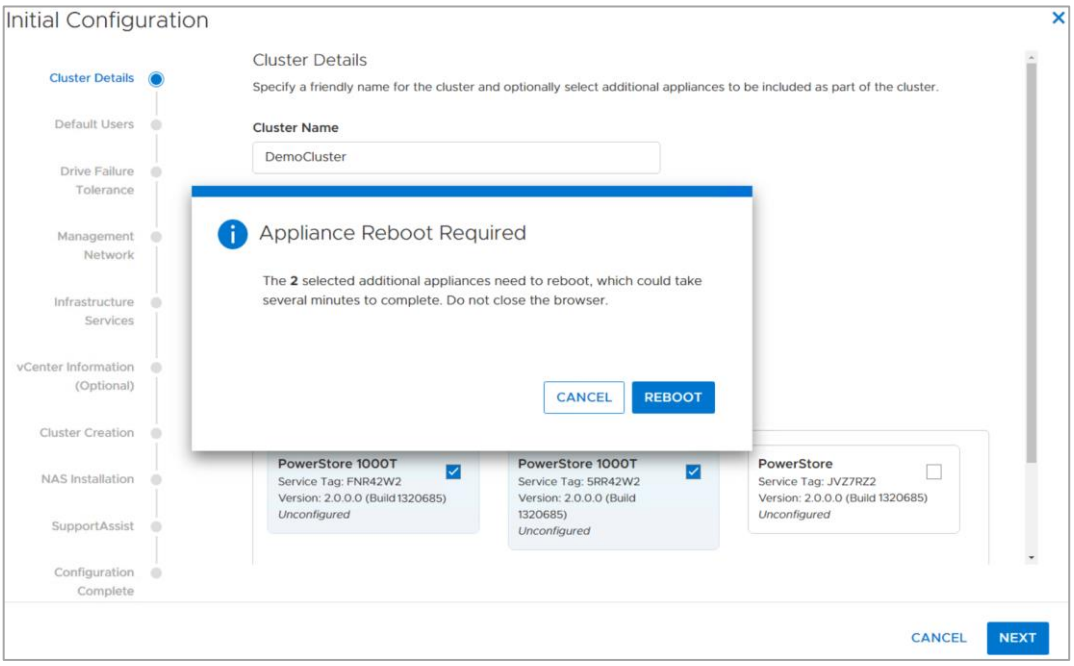

図5.　非プライマリー アプライアンスをブロック最適化構成で再起動する

ファイル サービスを使用しない場合は、クラスターをブロック最適化構成にすることで、ブロック パフォーマンスを最大限に高めることができます。次の図は、ブロック最適化として構成されるマルチアプライアンス クラスターの例を示しています。プライマリー アプライアンスはブロック最適化構成に設定されます。これにより、選択したすべてのアプライアンスがブロック最適化構成で自動的に再起動してからでないと、クラスターを構成できなくなる。

メモ：クラスターの構成タイプ（統合またはブロック最適化）は、クラスターの構成後に変更することはできません。

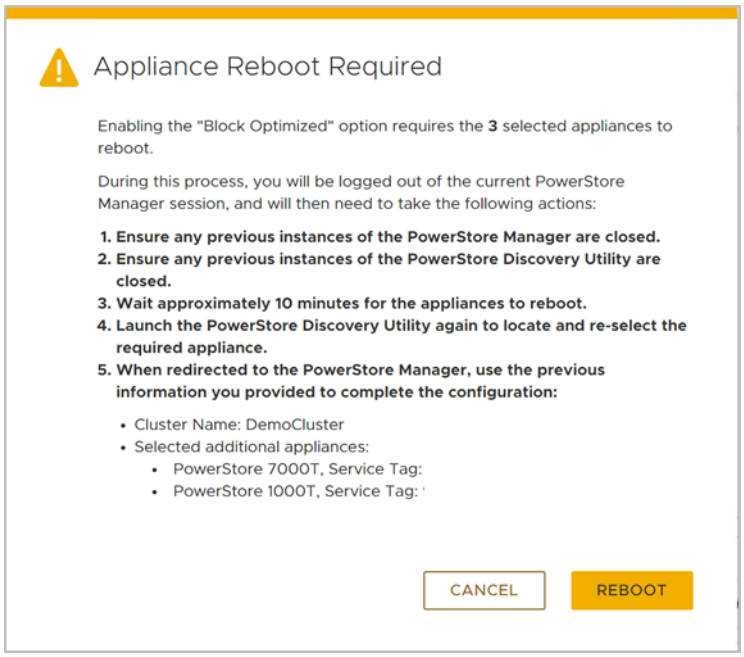

図6.　ブロック最適化クラスター内のすべてのアプライアンスを再起動する

PowerStore X クラスター

以下は、初めてマルチアプライアンス PowerStore X[1]クラスターを作成するときに考慮すべき一般的なガイドラインです。

- 最終的なクラスター サイズが 4 アプライアンスを超えてはならない。
- ユーザーは正常で初期化されていないアプライアンスのみを選択できる。
- 各アプライアンスは、プライマリー アプライアンスから同期的に構成される。
- アプライアンスは、PowerStore システムのボンディング インターフェイス上でタグなし VLAN および IPv6 アドレスを使用して相互に通信できなければならない。
- すべてのアプライアンスの 4 ポート カードの最初の 2 つのポートは、同じネイティブ VLAN 上にある必要がある。

アプライアンスの追加

PowerStore を使用すると、クラスターを個別にスケールアップおよびスケールアウトして拡張できる。容量を追加するには、拡張エンクロージャを追加してスケールアップする。コンピューティング、メモリー、接続を追加するには、既存のクラスターにアプライアンスを追加してスケールアウトする。PowerStore Manager、REST API、および PSTCLI を使用して、アプライアンスを追加できます。

既存のクラスターにアプライアンスを追加する際は、次の一般的なガイドラインを考慮してください。

- 最終的なクラスター サイズが 4 アプライアンスを超えてはならない。
- 初期化されていないアプライアンスのみ追加できる。
- 一度に追加できるアプライアンスは 1 つだけである。
- 新しいアプライアンスと既存のクラスターは、正常な状態でなければならない。
- アプライアンスはすべて、PowerStore システムのシステム ボンド上でタグなし VLAN および IPv6 アドレスを使用して相互に通信できる必要がある。
- すべてのアプライアンスのシステム ボンドが同じネイティブ VLAN 上になければならない。

アプライアンスの選択

PowerStore Manager は、既存のクラスターへのアプライアンスの追加をシンプルにする。次の図は、既存のクラスターにアプライアンスを追加する場合の例を示しています。アプライアンスを追加するには、［**Hardware**］ページに移動して、［**Add**］ボタンをクリックする。これにより、追加できる使用可能な未構成アプライアンスが表示される。PowerStore T モデルをクラスターに追加する場合、アプライアンスは自動的にブロック最適化構成で再起動されます。PowerStoreOS 3.0 以降を実行している PowerStore クラスターの場合、そのクラスターに追加される未構成のアプライアンスは、クラスターで実行されているのと同じバージョンの PowerStoreOS と自動的に同期されます。未構成のアプライアンスを選択した後、ユーザーは［Synchronize］ボタンをクリックするように求められます。同期が完了したら、ユーザーは引き続きアプライアンスの追加ウィザードを使用して、未構成のアプライアンスを追加できます。アプライアンスを追加する際にソフトウェアを同期させるには、未構成のアプライアンスのメジャー リリース番号が前後 1 以内（N-1 または N+1）である必要があります。

[1]PowerStoreOS 3.2 では、PowerStore X アプライアンスのマルチアプライアンス クラスタリングがサポートされなくなりました。

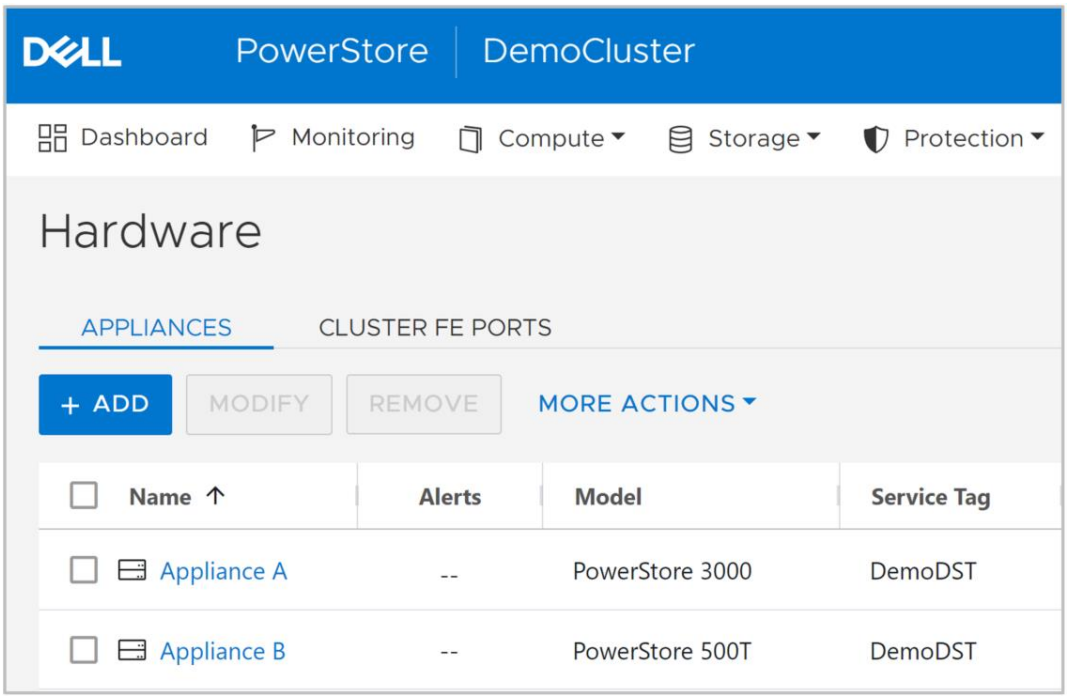

図7.　アプライアンスの追加方法

IP アドレスの構成

アプライアンスの選択後、使用可能な IP アドレスがない場合は、クラスターに IP アドレスを追加する必要があります。IP アドレスは 2 通りの方法で追加できる。たとえば、図 8 と図 9 は、**アプライアンスの追加**ワークフローの一環として IP アドレスを追加する方法を示している。この例では、単一の PowerStore T モデル アプライアンスがクラスターに追加されている。クラスター拡張では、クラスターを拡張するために、管理 IP アドレス 3 つとストレージ IP アドレス 2 つ（図には表示されていない）をさらに追加する必要があります。［**ネットワーク IP の追加**］をクリックすると、管理ネットワークまたはストレージ ネットワークの新しい IP アドレスを指定できます。

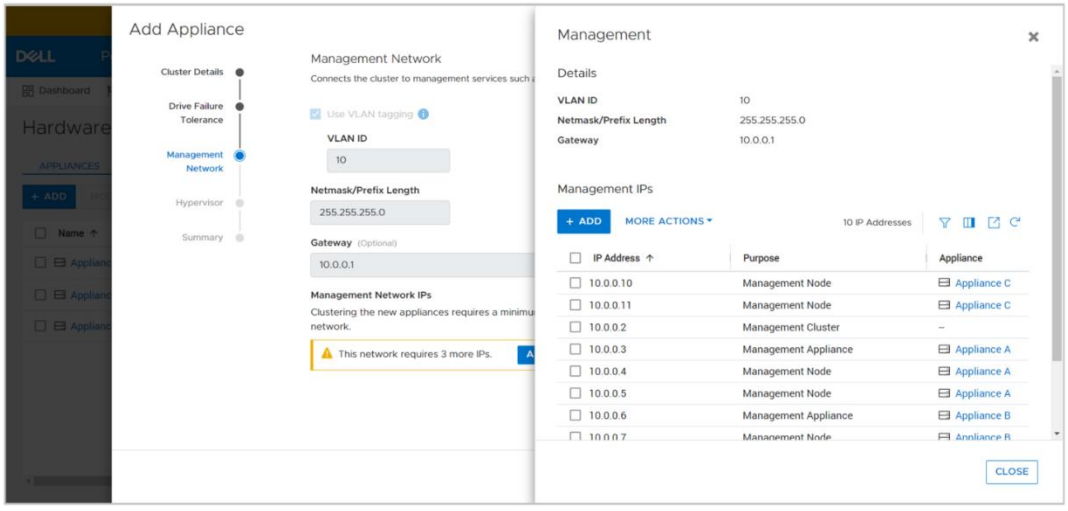

図8.　［Add Appliance］ウィザードから IP を追加する方法

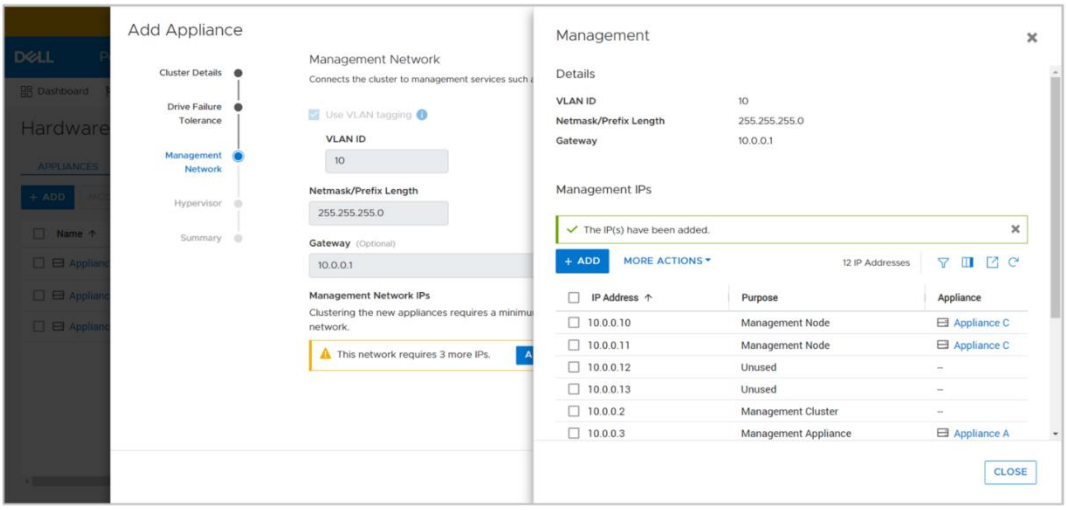

図9.　管理 IP の追加

IP アドレスを追加するもう 1 つの方法は、［**設定**］＞［**ネットワーキング**］＞［**ネットワーク IP**］に移動することです。次の図に、このページの例を示します。ユーザーは、一連の追加の IP アドレスを予約しておいて、それらの少数のアドレスのみが PowerStore システムで最初に使用されるようにすることができます。

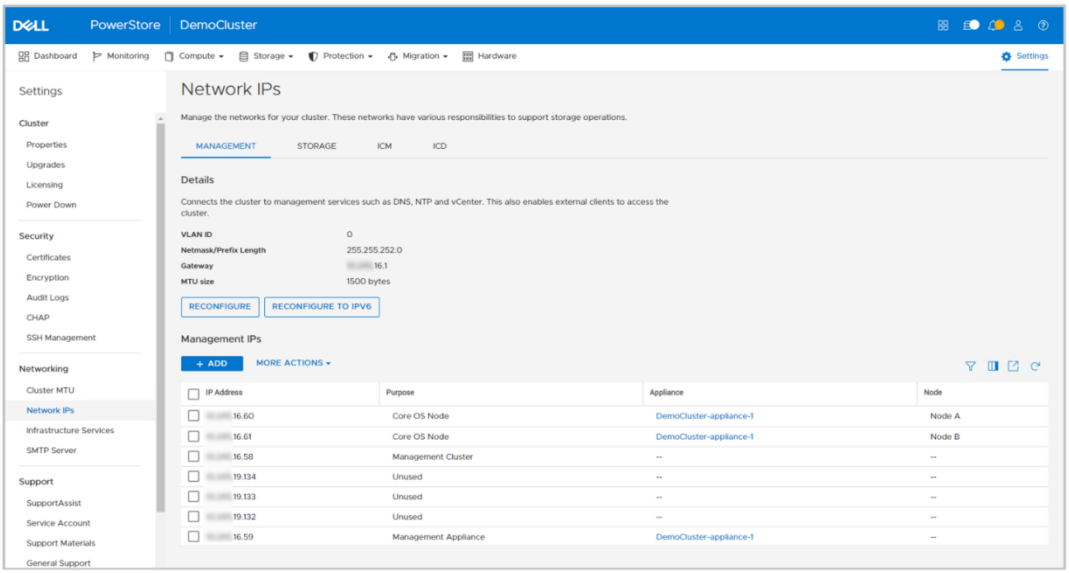

図10.　［Settings］ページから IP を追加する方法

［**Add Appliance**］＞［**Summary**］ページで、アプライアンスの追加ジョブを開始する前に、ネットワーク検証チェックを実行する。ネットワーク検証チェックの実行後、警告またはエラーが表示されることがあります。クラスターの作成を開始する前に警告に対処することをお勧めしますが、警告はバイパスすることもできます。ただし、何らかのエラーが発生した場合は、アプライアンスの追加ジョブを開始する前にエラーを修正しておく必要があります。エラーが解決されていない場合、ユーザーはクラスター作成ジョブを開始できなくなる。図 11 は、［サマリー］ページの例を示しています。ここでは、ネットワークを検証できます。

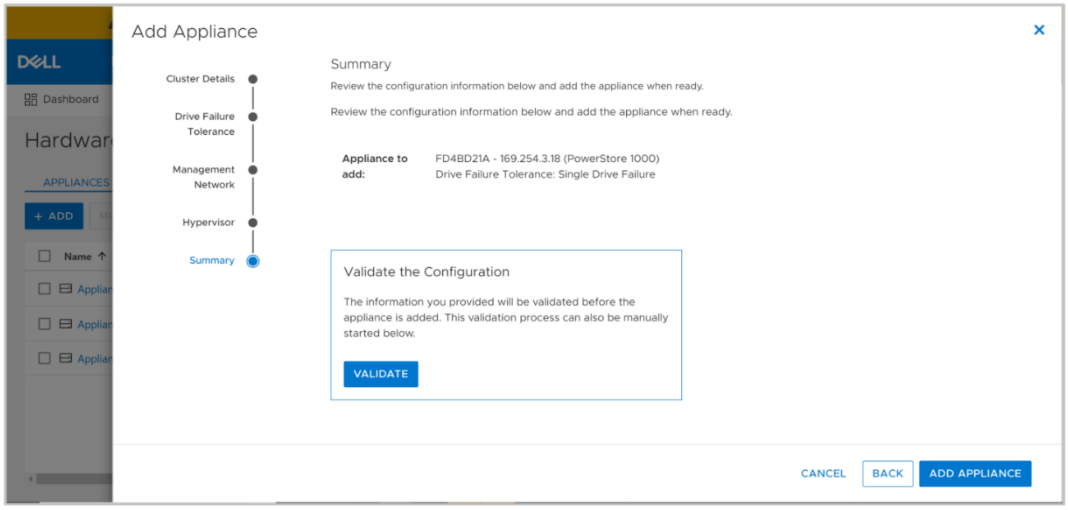

図11. アプライアンスの追加とネットワーク検証

ジョブの確認

アプライアンスの追加ウィザードが完了すると、新しいジョブが作成されます。ジョブの詳細を表示するには、PowerStore Manager の右上にある［**Jobs**］アイコンをクリックする。図 12 は、バックグラウンドで実行されているアプライアンスの追加ジョブの例を示しています。アプライアンスの追加ジョブの実行中は、PowerStore Manager で他のタスクを実行できます。

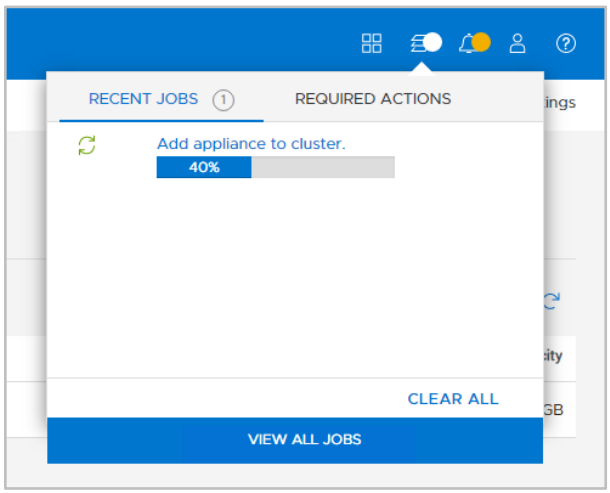

図12. アプライアンスの追加ジョブ

アプライアンスを削除する

クラスターから個々のアプライアンスを削除することで、PowerStore T と PowerStore X の両方のモデルのクラスターをスケールダウンできます。アプライアンスをクラスターから削除するには、慎重な計画と検討が必要となる。このプロセスを開始する前に、アプライアンスからストレージ リソースを自動または手動で移行することができます。クラスターからアプライアンスを削除する前に、ストレージ リソースが残っている場合は手動で削除できます。クラスターから削除されたアプライアンスは、工場出荷時の元の設定に戻ります。

クラスターからアプライアンスを削除する際、PowerStore アプライアンスのモデルによっては、ユーザーが知っておく必要のある別の手順が存在する。表 2 に、PowerStore T モデル アプライアンスをクラスターから削除するための手順の概要を示します。ただし、PowerStore X モデル アプライアンスの場合、表 3 に示すように、ユーザーはストレージ リソースと仮想マシン(VM)をクラスター内の他のアプライアンスに移行する必要があります。PowerStore X モデル アプライアンスの詳細な削除手順については、PowerStore Manager のオンライン ヘルプを参照してください。

表2. PowerStore T モデル アプライアンスを削除する手順の概要

	Entity	操作
1	PowerStore Manager	ストレージ リソースをクラスター内の他のアプライアンスに移行する
2	PowerStore Manager	サポート通知を無効にする
3	PowerStore Manager	クラスターからアプライアンスを削除する

表3. PowerStore X モデル アプライアンスを削除する手順の概要

	Entity	操作
1	PowerStore Manager	ストレージ リソースをクラスター内の他のアプライアンスに移行する
2	PowerStore Manager	サポート通知を無効にする
3	PowerStore Manager	非プライマリー ノードを特定する
4	vSphere Web クライアント	非プライマリー ノード上にある vMotion VM をクラスター上の他のアプライアンスに配置する
5	vSphere Web クライアント	非プライマリー ノードのメンテナンス モードを有効にする
6	vSphere Web クライアント	ESXi クラスターから非プライマリー ノードを削除する
7	vSphere Web クライアント	非プライマリー ノードのメンテナンス モードを終了する
8	vSphere Web クライアント	コントローラーVM の電源を入れる
9	vSphere Web クライアントと PowerStore Manager	プライマリー ノードに対して上記の手順を繰り返す。
10	vSphere Web クライアントと PowerStore Manager	クラスターからアプライアンスを削除する

PowerStore Manager

PowerStoreOS 2.0 以降では、クラスターからアプライアンスを削除する際の推奨方法は PowerStore Manager から確認できます。この操作は、2 ステップのプロセスで実行できます。手順の最初のステップでは、削除するアプライアンスからストレージ リソースを移行します。この手順は、PowerStore Manager から簡単に実行できます。［**ハードウェア**］ページで、削除するアプライアンスを選択し、［**その他のアクション**］＞［**移行**］の順に選択します。ウィザードベースのワークフローに沿って、ストレージ リソースをクラスター内の他のアプライアンスに自動的に移動します。アプライアンス スペース内の除去の詳細については、本ドキュメントで後ほど説明します。

2 番目のステップでは、クラスターからアプライアンスを削除する。次の図に示すように、［**ハードウェア**］ページで、削除するアプライアンスを選択し、［**削除**］を選択します。

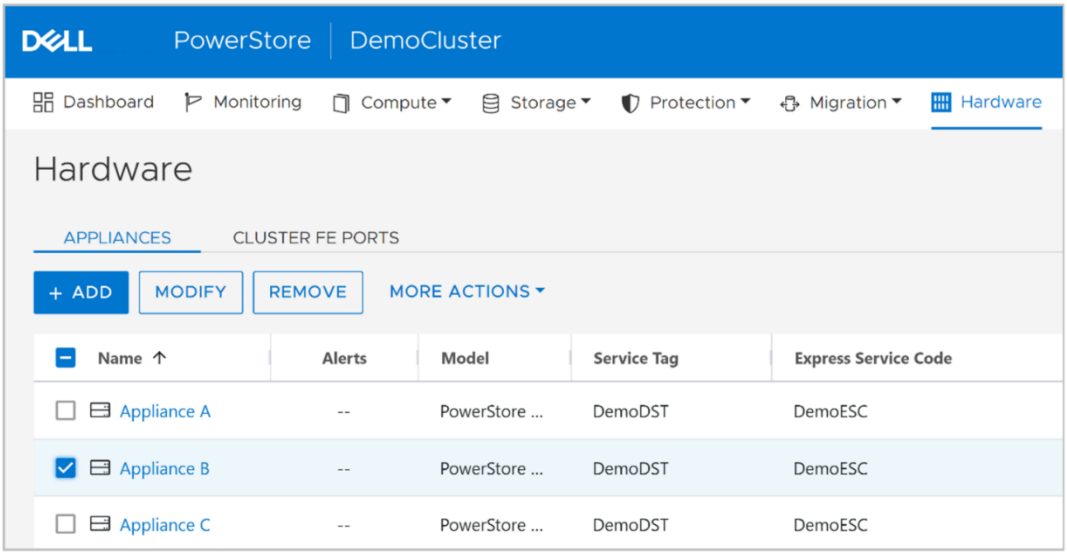

図13. アプライアンスを削除する

サービス スクリプト

PowerStoreOS 2.0 以降を実行していない場合、ユーザーは、サービス スクリプトを実行してクラスターからアプライアンスを削除するための以下のガイドラインとワークフローを参照のこと。

- この操作の進行中にストレージ リソースまたは仮想マシンを作成しないように、すべてのユーザーに通知する。
- **svc_remove_appliance** スクリプトを実行し、アプライアンスのストレージ リソースとワークロードを確認する。
- スクリプトのプロンプトを使用して、スケジューラと自動ストレージ配置を確認して停止し、アクティブなジョブをすべて完了できるようにする。
- ストレージ リソースまたは仮想マシンを、クラスター内の別のアプライアンスに移行する。
- スクリプト内のプロンプトを繰り返し表示して確認し、アプライアンスにアクティブなワークロードがないことを確認する。
- プロンプトに従って、アプライアンスの削除を開始することを確認します。

メモ：svc_remove_appliance サービス スクリプトは、削除されるアプライアンスからのストレージ リソースの自動的な移行は行わない。

VMware 統合

PowerStore T モデルと PowerStore X モデルは、VMware vSphere と緊密に統合できます。これらの統合には、VAAI と VASA のサポートとして、イベント通知、スナップショット管理、Virtual Volumes（VVols）のストレージ コンテナ、PowerStore Manager での仮想マシンの検出と監視が含まれます。デフォルトでは、PowerStore T モデルまたは PowerStore X モデルが初期化されると、ストレージ コンテナがアプライアンス上に自動的に作成される。ストレージ コンテナは、PowerStore から vSphere に VVols ストレージを提供するために使用される。vSphere は、ストレージ コンテナを VVols データストアとしてマウントし、VM ストレージで使用できるようにします。すると、この VVols データストアが内部 ESXi ノードまたは外部 ESXi ホストからアクセス可能になります。管理者がマルチアプライアンス クラスターを管理している場合、ストレージ コンテナの合計容量は、クラスター内の全アプライアンスの集計とな

ります。PowerStore X モデル アプライアンスの場合、クラスター分散型リソース スケジューラ(DRS)が部分自動化モードに設定されています。アプライアンスはこの構成を想定して最適化されているため、これらの設定の変更はサポートされていません。VMware と PowerStore の統合の詳細については、ホワイト ペーパー『PowerStore：仮想化の統合』を参照してください。

リソース バランサー

モダン データ センターで、ミッションクリティカルなアプリケーションをサポートするには、管理者は迅速かつ機敏である必要があります。PowerStore は、PowerStore のオペレーティング システムに組み込まれているインテリジェントな分析エンジンを提供する。このエンジンには、データの初期配置に基づく管理者の意思決定を支援したり、アプライアンス内でノードのパフォーマンスを自動的に分散したりするなど、多くのメリットがあります。また、分析や容量予測に基づくボリューム移行やストレージ拡張を支援します。

データの初期配置

ストレージ管理者がストレージ リソースを作成する際、リソース バランサーで多くのメリットが得られます。たとえば、マルチアプライアンス クラスターを導入すると、リソース バランサーによってインテリジェントかつ自動的に、新規作成されたボリュームがクラスター内の異なったアプライアンスに配置されます。ここでの配置は、未使用の容量が最も大きいアプライアンスはどれであるかに基づいています。自動配置の決定をオーバーライドすることを選択した場合は、ボリュームを配置するアプライアンスを手動で選択できます。

次の図は、特定のアプライアンスに自動的に配置されるボリュームの例を示しています。ボリュームまたはボリューム グループを作成するときは、それらが単一のアプライアンスにのみ配置され、クラスター内の複数のアプライアンスにまたがっていないことに注意します。これらのストレージ リソースは、作成された後、クラスター内の別のアプライアンスに手動で移行されない限り移動しません。クラスター内のストレージ リソースの移行については、本ホワイト ペーパーの以降のセクションで説明する。

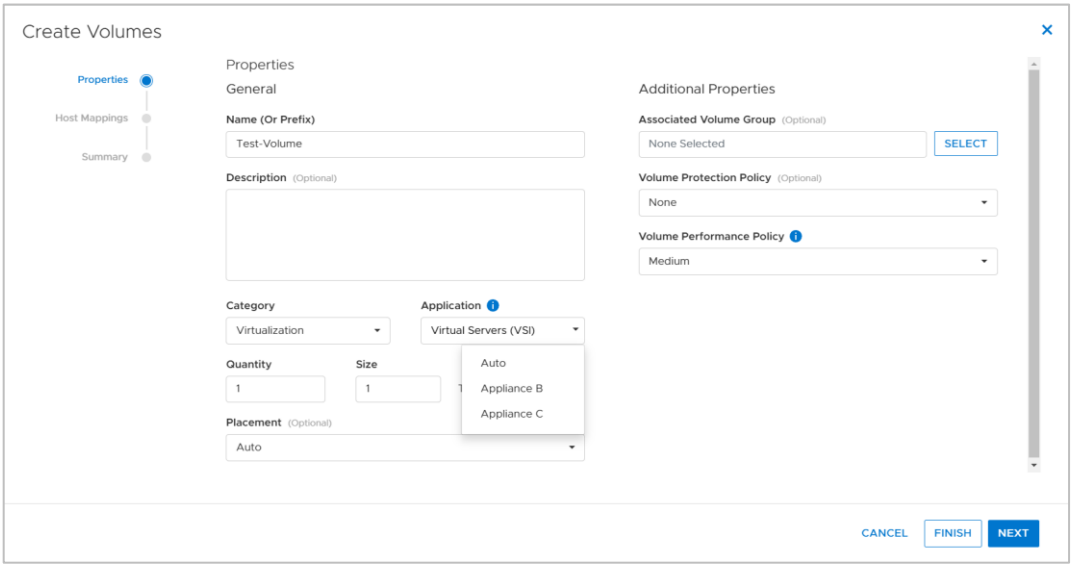

図14. ボリュームの自動配置または手動による割り当て

ボリューム グループ

PowerStore では、初めてボリューム グループを作成すると、ボリューム グループ内の対応するすべてのメンバーが同じアプライアンスに配置されます。既存のボリュームがある場合は、これらのボリュームがクラスター内の複数の異なるアプライアンスに自動的に配置される可能性があります。このような状況では、ストレージ リソースをボリューム グループに配置する前に、単一のアプライアンスに手動で移行しておく必要があります。

移行の概要

アプライアンスでボリュームまたはボリューム グループを作成した後、時間の経過とともに容量と需要が変化する場合があります。PowerStore では、ボリューム、ボリューム グループ、または VVols をクラスター内の別のアプライアンスに移動することがサポートされている。この操作は、PowerStore Manager、REST API、PSTCLI を通じて、PowerStore クラスター内の手動移行、アシスト移行、またはアプライアンス スペースの除去を使用して実行できます。これらの方法については、本ホワイト ペーパーの次のセクションで詳しく説明します。VVols 移行の詳細については、『PowerStore：仮想化の統合』を参照してください。

移行ジョブを開始する前に、管理者は次の手順を確認しておく必要がある。

- 該当する場合は、ホスト、ソース アプライアンス、および宛先アプライアンス間の FC スイッチのゾーニングを確認または設定する。
- ホストが iSCSI、FC、NVMe/TCP、または NVMe/FC プロトコルを使用してソース アプライアンスと宛先アプライアンスに接続できることを確認する。
- PowerStore Manager のホスト オブジェクトが、ソース アプライアンスと宛先アプライアンスに移動するイニシエーターを示していることを確認する。
- 移行されるストレージ オブジェクトの再スキャンをホストから実行する。

上記の手順を確認したら、ホストに対して透過的な移行ジョブを中断なしで開始できます。エンド ユーザーには表示されませんが、ストレージ リソースは、非同期レプリケーション テクノロジーを使用してクラスター内の新しいアプライアンスに転送されます。ストレージ リソースが宛先に移動されるときに、ストレージ リソースに関連づけられているスナップショットやクローンなど、関連するすべてのストレージ オブジェクトも宛先に移動されます。データ転送が完了すると、ホストはパスを宛先アプライアンスに自動的に切り替え、移行ジョブが完了します。

次の図は、4 つのアプライアンスからなるマルチアプライアンス クラスターの例を示しています。この例では、ストレージ リソースを PowerStore 1000T モデルから PowerStore 5000T モデルに移動します。ホストは、iSCSI、FC、NVMe/TCP、または NVMe/FC プロトコルを使用して 2 つのアプライアンスに接続できます。ただし、PowerStore 3000T および 7000T モデルは、ホストに接続できません。そのため、ホストとの接続が確立されないと、ストレージリソースをこれらのアプライアンスに移行できない。

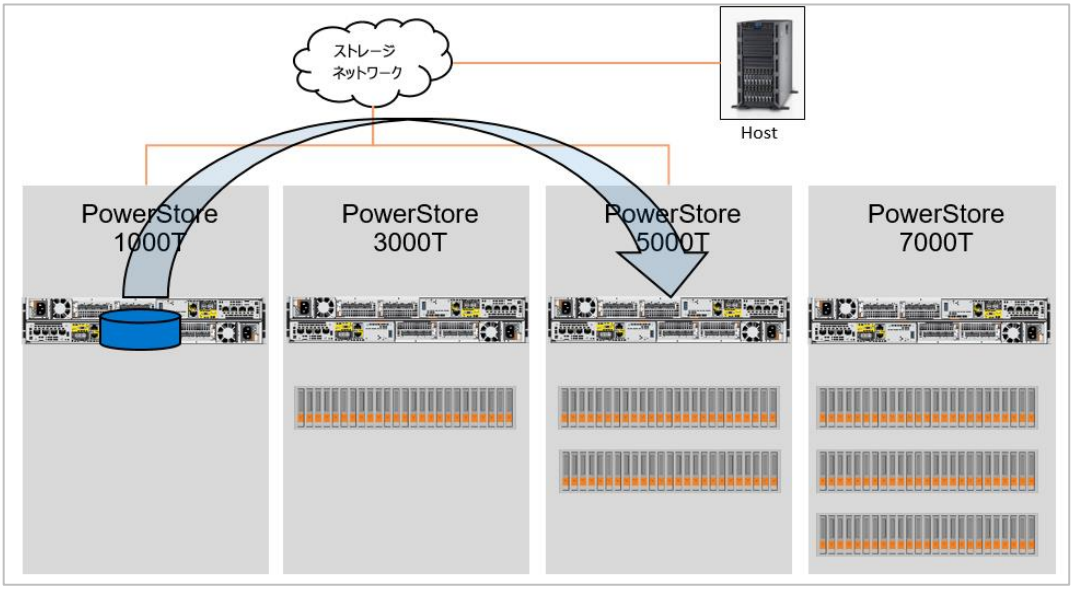

図15. ボリューム移行の例

手動での移行

PowerStore では、ボリューム、ボリューム グループ、VVols などのストレージ リソースをクラスター内の別のアプライアンスに移動することがサポートされている。ストレージ リソースを移動する前に、PowerStore Manager からパスが認識され、ホストの再スキャンが実行されていることを確認する。ホスト イニシエーターのパスが検証されたら、ストレージ リソースで手動移行を実行できます。

ボリューム、ボリューム グループ、または VVols の PowerStore での手動移行の方法は、ストレージ リソースによって異なるため、次の点を考慮してください。

- ボリューム：［**ストレージ**］＞［**ボリューム**］を選択し、クラスター内の使用可能なすべてのボリュームのリストを表示します。移行するボリュームを選択して、［**More Actions**］＞［**Migrate**］をクリックする。
- ボリューム グループ：［**ストレージ**］＞［**ボリューム グループ**］を選択し、クラスター内の使用可能なすべてのボリュームのリストを表示します。移行するボリューム グループを選択して、［**More Actions**］＞［**Migrate**］をクリックする。
- VVols：［**ストレージ**］＞［**ストレージ コンテナ**］を選択し、仮想マシンが導入されているストレージ コンテナを選択します。［**Virtual Volumes**］タブを選択して、クラスター内の関連するすべての vVols を表示し、移行する vVols を選択して、［**移行**］をクリックします。

次の図に示す例では、「Test-Volume-001」という名前のボリュームが、2 台のアプライアンスからなるクラスター上に作成されています。この例では、ボリュームはアプライアンス 2 で作成されており、ボリュームを移行できるクラスター内の使用可能なアプライアンスのリストがウィザードに表示されている。［ボリュームの移行］ウィザードが開くと、2 台のアプライアンスからなるクラスター内のもう一方のアプライアンスがアプライアンス 1 として表示され、管理者は［**移行の開始**］をクリックして移行ジョブを開始できます。

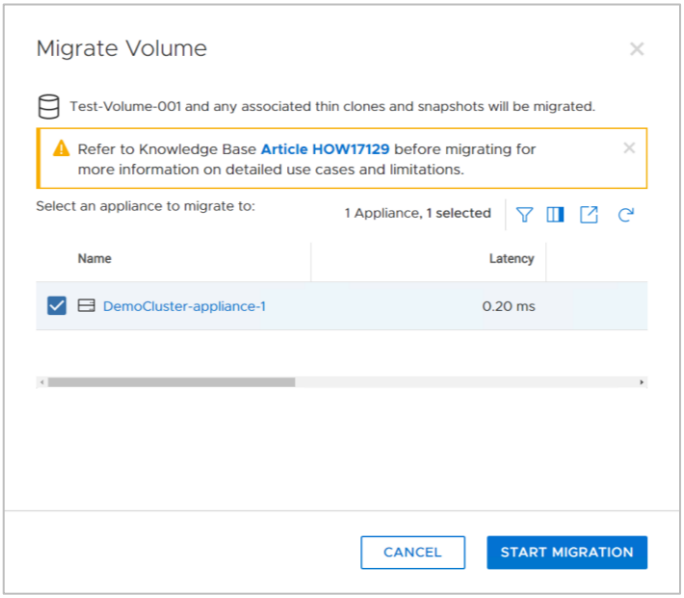

図16. ボリューム移行ジョブの開始

アシスト移行

システムの実行中は、クラスター内のすべてのアプライアンスでストレージ リソースの使用率が定期的に監視される。時間の経過とともに、アプライアンスが最大有効容量に近づくことがあります。このような場合、移行の推奨事項がシステムによって生成されます（次の図を参照）。これらの推奨事項は、ドライブの摩耗、アプライアンスの容量、稼働状態などの要因に基づいて生成される。管理者は、アシスト移行の推奨事項をアラートから表示することも、PowerStore Manager の［**移行**］＞［**移行アクション**］ページに移動して表示することもできます。管理者が移行の推奨事項に同意すると、移行セッションが自動的に作成される。元のストレージ リソースに関連づけられているスナップショットまたはクローンは、手動のボリューム移行と同様に新しいアプライアンスに移行されます。

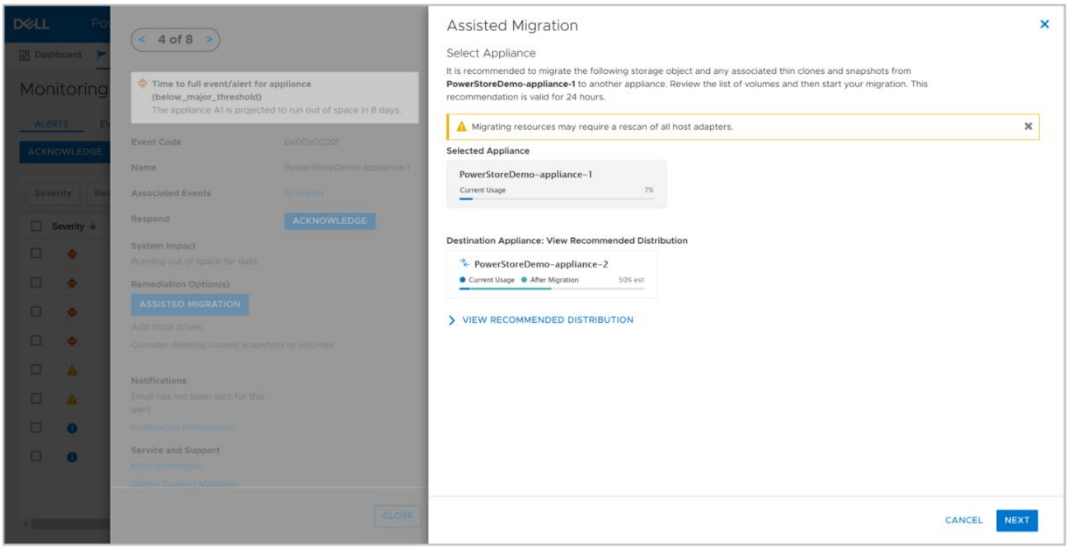

図17. アシスト移行の推奨事項

アプライアンス スペース内の除去

PowerStoreOS 2.0 では、ボリューム、ボリューム グループ、VVols などのストレージ リソースをアプライアンスから自動的に移動させる機能（すべて PowerStore Manager から操作）が導入されています。この使いやすいウィザードベースのワークフローは、アプライアンスの電源をオフにしたり、アプライアンスをクラスターから削除したり、一杯になったアプライアンスのストレージ リソースを移行したりするのに役立ちます。

このタスクを実行するには、［**ハードウェア**］ページに移動し、アプライアンスを選択して［**その他のアクション**］＞［**移行**］をクリックします。次の図に示すように、このアクションにより新しいワークフローが開始されます。ウィザードには、ボリューム、ボリューム グループ、vVol を含むストレージ リソースのリストが表示される。表示されているすべてのストレージ リソースを選択することも、個別のストレージ リソースを手動で選択することもできます。すべてのストレージ リソースがウィザードの実行時に使用可能であるとは限りません。アクティブなインポート セッション中またはアクティブな内部移行セッション中のストレージ リソースは、移行の対象外としてマークされます。インポート セッションまたは移行セッションの完了後は、ウィザードを再実行し、ウィザードで他のアプライアンスにストレージ リソースを移動することができます。

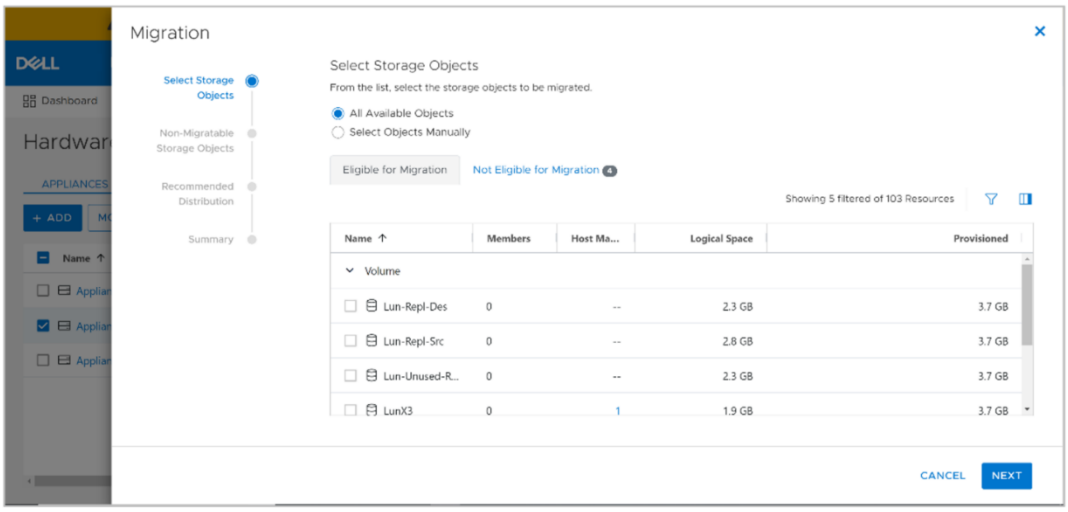

図18. 移行ウィザード

ストレージ リソースを選択すると、PowerStore は、さまざまなストレージ リソースに最適なアプライアンスをインテリジェントに計算します。次の図に示すように、PowerStore は LunX3 を宛先アプライアンス C に移動することを推奨しています。PowerStore はこれらの推奨事項を自動的に作成しますが、状況によっては特定のアプライアンスにストレージ リソースを設置する必要が生じることがあります。次の図に示すように、アプライアンス C の横に表示されている鉛筆アイコンをクリックして宛先をオーバーライドできます。

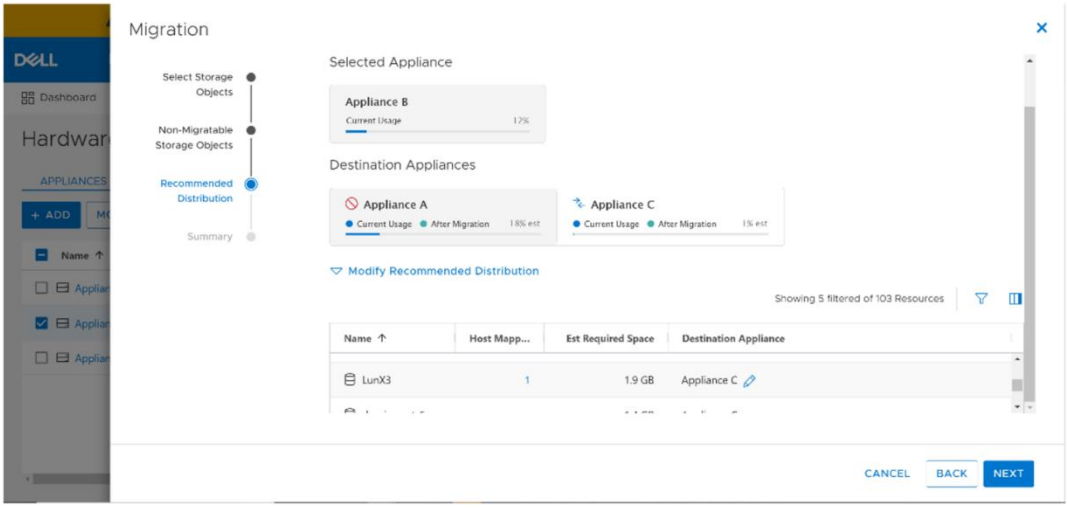

図19. ストレージ リソース アプライアンスの推奨事項

ウィザードの最後に、再スキャンが必要な関連ホストのリストが表示される場合があります。再スキャンが完了すると、自動的に移行セッションが生成され、移行ジョブが開始されます。

移行状態

ストレージ リソースの移行が生成されたら、管理者はクラスター内でアクティブになっているさまざまなセッションを表示できます。管理者は、PowerStore Manager、REST API、または PSTCLI を使用して、アクティブな移行セッションを一時停止、再開、または削除できる。次の表は、移行セッションで起こりうるすべての状態を示しています。

表4. 移行セッションの状態

State	定義
Initializing	移行セッションが開始され、セッションの初期化が完了するまでこの状態のままになる
Initialized	セッションの初期化が完了すると、移行セッションはこの状態に遷移する
Synchronizing	バックグラウンド コピーが進行中
Idle	最初のバックグラウンド コピーが完了すると、移行セッションはこの状態に遷移する
Cutting_Over	コピーの最後の部分がこの状態で実行され、ストレージ リソースの所有権が新しいアプライアンスに移転される
削除中	移行セッションを削除中
Completed	移行セッションが完了し、セッションを安全に削除できる
Pausing	一時停止コマンドが発行されると、移行セッションはこの状態に遷移する
Paused	移行セッションが一時停止されており、セッションを再開するにはユーザー介入が必要となる

State	定義
System_Paused	エラーが発生した場合、移行セッションはこの状態に遷移する。ユーザーはエラーを解決した後で移行を再開または削除できる
Resuming	移行セッションのバックグラウンド コピーが再開された
フェイル	移行セッションでエラーが発生した

容量フォーキャスト

PowerStore では、クラスターの領域が不足する可能性があるタイミングの詳細が表示されます。これらの詳細は、ビジネス ニーズに基づいてクラスターの拡張を計画する際に役立てることができるこれらの判断に必要な情報を提供するため、PowerStore は、アプライアンスで利用されたストレージに関する履歴統計を維持し、このストレージが時間の経過とともにどのように使用されるかをインテリジェントに予測する。

可能な限り正確な情報を提供できるよう、統計は 15 日間にわたって収集される。15 日後、PowerStore Manager は、システムの領域がいつ不足する可能性があるかについての予測を表示します。次の図は、PowerStore Manager での容量予測の表示例を示しています。

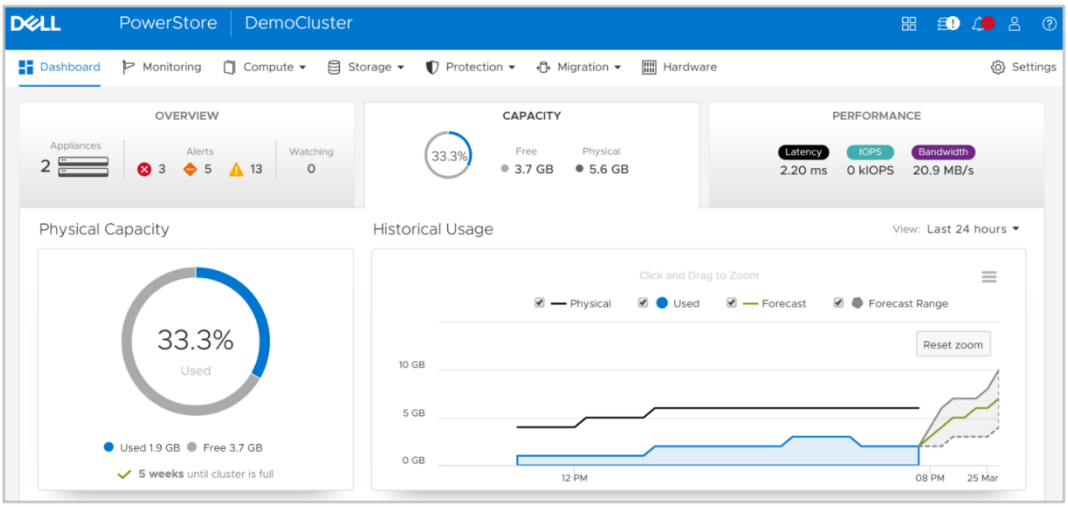

図20. 容量フォーキャスト

動的ノード親和性

ボリュームをホストにマッピングする際、PowerStore は、次の図に示すように、ホストが優先的にアクセスするノード（ノード親和性とも呼ばれる）を動的に設定します。時間の経過とともに多くのストレージ リソースが作成されて使用される中で、1 つの PowerStore ノードがピア ノードよりも頻繁に使用される状況が生じることがあります。このような状況では、特定のパフォーマンス メトリック チェックが満たされている場合、PowerStore が 1 つ以上のストレージ リソースのノード親和性を自動的に変更することがあります。

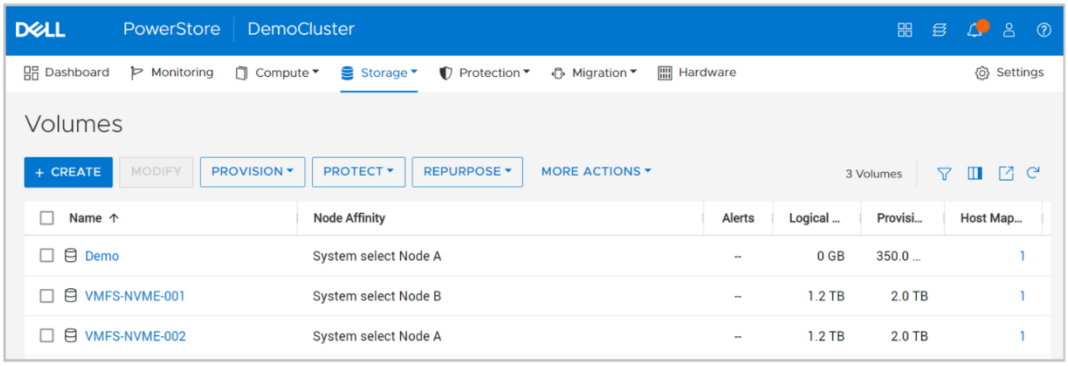

図21. ノード親和性

最初のパフォーマンス チェックでは、ソース ノードと宛先ノードの読み取り/書き込み IOPS を比較します。2 つ目のパフォーマンス チェックでは、少なくとも 1 つのノードで CPU の利用率が 50%を超えていることを確認します。最後のパフォーマンス チェックでは、システムがデータ パス エンジンによって報告されたレイテンシー チェックを上回ろうとしているかどうかを確認します。PowerStore は、30 分ごとにこれらのパフォーマンス メトリックをチェックします。これら 3 つのパフォーマンス チェックがすべて、30 分間にわたって満たされている場合、PowerStore はボリュームのノード親和性を動的に変更します。この変更は、ホストに対して透過的かつ無停止で行われます。

高可用性

はじめに

PowerStore アプライアンスはマルチアプライアンス構成をサポートしてクラスター全体の冗長性とフォールト トレランスを向上させるが、各 PowerStore アプライアンスには HA ペアを構成する 2 つのノードもある。PowerStore には、完全に冗長なハードウェアが備わっており、高可用性機能がいくつか含まれている。これらの機能は、システム自体のコンポーネント障害と環境内のコンポーネント障害に対する耐性を持つよう設計されている。障害の例としては、ネットワークまたは電源の障害などが挙げられる。個々のコンポーネントで障害が発生しても、ストレージ システムはオンラインのままで、データを提供し続ける。
別々のコンポーネント セットで障害が発生した場合、システムは複数の障害にも耐えることができる。障害アラートが発行された後で、障害のあるコンポーネントを注文して、影響を与えることなく交換できる。このセクションでは、プラットフォームとクラスター内の PowerStore の冗長性とフォールト トレランスについて説明する。

ハードウェアの冗長性

PowerStore は、2 つの同一ノードを含むデュアルノード アーキテクチャによって冗長性を実現し、両方のノードが同時に I/O を処理するアクティブ/アクティブ コントローラー構成を備えている。この機能により、アイドル状態のスタンバイ ハードウェアが不要になるため、ハードウェアの効率性が向上する。ベース エンクロージャには、これらのノードと最大 25 台の 2.5 インチ ドライブが含まれている。

図 22 と図 23 に、以下を含む PowerStore 1000～9000 モデルの内部コンポーネントを示します。

- PSU
- BBU（バッテリー バックアップ ユニット）
- インテル CPU
- インテル QAT ハードウェア オフロード エンジン
- マザーボード
- 冷却ファン

- M.2 SSD x 2
- メモリーDIMM
- I/O モジュール（オプション）

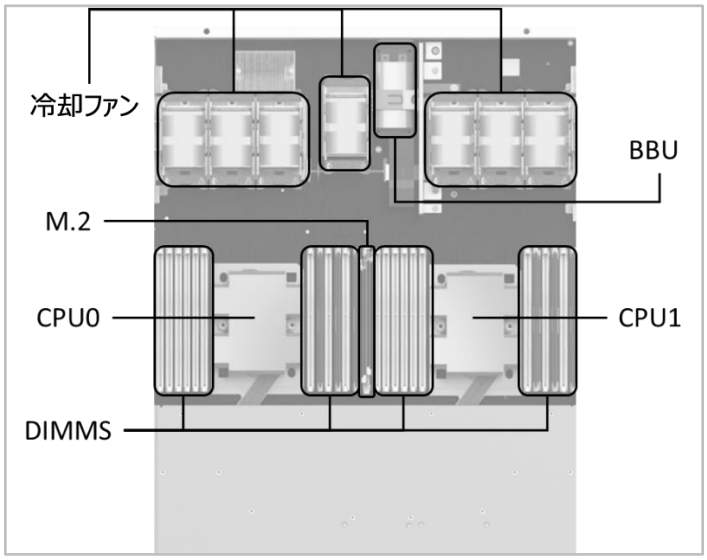

図22. PowerStore 1000～9000 ノードの内部コンポーネント（上面図）

図23. PowerStore 1000～9000 ノードの背面図

PowerStore 1000～9000、PowerStore 1200～9200（PowerStoreOS 3.0 以降で使用可能）、および PowerStore 500 アプライアンスのコンポーネントの詳細については、次のドキュメントを参照してください。

- dell.com/powerstoredocs の『PowerStore ハードウェア情報ガイド』
- dell.com/powerstoredocs の『500T 向け PowerStore ハードウェア情報ガイド』

また、PowerStore Info Hub のホワイト ペーパー『PowerStore：プラットフォームの概要』も参照してください。

管理ソフトウェア

管理ソフトウェアは、PowerStore Manager、管理インターフェイス（クラスターIP）、および REST API などのサービスを実行します。PowerStoreOS 3.0 より前のバージョンで実行されている PowerStore システムの場合、すべての管理ソフトウェアがアプライアンス内の 1 つのノードで一度に実行されます。このような以前のバージョンの PowerStoreOS では、管理ソフトウェアを実行するノードが PowerStore Manager でプライマリー ノードとして指定されています。図 24 は、ハードウェア プロパティ ページのプライマリー ノードの例を示しています。これは、［**ハードウェア**］＞［**背面図**］に移動するとハイライト表示されます。

PowerStoreOS 3.0 以降では、各アプライアンス ノードの CPU の利用率を向上させるために、管理リソースがアプライアンスの両方のノード間で分割されバランスが取られています。PowerStoreOS 3.0 以降では、PowerStore Manager はセカンダリ ノードで実行され、他の管理サービスはプライマリ ノードで実行されます。

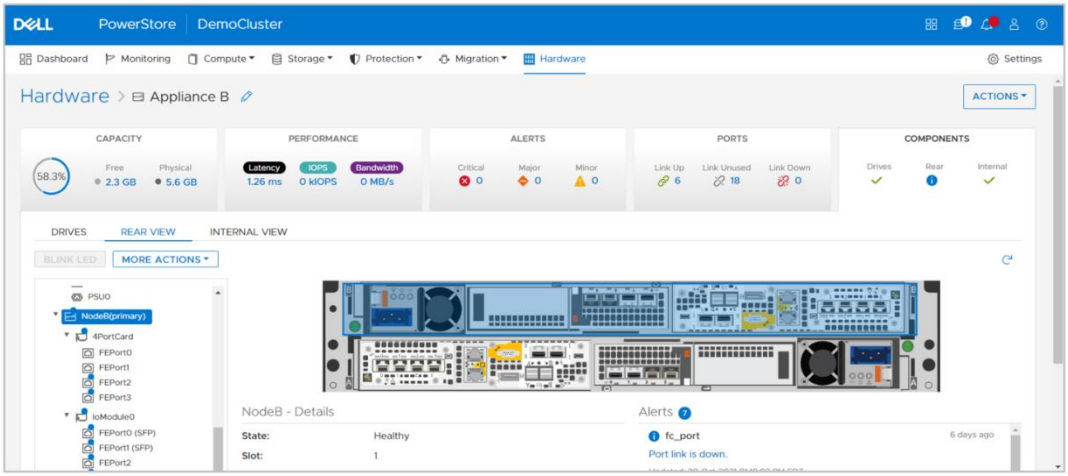

図24. プライマリー ノード

ノードが再起動するか、クラッシュするか、または管理接続が停止した場合、そのノードで実行中のサービスは自動的にピア ノードにフェールオーバーします。フェールオーバー後、すべてのサービスが完全に起動するまで数分かかる場合があります。フェールオーバー中に PowerStore Manager にログインしているユーザーには、接続が失われたことを示すメッセージが表示されることがあります。フェールオーバー プロセスが完了すると、PowerStore Manager へのアクセスが自動的にリストアされますが、ユーザーは手動でブラウザーを更新してアクセスを回復することもできます。ストレージ リソースへのホスト アクセスは優先され、PowerStore Manager にアクセスできるようになる前に使用可能になる。

フェールオーバー後、サービスは一時的に新しいノードで実行されます。ピア ノードがフェールオーバー イベントからリカバリーされると、サービスは 5 分後に自動的に再バランシングされます。再バランシング中に、サービスがリカバリーされたノードに戻される間、PowerStore Manager の接続が最大 3 分間失われる場合があります。フェールオーバー イベントの後、プライマリー ノードがピア ノードに変更される場合があり、システムが再起動されるか再度フェールオーバーされるまで、この状態が続くことに注意してください。

システム ボンド

PowerStore システムでは、システム全体に単一障害点がないことが保証されている。PowerStore T モデル アプライアンスの場合は、次の図に示すように、4 ポート カードの最初の 2 つのポートをトップオブラック スイッチにケーブル接続することをお勧めします。PowerStore T モデルでは、これらのポートは内部的に**システム ボンド**と呼ばれる。このインターフェイスは、単一障害点がなく、データ サービスと本番データが常に利用可能であることを保証するために作成される。以下は、PowerStore システムでデータ サービスを利用する場合に、システム ボンドを通過する各種トラフィックのリストです。

- 管理トラフィック（PowerStore X モデルの場合）
- クラスター通信
- iSCSI または NVMe/TCP ホスト トラフィック（オプション）
- レプリケーション トラフィック（オプション）
- NAS トラフィック（PowerStore T モデルの場合）（オプション）

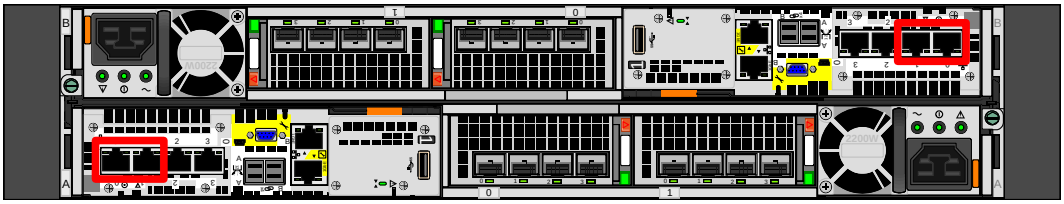

図25. 4 ポート カードの最初の 2 ポート（システム ボンド）

PowerStore T モデル アプライアンスの場合、システム ボンド内の個々のポートは、アクティブ/アクティブ モードまたはアクティブ/パッシブ モードで実行されている可能性がある。この状態は、ネットワーク スイッチで Link Aggregation Control Protocol (LACP)が構成されているかどうかによって異なります。耐久性とネットワーク パフォーマンスを最大限に高めるには、ネットワークで LACP を有効にすることをお勧めします。PowerStore での LACP の詳細については、このドキュメントの「LACP（Link Aggregation Control Protocol）」を参照してください。PowerStore T モデル アプライアンスの LACP の構成に関する詳細については、dell.com/powerstoredocs の『PowerStore T モデルの PowerStore ガイド』を参照してください。

PowerStore X モデル アプライアンスの場合、リンク アグリゲーションとフェールオーバーは VMware ハイパーバイザー内の仮想スイッチによって処理されるため、システム ボンドは存在せず、LACP 構成も不要です。

PowerStore を使用すると、ポートを追加して帯域幅を増やしたり、フォールト トレランスを高めたりすることができます。管理者は、クラスター内のアプライアンスに関連づけられている追加の物理ポートまたは仮想ポート（該当する場合）をマップすることにより、ストレージ トラフィックを拡張できます。PowerStore T モデル アプライアンスの場合、レプリケーション トラフィックが 4 ポート カードまたは I/O モジュールの他のポートにすでにタグ付けされていれば、管理者はシステム ボンド ポートからレプリケーション トラフィックのタグ付けを解除して削除することもできます。ホスト接続用のポートを追加する方法の詳細については、「Ethernet 構成」を参照してください。レプリケーション トラフィックを有効化して拡張する方法については、ホワイト ペーパー『PowerStore：レプリケーション テクノロジー』を参照してください。

メモ： NAS インターフェイスは、システム ボンド インターフェイスに永続的に固定されており、PowerStore システムの他のポートに移動したり割り当てたりすることはできない。

Dynamic Resiliency Engine（DRE）

エンタープライズ クラスのストレージ システムには、高レベルの信頼性に加え、データ ロスや潜在的なドライブ障害からの保護が必要です。従来のデータ保護スキームは、ボリューム データを保護する固定レイアウトの RAID グループに基づいている。この従来型の設計では、帯域幅と再構築速度は、グループに参加するドライブの数によって制限されます。また、再構築の速度は、その RAID レベルのドライブで許容される障害の数によっても制限されます。たとえば、6 + 2 では、6 台のドライブの読み取り速度を達成できるが、デュアル ドライブに障害が発生した場合は、2 台のドライブに相当する再構築速度しか維持できない。

保護されるデータの信頼性は、ドライブのビット エラー レート(BER)、再構築する必要のあるデータの量、ドライブの数に大きく依存します。容量とドライブ数が増えると、従来の RAID 保護スキームを使用した場合に信頼性を維持することは困難になります。

さらに、経済性と信頼性は購入時の重要な決定要因となる。ストレージ ソリューションのコストは、ドライブのコストによって決まり、最終的には、パフォーマンスと実効容量に基づく費用対効果（IOPS あたりの金額と実効容量あたりの金額）によって決まります。ストレージとドライブの容量が増加するに従い、以下の事象が発生します。

- パフォーマンスが拡張する

- コントローラーの相対コストが減少する
- ドライブ障害が発生する可能性が高くなる
- 保護とシステム メタデータのオーバーヘッドが増加し、実効ストレージ容量に影響を与える
- 信頼性を維持するために、再構築速度の向上が必要となる

DRE の概要

PowerStore Dynamic Resiliency Engine (DRE)は、冗長性に対する 100%ソフトウェア ベースのアプローチで、従来の RAID よりも分散性、自動化、効率性に優れています。RAID 6 や RAID 5 のパリティ要件を満たし、優れた耐障害性と低コストを実現しています。PowerStore が実装している独自のアルゴリズムにより、すべてのドライブが複数の仮想チャンクに分割され、複数のドライブにまたがるチャンクを使用して冗長エクステントが作成されます。

また、アプライアンス内のドライブが自動的に利用され、すべてのドライブを使用して適切な冗長性が確保される。このプロセスにより、全体的なパフォーマンスが向上し、アプライアンスにドライブが追加されるにつれてパフォーマンスを拡張できるようになります。1 つのボリュームに書き込まれたデータは、アプライアンス内の任意の数のドライブに分散させることができる。新しいドライブが追加されると、データは自動的に再バランシングされます。

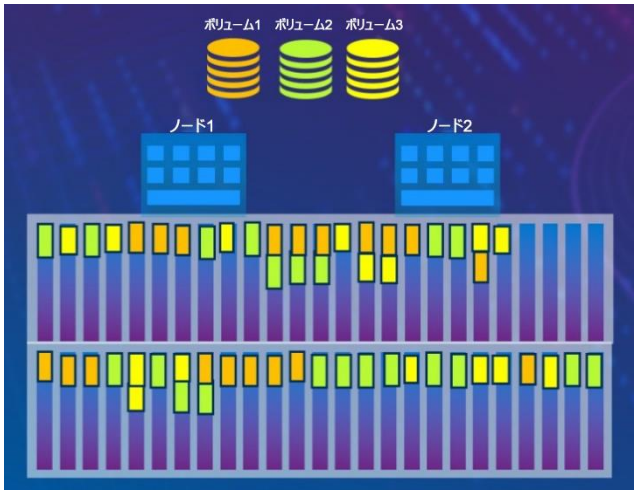

図26. PowerStore でのデータ配置

分散スペアリング

従来の RAID 保護戦略とは異なり、PowerStore には専用のスペア ドライブが必要とならない。スペア領域はアプライアンス全体に分散され、ドライブに障害が発生した場合のスペアリングに使用するための小さな領域が各ドライブから予約されます。アプライアンスの耐久性セットごとに、ドライブ 1 台分に相当するスペア領域が予約されます。耐久性セットについては、このセクションで後ほど説明します。

ドライブで障害が発生すると、データが書き込まれたドライブの部分のみが再構築される。これにより、必要な領域のみを利用して、スペア容量を効率的に管理できるようになる。この機能により、ドライブに書き込まれたデータしか再構築しなくてよくなるため、再構築時間も短縮されます。

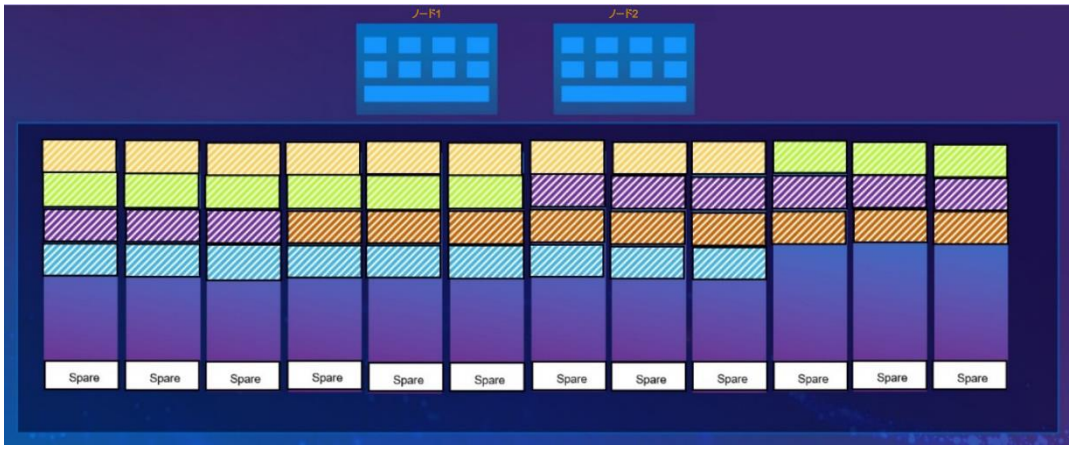

図27. 分散スペアリング

耐久性セット

PowerStore は、スペア オーバーヘッドを最小限に抑えながら信頼性を向上させるために、耐久性セットを実装しています。複数の障害ドメイン（耐久性セットとも呼ばれる）があると、障害が同時に発生した場合に、アプライアンスが各セット内のドライブ障害を許容できるため、システムの信頼性が向上します。

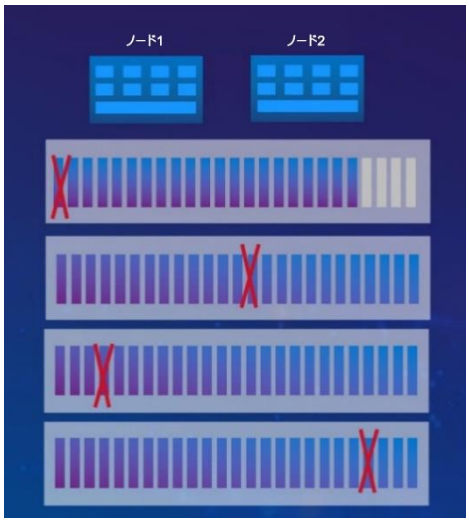

図28. 複数の耐久性セットにおける同時ドライブ障害に対するトレランス

複数の異なるインスタンスで障害が発生した場合（最初の障害が発生したドライブの再構築が完了した後に 2 番目のドライブで障害が発生した場合）、アプライアンスは、同じ耐久性セット内でも複数のドライブ障害に耐えることができます。PowerStoreOS 2.0 以降では、アプライアンスの初期構成時に、25 ドライブ耐久性セットの単発ドライブ障害トレランスまたは 50 ドライブ耐久性セットの二重ドライブ障害トレランスを選択できます。マルチアプライアンス クラスターを導入する場合、アプライアンス上で異なるドライブ障害トレランスを混在させることができます。次の図に示すように、アプライアンス A を単発ドライブ障害トレランスに設定し、アプライアンス B を二重ドライブ障害トレランスに設定できます。

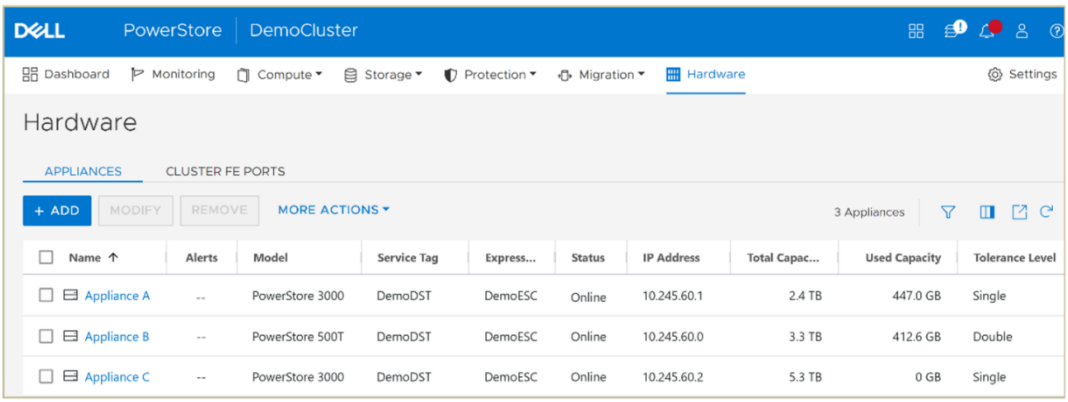

図29. アプライアンスのフォールト トレランス レベル

時間の経過とともにシステムに容量とドライブが追加されるにつれ、耐久性セットは動的に増加します。たとえば、単発ドライブ フォールト トレランスが設定されたアプライアンスがあるとします。これは、システムが 25 ドライブ耐久性セットを使用して構成されていることを意味する。26 番目のドライブをシステムに追加すると、耐久性セットは動的に 2 つのセットに分割されます。さらに、耐久性セットは、アプライアンス内のドライブ数に基づいて複数の物理エンクロージャにまたがることができ、ドライブ サイズを混在させることができる。

耐久性セットには、主に次のようなメリットがあります。

- エンタープライズ クラスの可用性
  - 分散スペアリングによる再構築時間の短縮
  - ドライブを小さなチャンクに分けた再構築と、アプライアンス内の複数のドライブの再構築を同時に進行

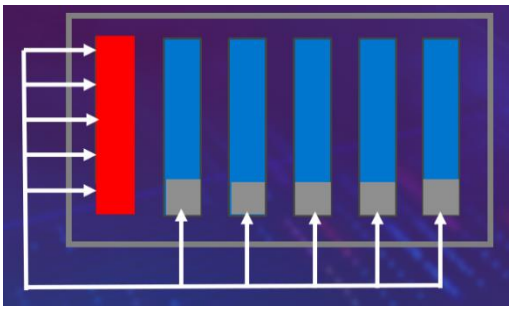

図30. 単一ドライブの再構築と領域の分散スペアリングを並列処理

- インテリジェントなインフラストラクチャ
  - 未使用のユーザー領域を自動的に割り当てて、複数の障害に対処するためのスペア領域を補充

    アプライアンスで使用可能な未使用容量が十分にある場合、DRE は未使用ユーザー容量を動的に転送して、スペア容量を補充する。

  - 可用性を維持しながら、着信 IO トラフィックに基づいて再構築速度をインテリジェントに変更

    PowerStore は機械学習アルゴリズムを使用して、ドライブに障害が発生した場合にホスト IO を優先するように再構築速度を自動的に調整し、信頼性を維持しながらパフォーマンスを最適化します。

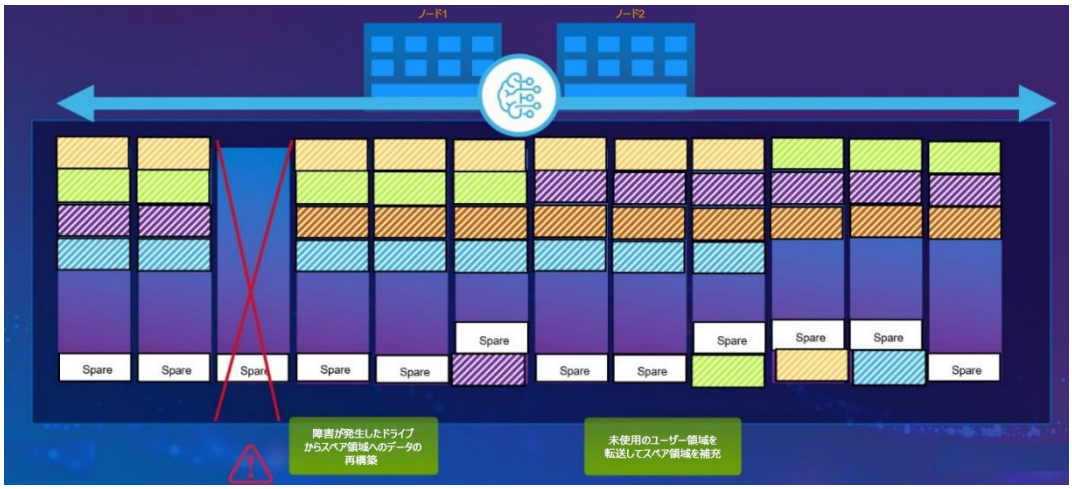

図31. ドライブ障害後にデータ チャンクをスペア領域に再構築し、スペア領域を未使用のユーザー領域で補充

- 柔軟な構成
  - ドライブを 1 台ずつ追加してストレージを拡張することで TCO を削減

    アプライアンスには最小で 6 台のドライブを搭載でき、最大 96 台のドライブまで拡張できます。容量は中断なしで追加でき、お客様は、必要に応じて 1 台または複数のドライブを追加することにより、ストレージを柔軟に拡張できます。

  - ストレージ ニーズに基づいてさまざまなドライブ サイズを追加できる柔軟なオプション

    PowerStore は、複数のドライブにわたって冗長エクステントの分散を最適化することにより、さまざまなサイズのドライブを管理する独自のアルゴリズムを実装している。

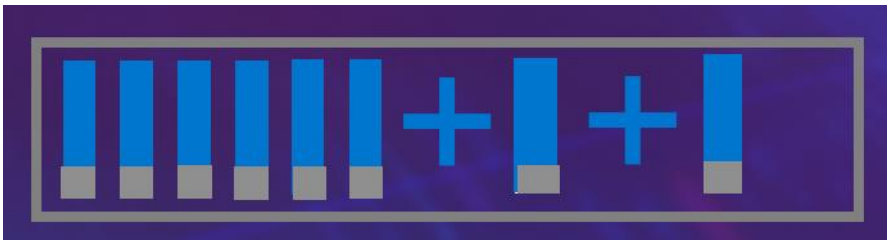

図32.単一ドライブ拡張

ブロック ストレージ

PowerStore アーキテクチャは、完全にアクティブ/アクティブなシン プロビジョニングされたマッピング サブシステムを実装しているため、任意のパスからの IO をコミットし、リダイレクトを必要とせずに完全な整合性を維持できます。この実装の主な利点は、一部のパスが使用できなくなった場合に、I/O 要求を処理する時間を短縮できることである。このアプローチはリダイレクトと単一ノード ロックを使用する競合アーキテクチャとは対照的です。競合アーキテクチャでは、トレスパス時間が長くなるため、フェールオーバー中にデータ欠損になる可能性が高くなります。

PowerStore は、SCSI ベースのホスト アクセスに非対称論理ユニット アクセス(ALUA)を使用し、NVMe-oF ホスト アクセスに非対称ネームスペース アクセス(ANA)を使用する、デュアルノードの共有ストレージ アーキテクチャを実装します。任意のパス（アクティブ/最適化またはアクティブ/非最適化）で受信した I/O 要求は、ローカルでコミットされ、ピア ノードと完全に整合します。I/O は通常、アクティブ/最適化パスで送信されます。ただし、すべてのアクティブ/最適化パスが使用不可になるというまれな状況では、アクティブ/非最適化パスを使用して I/O 要求がローカル ノードで処理され、ピア ノードにデータを送信する必要がありません。次に、I/O は共有 NVRAM 書き込みキャッシュに書き込まれ、そこで重複排

除および圧縮されてからドライブに書き込まれます。ドライブへの I/O の書き込み方法の詳細については、『PowerStore：データ効率性』を参照してください。

ALUA および ANA マルチパスにより、高可用性が確保されます。また、基盤となる完全に対称的なシンプロビジョニング アーキテクチャは、正常に稼働しているパスでボリュームを使用可能にすることに伴う複雑さとオーバーヘッドを排除して、I/O の処理時間を短縮します。ALUA と ANA を使用するには、Dell PowerPath などのマルチパス ソフトウェアをホストにインストールする必要があります。マルチパス ソフトウェアは、最適化されたパスをまず使用し、最適化された使用可能なパスがない場合にのみ、最適化されていないパスを使用するように構成する必要があります。可能な場合は、2 つの別々のネットワーク インターフェイス カード（NIC）または Fibre Channel ホスト バス アダプター（HBA）をホストで使用する。こうすることで、カードとサーバー カード スロットの単一障害点を回避できる。

物理ポートは両方のノードで常に一致している必要があるため、フェールオーバーが発生した場合のホスト アクセスには、常に同じポート番号が使用されます。たとえば、ノード A の 4 ポート カード ポート 3 がホスト アクセスに現在使用されている場合は、障害が発生すると、ノード B で同じポートが使用される。そのため、マルチパスと冗長性を確保するには、両方のノードで同じポートをホストの複数のスイッチに接続する必要があります。

Ethernet 構成

PowerStoreOS 2.1 では、PowerStore T モデル アプライアンスで NVMe/TCP のサポートが導入されました。これにより、ユーザーは iSCSI または NVMe/TCP ホスト接続用の Ethernet インターフェイスを構成することができます。PowerStore X モデル アプライアンスについては、NVMe/TCP ホスト接続は PowerStoreOS 2.1.1 以降でサポートされています。Ethernet インターフェイスはフェールオーバーしないため、インターフェイスはミラーリング ペアで両方の PowerStore ノードに導入できます。この構成により、1 つのノードが使用不可になった場合でも、ホストはブロックレベルのストレージ リソースに継続的にアクセスできるようになる。PowerStore T モデル アプライアンスの場合、ストレージ ネットワークはクラスターの作成後に構成します。堅牢な HA 環境を実現するには、クラスター作成後に他のポートで追加のインターフェイスを作成する。

PowerStore X モデル アプライアンスの場合、ICW の実行時に iSCSI または NVMe/TCP を有効にします。図 33 に示すように、［**ストレージ ネットワーク**］ページに関する IP アドレス情報を入力します。これらの IP アドレスは、ノード A とノード B の 4 ポート カードの最初の 2 つのポートに割り当てられます。

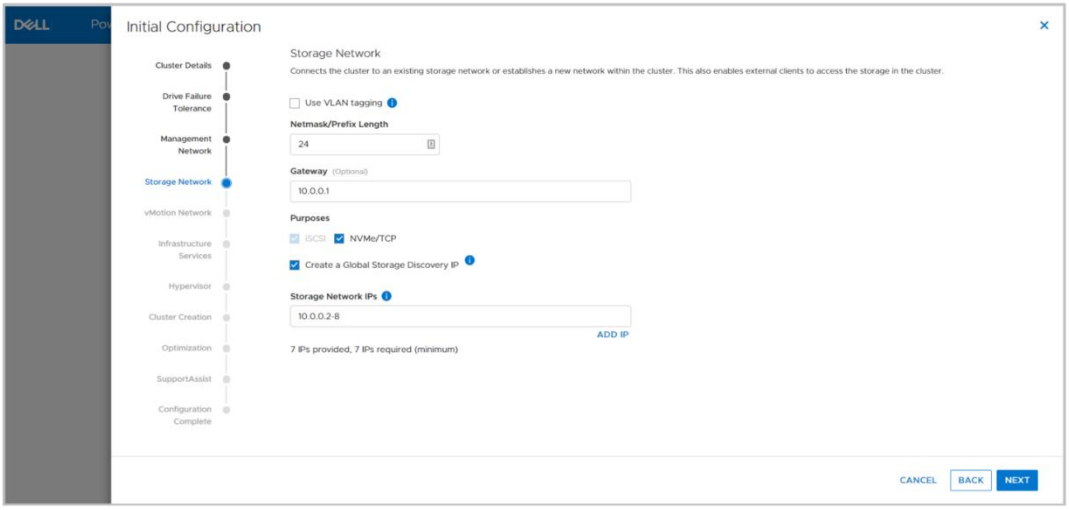

図33. PowerStore X モデル アプライアンスのストレージ ネットワーク

ホストおよびスイッチ レベルで単一障害点を排除するには、PowerStore システムの 4 ポート カードの最初の 2 つのポートがスイッチにケーブル接続されていることを確認します（次の図を参照）。

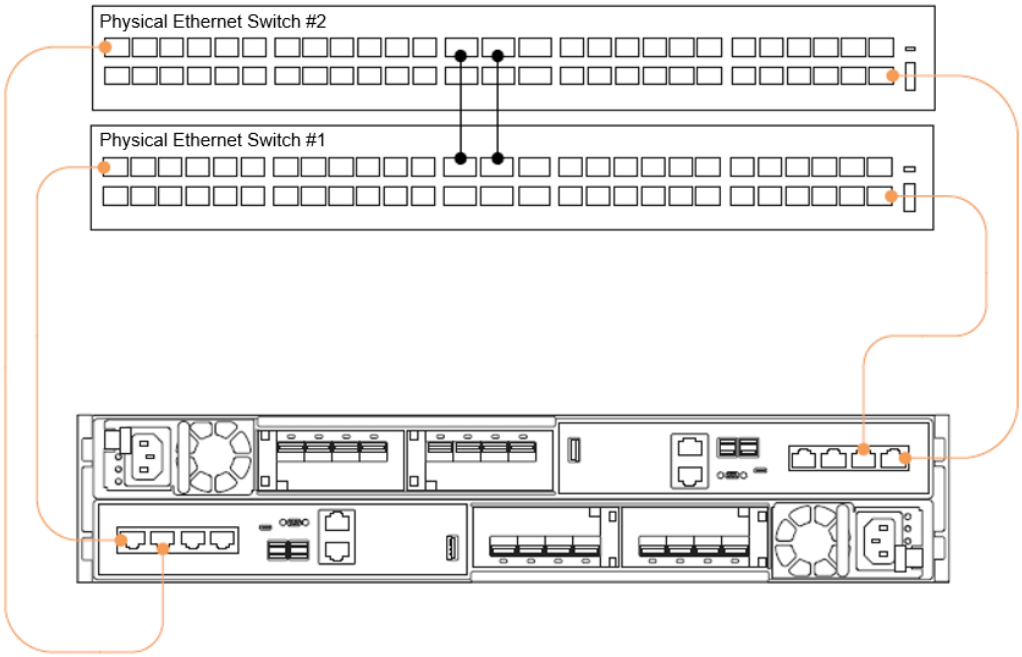

図34. トップ オブ ラック スイッチにケーブル接続された 4 ポート カードの最初の 2 ポート

アプライアンスを構成したら、追加の帯域幅、スループット、専用ホスト接続、専用レプリケーション トラフィックなどのための他のポートを有効にします。これらのポートについては、4 ポート カードまたは I/O モジュールで使用可能な他のポートをケーブル接続する。物理ポートは両方のノードで一致しているので、ホストのマルチパスと冗長性を確保するために、複数のスイッチにポートがケーブル接続されていることを確認します。レプリケーション トラフィック用の追加のインターフェイスを有効にする方法の詳細については、『PowerStore：レプリケーション テクノロジー』を参照してください。

ケーブル接続が完了したら、PowerStore Manager で IP アドレスを追加する。図 35 は、ストレージ ネットワーク上に未使用の IP アドレスが構成されていることを確認するための最初のステップを示している。使用可能な IP が他にない場合は、［**Add**］ボタンをクリックして IP アドレスを追加する。次の図は、ホスト接続用の追加ポートをマップするための、［Ports］ページでの次のステップを示しています。［Ports］ページは、モデルのタイプによって異なり、次のようにしてアクセスできる。

- PowerStore T モデル：［**ハードウェア**］＞［**アプライアンスの詳細**］＞［**ポート**］
- PowerStore X モデル：［**ハードウェア**］＞［**アプライアンスの詳細**］＞［**ポート**］＞［**仮想**］

これらのポートは、次に使用可能な未使用 IP アドレスをストレージ ネットワークから取得して、新しく構成されたポートに割り当てる。この例に示すとおり、ホスト トラフィック用に割り当てられている新しく構成されたポートは、ミラーリング ペアで有効になります。ポートを 1 つだけ選択した場合は、ピア ノードの関連ポートでもホスト トラフィックが有効になっていることを示す通知が表示される。PowerStore X モデル アプライアンス用の仮想ポートの詳細については、『PowerStore：仮想化の統合』を参照してください。

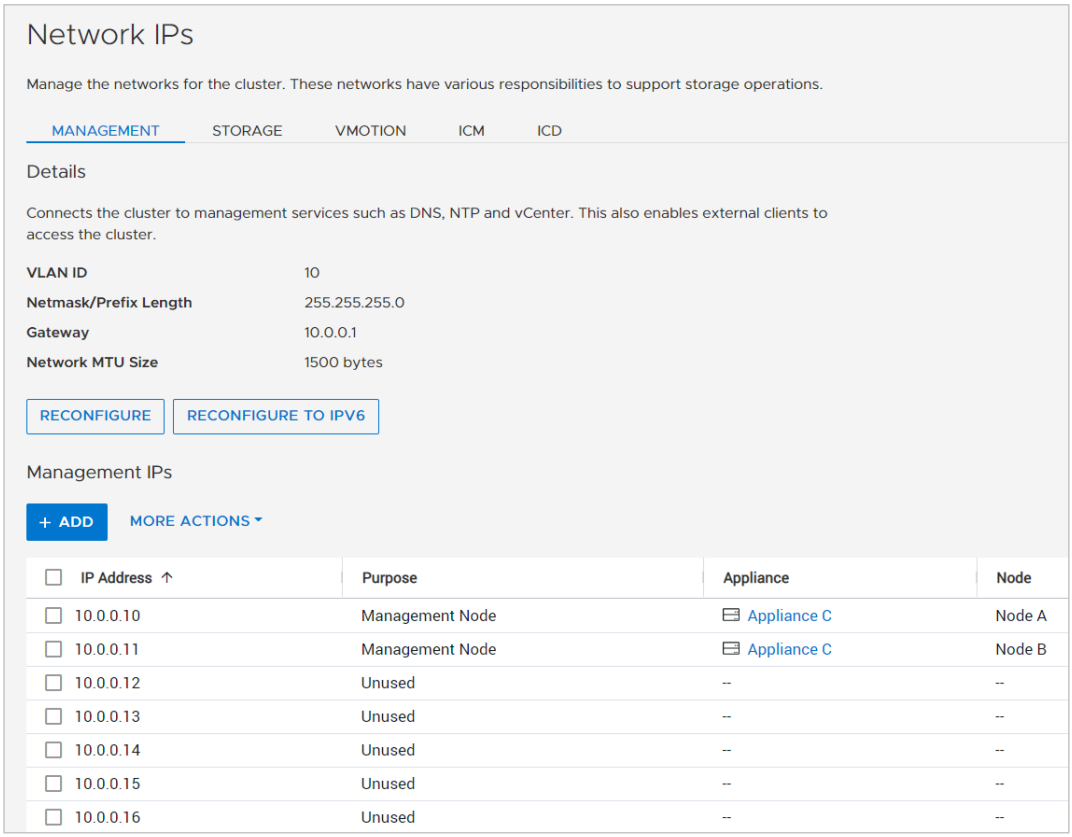

図35. 未使用のストレージ ネットワーク IP アドレスを確認する

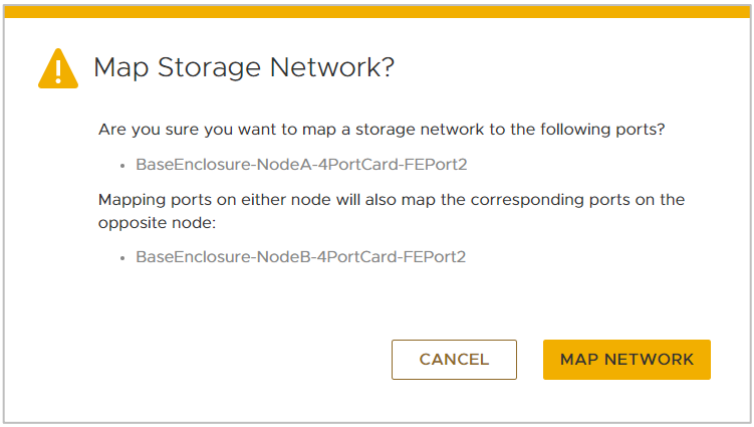

図36. ストレージ ネットワークのマッピング

Fibre Channel 構成

Fibre Channel (FC)および NVMe/FC で高可用性を実現するには、各ノードへの接続を少なくとも 1 つ構成します。こうすることで、1 つのノードが使用不可になった場合でも、ホストはブロックレベルのストレージ リソースに継続的にアクセスできるようになる。

FC と NVMe/FC の場合、ホストと PowerStore アプライアンス間の通信を可能にするために、スイッチでゾーニングを構成する必要があります。FC を使用しているアプライアンスがクラスター内に複数ある場合は、ホスト通信を有効にするように構成されたゾーンが各 PowerStore アプライアンスにあることを確認する。クラスター内の各アプライアンスの各ノード FC ポートに対して、ホスト HBA ポートごとのゾーンを

作成する。堅牢な HA 環境を実現するには、ゾーニングする FC ポートを増やして、PowerStore アプライアンスにより多くの追加パスを提供します。

PowerStore Manager では、2 つの場所でポートの World Wide Name (WWN)を確認できます。次の図に示すように、1 つは［**ハードウェア**］＞［**クラスターFE ポート**］ページです。図 38 に示すように、もう 1 つは［**コンポーネント**］＞［**背面図**］ページです。ホスト構成の詳細については、dell.com/powerstoredocs の『PowerStore ホスト構成ガイド』を参照してください。

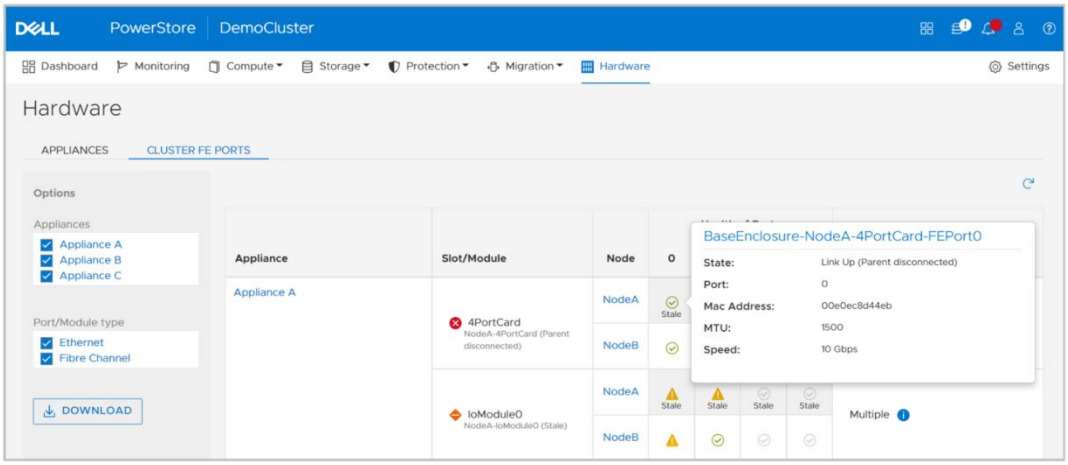

図37. PowerStore Manager のハードウェア クラスターのフロントエンド ポート ページ

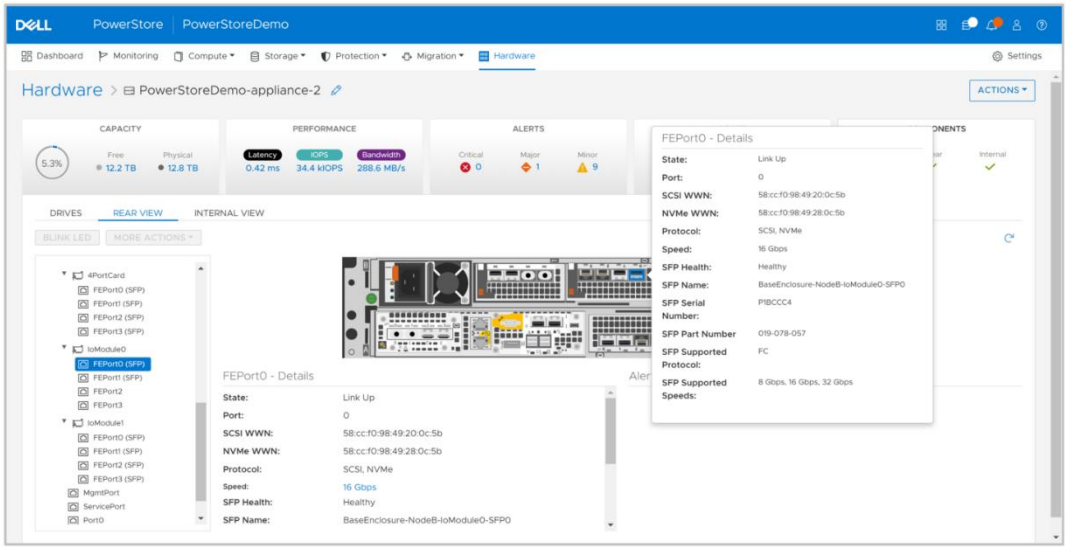

図38. PowerStore Manager のコンポーネント背面図

ブロックの例

高可用性インフラストラクチャを設計する場合は、ストレージ システムに接続するコンポーネントも冗長でなければならない。この設計には、接続の問題によるデータ欠損を回避するために、ホストおよびスイッチ レベルで単一障害点を排除することが含まれる。次の図は、単一障害点を持たない PowerStore T モデル システムの高可用性構成の例を示しています。ネットワーク インフラストラクチャのケーブル接続と構成の詳細については、dell.com/powerstoredocs の『PowerStore ネットワーク プランニング ガイド』および『PowerSwitch シリーズ向け PowerStore ネットワーク構成ガイド』を参照してください。

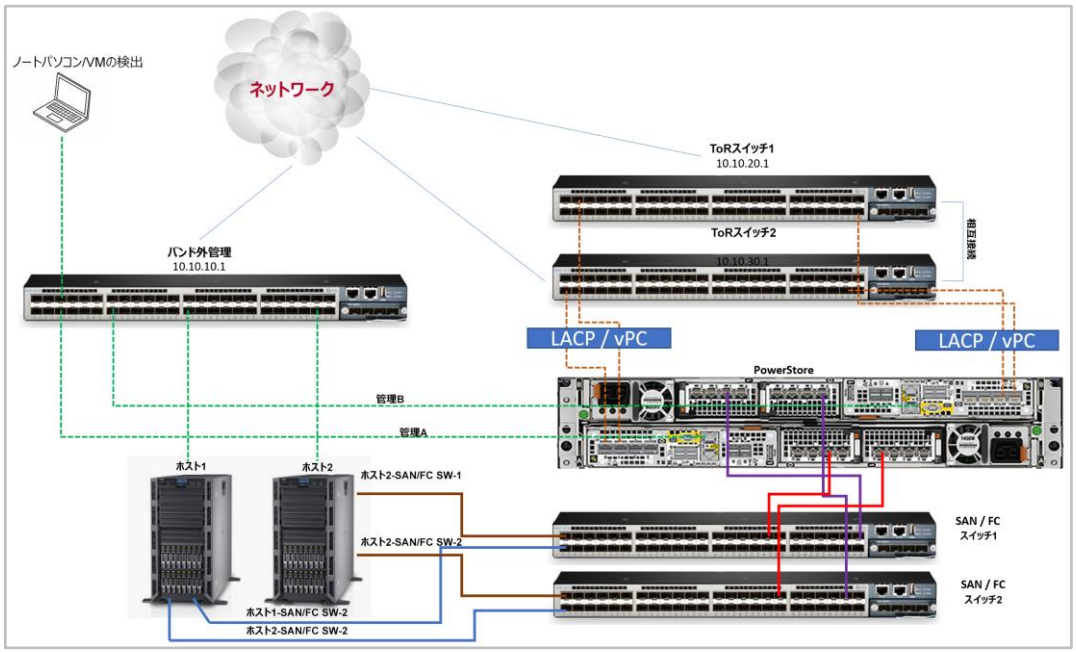

図39. 高可用性ブロック構成

ファイル ストレージ

PowerStore T モデル システムのファイルレベルのリソースを共有する前に、4 ポート カードの最初の 2 つのポートでデフォルトのシステム ボンドを使用するか、ユーザー定義のリンク アグリゲーション グループを使用して、NAS サーバー インターフェイスを構成する必要があります。PowerStoreOS 3.0 以降では、ユーザー定義のリンク アグリゲーション グループを使用できます。リンク アグリゲーションの詳細については、「LACP（Link Aggregation Control Protocol）」を参照してください。

管理者は、ファイル システムへの SMB、NFS、FTP、SFTP アクセスに関する構成情報を保持する NAS サーバーを作成できます。新しい NAS サーバーは、使用可能なノードにラウンド ロビン方式で自動的に割り当てられる。優先ノードは、このアルゴリズムに基づいて、NAS サーバーが実行されているノードを示すマーカーとして機能します。プロビジョニング後は、NAS サーバーの優先ノードは変更されません。現在のノードは、NAS サーバーが実行されているノードを示します。現在のノードを変更すると、NAS サーバーが別のノードに移動し、このノードを負荷分散の目的で使用できるようになります。NAS サーバーを新しいノードに移動すると、その NAS サーバー上のすべてのファイル システムもともに移動されます。

次の図は、PowerStore Manager の現在のノードの列と優先ノードの列を示しています。

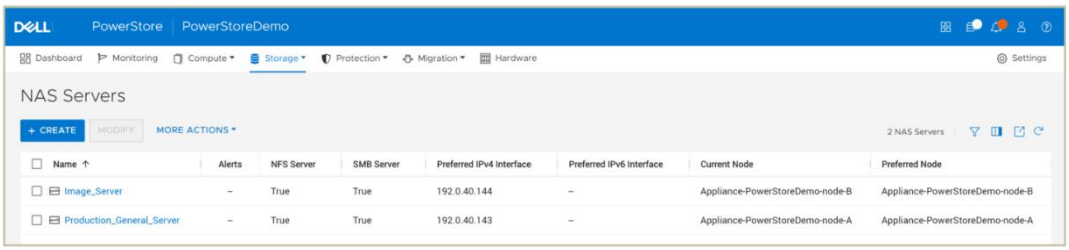

図40. 現在のノードと優先ノード

PowerStore ノードで障害が発生すると、NAS サーバーは正常に稼働しているノードに自動的にフェールオーバーする。ホストのタイムアウトを回避するため、このプロセスは通常 30 秒以内に完了する。障害が発生したノードがリカバリーされたら、NAS サーバーをフェールバックしてバランスの取れた構成に戻すための手動プロセスが必要となります。

NAS サーバーは、アップグレード プロセス中に自動的にピア ノードに移動される。アップグレードが完了すると、NAS サーバーはアップグレードの開始時に割り当てられていたノードに戻ります。NAS サーバーの詳細については、『PowerStore：ファイル機能』を参照してください。

LACP（Link Aggregation Control Protocol）

リンク ロスは、ケーブルやスイッチ ポートの障害など、多数の環境的要因によって発生する可能性があります。そのようなタイプの障害から保護するには、ポートに高可用性を構成します。

リンク アグリゲーションでは、複数のネットワーク接続を 1 つの論理リンクに統合します。これにより、複数の接続にトラフィックを分散することでスループットを増やすことができます。また、1 つの接続が失敗した場合に備えて冗長性を提供することもできます。接続の喪失が検出されると、そのリンクは即時に無効になり、アグリゲーション内でまだ機能しているリンクにトラフィックが自動的に移動され、システムの停止が回避されます。接続が回復するとアグリゲーション内にポートが戻されるように、スイッチを適切に構成する必要があります。リンク統合によってより全体的な帯域幅を得られますが、それでもなお個々のクライアントは同じ単一のポートを通ります。Dell PowerStore システムでは、Link Aggregation Control Protocol (LACP) IEEE 802.3ad 標準が使用されます。

NAS サーバーには、クライアントにアクセスするために Ethernet ポート上に作成された、1 つ以上のネットワーク インターフェイスが含まれています。3.0 より前の PowerStoreOS バージョンでは、LACP は、組み込みモジュールの 4 ポート カード上のシステム ボンド ポート（ポート 0 および 1）を介してのみ使用できます。PowerStoreOS 3.0 以降では、異なる I/O モジュールの 2～4 個のポートを備えた PowerStore T システムでリンク アグリゲーションを構成できます。また、組み込みモジュールの I/O モジュールと Ethernet ポートの間にもリンク アグリゲーションを構成できます。統合内のすべてのポートは、速度、二重モードの設定、MTU のサイズが同じである必要があります。リンク アグリゲーションは、PowerStore Manager の［ハードウェア］＞［アプライアンス］＞［ポート］ページで作成できます（図 41）。

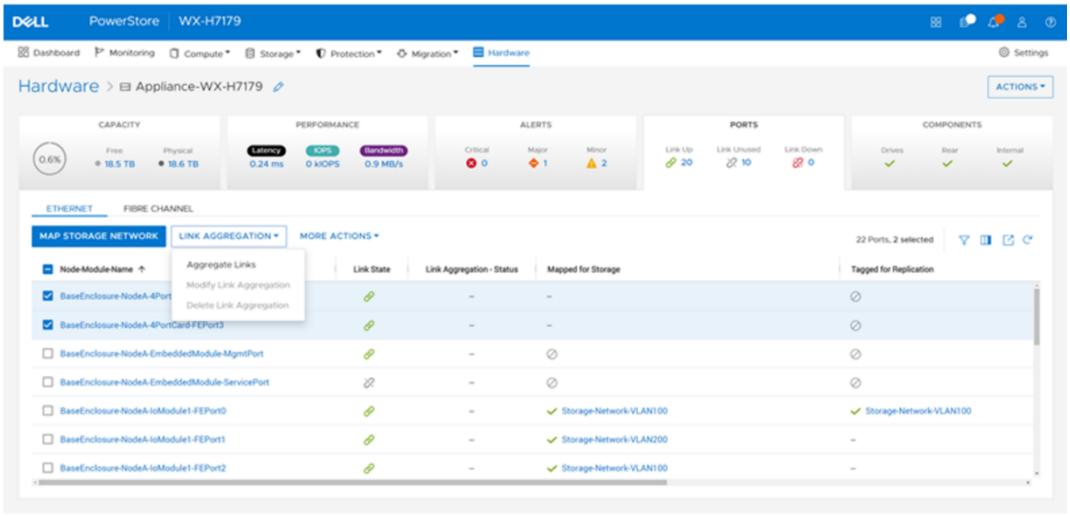

図41. リンク アグリゲーションの作成

リンク アグリゲーションは、NAS サーバーとファイル インポート インターフェイスに対して使用できます。PowerStore は、同じインターフェイスでのプロトコルの共有をサポートしているため、iSCSI、レプリケーション、インポート トラフィックに同じ個別ポートを使用しながら、複数のポート間でリンク アグリゲーションを行うことができます。ブロック アクセスにはマルチパスが使用されるため、iSCSI トラフィックをリンク アグリゲーションに直接マッピングすることはできません。ベスト プラクティスとして、最適な帯域幅とパフォーマンスを維持するために、可能な場合はブロック トラフィックとは別のポートでリンク アグリゲーション トラフィックを維持することをお勧めします。

リンク アグリゲーションを構成する際は、同じポートがシステムの両方のノードでケーブル接続されていることを確認します。これが必要なのは、フェールオーバーの場合に、ピア ノードが同じポートを使用するからです。また、ノードに接続されている適切なスイッチ ポートがリンク アグリゲーション用に構成されていることを確認します。スイッチが正しく構成されていない場合や、ケーブル接続が一致しない場合、通信の問題が発生する可能性があります。

PowerStore のポートが同じスイッチに接続している場合も、異なるスイッチに接続している場合も、単一のリンク アグリゲーションを使用できます。スタックされたスイッチの場合、リンク アグリゲーションは冗長性の目的で複数のスイッチへの複数のパスを提供します。図 42 は、システムの最初の 2 つのポート（4 ポート カードのポート 0 および 1）のデフォルトのシステム ボンド LACP 構成を示すケーブル接続図です。ノード A の各ポートは別の Dell スイッチに接続し、構成はノード B にミラーリングされています。スタックされたスイッチの場合は、VLT または別のスイッチ ベンダーの同等のテクノロジーを構成する必要があります。リンク アグリゲーション グループは、次の例の緑色のパスとは別個に、青色のパスで構成されます。

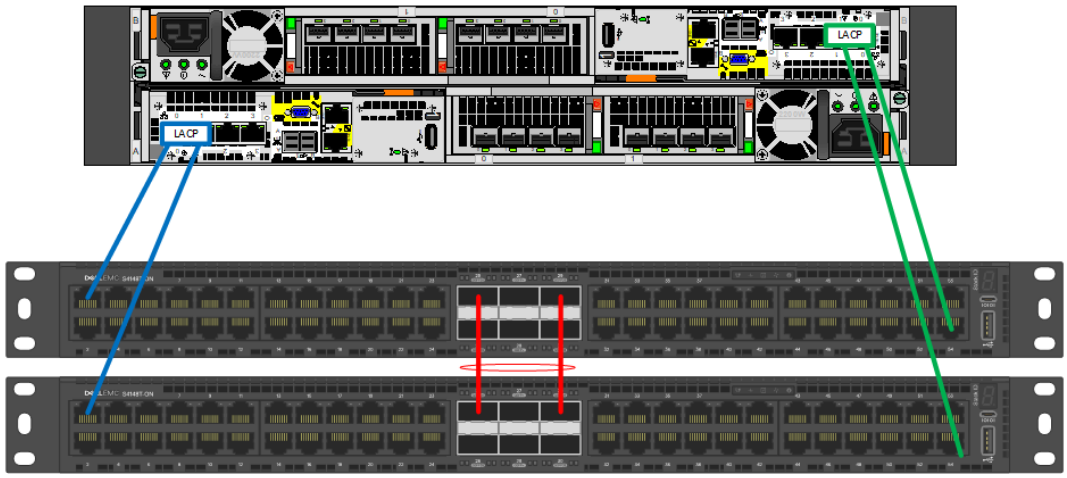

図42. リンク アグリゲーションの例 1

相互接続されていないスイッチを含むネットワークの場合、1 つのリンク アグリゲーションを使用して 1 つのネットワーク スイッチに冗長性を提供し、2 つ目のアグリゲーションを使用して別のネットワーク スイッチに冗長性を提供することができます。図 43 では、4 ポート カードのポート 0 と 1 に最初のリンク アグリゲーション（青色）が作成され、上のスイッチに接続されています。2 つ目のリンク アグリゲーション（緑色）は、4 ポート カードのポート 2 と 3 に作成され、下のスイッチに接続されています。次に、4 つのケーブル グループごとに、リンク アグリゲーション グループを構成する必要があります。ノード A では、リンク アグリゲーション グループは、下のスイッチに接続する緑色のパスとは別個に、上のスイッチに接続する青色のケーブルで構成されます。ノード B のポートについても、これを繰り返します。

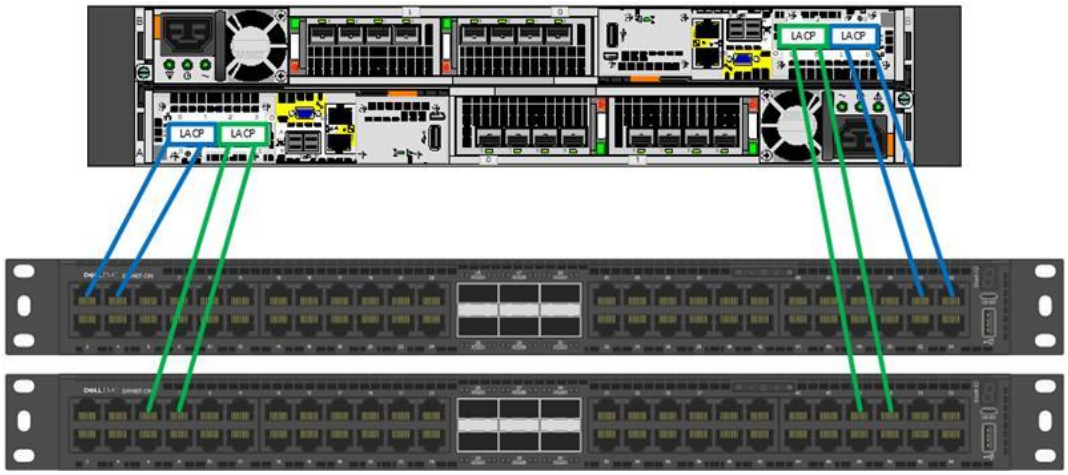

図43. リンク アグリゲーションの例 2

あるシステムから別のシステムにレプリケートする場合は、両システムに同じ方法でリンク アグリゲーションを構成することをお勧めします。同じ名前のリンク アグリゲーションが宛先システムで見つからない場合、宛先 NAS サーバー上のインターフェイスはポートの割り当てなしで作成されます。この場合、ユーザーが手動でポートを割り当てる必要があります。また、一致しない構成が必要な場合は、宛先 NAS サーバー上のインターフェイスをオーバーライドして、有効なポートに割り当てることができます。そうしなければ、フェールオーバーが実行されるとデータにアクセスできなくなります。

リンク アグリゲーションをホスト レベル/クライアント レベルで構成し、ポートやケーブルの障害に対する耐久性を確保する必要もあります。ベンダーによっては、リンク アグリゲーションをトランキング、ボンディング、または NIC チーミングと呼ぶ場合があります。詳細については、ベンダーのマニュアルを参照してください。

VMware 統合

次のサブセクションでは、PowerStore X モデルを使用したいくつかの HA 機能について説明します。PowerStore と VMware の統合の詳細については、『PowerStore：仮想化の統合』を参照してください。

コントローラーVM

PowerStore X モデル アプライアンスは、コントローラーVM の形式のノードで完全に冗長化されている。各ノードには、データ サービスを提供するコントローラーVM がある。各コントローラーVM は、ノード内の物理 M.2 デバイスを表すプライベート データストアに導入される。これらのデータストアは共有されていないため、ノードに障害が発生した場合、コントローラーVM はクラスター内の他のホストで再起動しません。ただし、PowerStore は完全に冗長であるため、他のコントローラーVM を実行しているピア ノードは、アプライアンスへの I/O を処理し続ける。2 つのコントローラーVM は、アプライアンスの HA ペアとして機能する。コントローラーVM を移動したり、設定を変更したりしないようにする必要があります。

vSphere HA

管理者が PowerStore T または PowerStore X モデル アプライアンスのいずれかで仮想マシンを使用している場合は、クラスターで vSphere HA を有効にすることをお勧めします。クラスター内のノードまたはホストの電源が失われた場合、仮想マシンは、クラスター内の正常に稼働しているホストまたはノードで再起動します。

vCenter 接続

PowerStore T モデル アプライアンスと PowerStore X モデル アプライアンスはいずれも、vCenter から管理および監視されることにより、VMware 環境との統合を実現します。PowerStore T モデル アプライアンスの場合、クラスターのセットアップ後の vCenter 接続はオプションである。PowerStore X モデル アプライアンスの場合、初期構成ウィザードの実行中に vCenter 接続が必要となる。クラスターが導入された後、PowerStore アプライアンスへの vCenter 接続が失われる場合がある。このような場合、vCenter を介した管理タスクが一時的に利用できなくなる可能性があります。PowerStore アプライアンス内のすべてのストレージ オブジェクトの I/O は、引き続き処理されます。

レプリケーション

システム レベルまたはデータセンター レベルでのアウテージからデータを保護するために、リモート サイトへのレプリケーションを使用することをお勧めします。このユース ケースには、計画的なメンテナンス イベント、計画外の停電、自然災害などが含まれます。PowerStore は、ブロック リソースのディザスター リカバリーをシンプルにするレプリケーションをサポートしています。PowerStoreOS 3.0 以降では、レプリケーション用にファイル リソースもサポートされています。PowerStoreOS 3.0 では、PowerStore アプライアンスはネイティブの metro ボリューム レプリケーションをサポートしています。これにより、metro 距離にある VMFS データストアの 2 つの PowerStore クラスターにまたがるブロック ストレージ ボリュームの同期レプリケーションが実現されます。PowerStore でレプリケーション機能をセットアップして使用する方法の詳細については、デル・テクノロジーズ情報ハブのホワイト ペーパー『Dell PowerStore：レプリケーション テクノロジー』および『Dell PowerStore：metro ボリューム』を参照してください。

プラットフォームの高可用性

PowerStore では、負荷分散と冗長性のため、各ストレージ リソースがノード A またはノード B のいずれかに割り当てられる。ストレージ リソースの割り当てに加えて、各ノードでは、PowerStore オペレーティング システムを構成するさまざまなコンテナが実行されている。1 つのノードが使用不可になった場合、そのリソース（ストレージとコンテナ）は、正常に稼働しているノードに自動的にフェールオーバーする。

フェールオーバー プロセスが完了するまでにかかる時間は、システムの使用率やリソースの数などの要因によって異なる。ピア ノードはリソースの所有権を引き継ぎ、長時間の停止を回避するために I/O の処理を継続する。フェールオーバーは、ノードで次のような状況が発生した場合に行われる。

- ノードの再起動：システムまたはユーザーがノードを再起動した。
- ハードウェアまたはソフトウェアの障害：ノードで障害が発生したため、交換する必要がある。
- サービス モード：システムまたはユーザーがノードをサービス モードに設定した。この状況は、ハードウェアまたはソフトウェアの問題が原因でノードが起動できない場合に自動的に発生する。
- 電源オフ：ユーザーがノードの電源をオフにした。

メモ：サービス モードへの手動でのノードの切り替えは、サービス スクリプトを使用した場合にのみ可能となり、PowerStore Manager、REST API、または PSTCLI を使用して行うことはできない。

ノードが使用不可になっている間、そのノードのすべてのリソースはピアによって処理される。ノードがオンラインに戻るか、障害が修正されると、ブロック ストレージ リソースは自動的に適切なノード所有者にフェールバックする。ファイル ストレージ リソースは、手動でフェールバックする必要がある。

コードのアップグレード中は、両方のノードが調整された方法で再起動する。再起動中のノード上のすべてのリソースは、ピア ノードにフェールオーバーされる。ピアがオンラインに戻ると、リソースは元の所有者にフェールバックされる。このプロセスが 2 番目のノードで繰り返される。ユーザーは、コード アップグレードを開始する前に、アップグレード前のヘルス チェックを実行することで、円滑なアップグレード プロセスを確保できる。

クラスターの高可用性

PowerStore アプライアンスは、クラスター リソースの管理に使用されるペースメーカー スタックを使用して、クラスターにクォーラムがあることを確認する。ペースメーカーはクラスターの司令塔として機能し、クラスターに対するアプライアンスの追加と削除などのクラスター イベントや、障害によって引き起こされるリソース イベントを処理し、それらのイベントに対応する。アプライアンスに障害が発生しても、クォーラムがある場合は、管理サービスが引き続き提供される。

クラスターのクォーラム

クォーラムは、アクティブな通信状態にある N/2+1 アプライアンスとして定義される。クォーラムがない場合、管理操作は一時的に失われますが、データは使用可能であれば引き続き提供されます。次の図は、Fibre Channel 接続を備えたホストに I/O を提供しているが、一時的に管理アクセスを失った、2 台のアプライアンスからなるクラスターの例を示しています。この例では、アプライアンス 1 とアプライアンス 2 の間にクォーラムはない。アプライアンスは Fibre Channel 接続を使用してホストに I/O を提供しているため、ホストへの影響はありません。ただし、クォーラムは失われるため、一部の管理サービスはクォーラムがリストアされるまで一時的に停止する。

- PowerStore Manager へのアクセス
- スケジュール設定されたスナップショットの実行
- レプリケーション セッションの同期

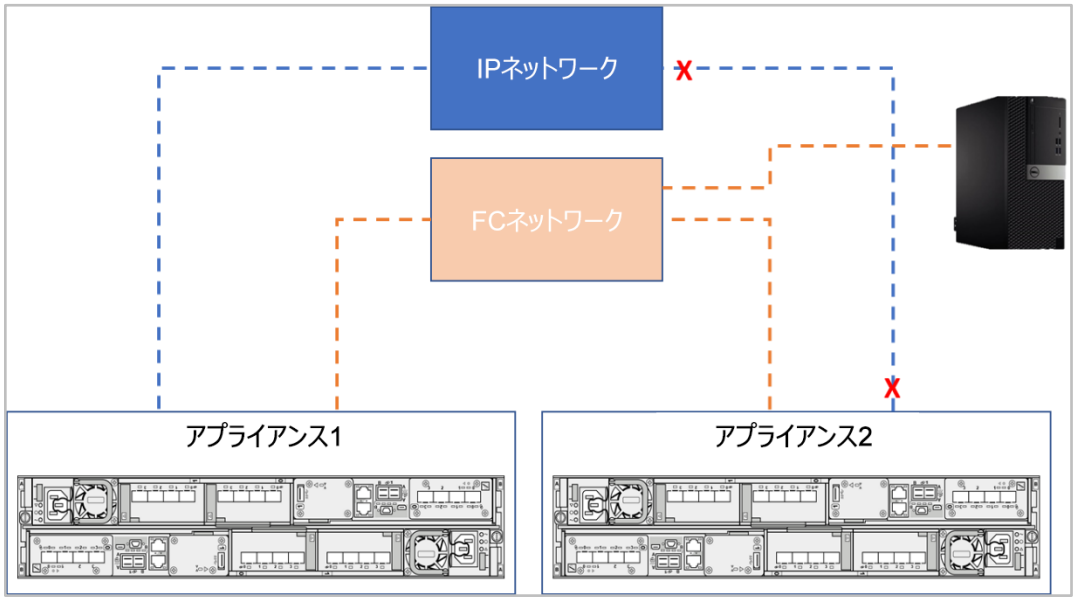

図44. 2 台のアプライアンスからなるクォーラムなしのクラスター

次の図は、3 台のアプライアンスからなるクラスターで、アプライアンス 3 がクラッシュした状態を示しています。このシナリオでは、2 台のアプライアンスがアクティブな通信状態であるため、クォーラムが満たされている。I/O は、アプライアンス 1 およびアプライアンス 2 から引き続きアクセスできる。ただし、アプライアンス 3 で障害が発生したため、そのアプライアンスの I/O にはアクセスできない。

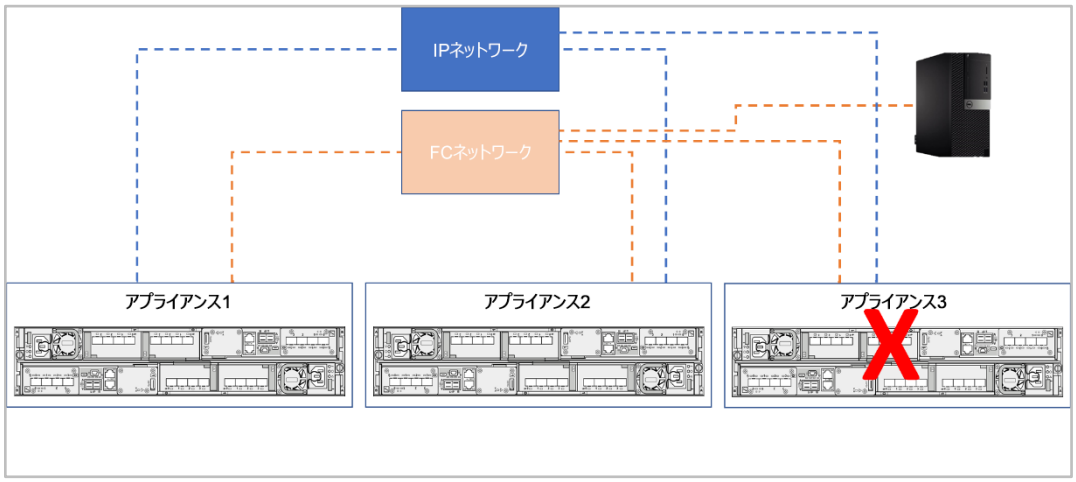

図45. 3 台のアプライアンスからなるクォーラムありのクラスター

クラスターIP

クラスターの IP アドレスは、クラスターの初期構成時に設定できる必須入力フィールドです。このクラスターは、PowerStore Manager へのアクセスに使用される高可用性 IP アドレスである。クラスターIP アドレスは通常、プライマリー アプライアンスのプライマリー ノードにあります。プライマリー アプライアンスでデュアルノード障害が発生するまれな状況では、クォーラムがある場合、クラスターIP アドレスはクラスター内の別のアプライアンスにフェールオーバーする。このクラスターIP アドレスがフェールオーバーすると、PowerStore Manager へのアクセスを回復するのに数分かかる場合があります。

グローバル ストレージ IP

グローバル ストレージ IP（GSIP）アドレスは、クラスターの初期構成中または初期構成後に設定できるオプションのフィールドであり、この IP アドレスは、グローバルなフローティング IP としてストレージの検出に使用される。iSCSI または NVMe/TCP ホストは、1 つの GSIP のみを必要とし、クラスター内のすべてのアプライアンスのすべてのストレージ パスを検出できます。それ以外では、いずれかの IP が停止した場合に、その次の IP を試すことができるように、iSCSI または NVMe/TCP ホストはストレージ IP のリストを要求します。

まとめ

高レベルの可用性を備えたインフラストラクチャを設計することで、ビジネスクリティカルなデータへの継続的なアクセスを確保できる。データが使用不可になると、日々の運用に影響が及び、生産性と収益が失われる可能性があります。PowerStore システムは、ハードウェア レベルとソフトウェア レベルの両方で、すべてのコンポーネントにわたって完全な冗長性を備えた設計になっている。これらの機能により、99.9999%の可用性を実現するようにシステムを設計できます。[2]PowerStore でクラスタリングと高可用性機能を使用することで、データ欠損のリスクを最小限に抑えることができます。

[2] 2020 年 4 月に策定されたデル・テクノロジーズの PowerStore 向け仕様に基づきます。実際のシステムの可用性は異なる場合があります。

参考資料

［デル・テクノロジーズ情報ハブ］ > ［ストレージ］サイトでは、お客様の Dell ストレージ プラットフォームでの成功を確実に実現するための専門知識を提供します。

Dell.com/powerstoredocs には、PowerStore システムをインストール、構成、管理する方法に関する詳細なドキュメントが用意されています。