



テクニカル レポート

## MetroCluster IP

### ソリューションのアーキテクチャと設計

NetApp

Stephen B. Galla

2023年5月 | TR-4689

## 概要

NetApp® MetroCluster™ は、FASシステムと AFFシステムで稼働するNetApp ONTAP®向けの継続的なストレージ ソリューションです。MetroCluster IPは、イーサネットベースのバックエンドストレージ ファブリックを使用した最新の進化型です。MetroCluster IPは、最も重要なビジネス アプリケーションのニーズを満たすために、高度な冗長の構成を提供します。MetroCluster IPはONTAPに含まれているため、個別のライセンスは必要ありません。また、ONTAPストレージを使用するクライアントとサーバにNASとSANの接続を提供します。

<<本レポートは機械翻訳による参考訳です。公式な内容はオリジナルである英語版をご確認ください。>>

## 目次

<b>MetroClusterの概要</b>	4
可用性の維持ソリューションの概要	4
MetroCluster IP と MetroCluster FC の比較	5
<b>MetroCluster IPアーキテクチャ</b>	5
ディザスタリカバリグループ	7
MetroCluster IPでのレプリケーション	7
ネットワーク	8
ストレージ	10
<b>解決策の設計</b>	11
サポートの確認	11
ハードウェアコンポーネント	11
解決策のサイジング	15
プラットフォームの制限	17
ネットワーク構成	17
クラスタファブリックスイッチ間リンク (ISL)	19
MetroCluster IP Inter-Switch リンクの設計	19
<b>運用と管理</b>	24
高可用性 (HA) とディザスタリカバリ (DR)	24
クオーラム監視	25
<b>相互運用性</b>	28
SnapMirror非同期	28
NetApp ONTAP FlexGroupボリューム	28
NetApp FlexCache	28
NetApp FabricPool	29
SVMディザスタリカバリ (SVM DR)	29
<b>詳細情報の入手方法</b>	30
<b>改訂履歴</b>	31
<b>表一覧</b>	
表1) MetroCluster IPコントローラモデル	12

表2) NetAppディスクシェルフの機能の比較 .....	12
表3) NetApp AFFコントローラとディスクシェルフの互換性.....	12
表4) NetApp FASコントローラとディスクシェルフの互換性 .....	13
表5) MetroCluster IPスイッチモデル .....	14
表6) NetApp AFFコントローラとスイッチの互換性.....	14
表7) NetApp FASコントローラとスイッチの互換性.....	15
表8) スイッチ間の距離 40GB 3~5m (ケーブルのおおよその長さ) .....	20
表9) スイッチ間の距離 100GB 3m~5m (ケーブルのおおよその長さ) .....	20
表10) 40GbEスイッチ用の短距離光モジュール .....	20
表11) 100GbEスイッチ用の短距離光モジュール .....	20
表12) 40Gbおよび100Gb光ケーブル .....	20

## 図一覧

図1) MetroCluster IPおよびVMware vSphere Metroストレージクラスタ .....	5
図2) MetroCluster IPアーキテクチャ .....	6
図3) ストレージとサーバ .....	6
図4) MetroCluster HAとディザスタリカバリ .....	7
図5) MetroCluster IP組み合わせファブリック .....	9
図6) AFF A700ネットワークの例 (1 サイト) .....	13
図7) ネットワークスイッチを使用しないAFF A700単一サイトの例 .....	14
図8) アクティブ/パッシブクラスタまたはサイト .....	16
図9) アクティブ/パッシブHA .....	16
図10) Hardware Universeプラットフォームの制限 .....	17
図11) RCFファイルジェネレータ .....	18
図12) サイトA (10Gb光モジュールとQuad Small Form-factor PluggableAdapter (QSA) を使用したパッシブDWDMの例) .....	22
図13) 10Gbポートアダプタを搭載したISL .....	23
図14) MetroClusterメディアエーターサイト .....	26
図15) MetroCluster Tiebreakerサイト .....	27
図16) Tiebreakerサイトリンク障害 .....	27
図17) Tiebreakerサイト障害 .....	28
図18) SVMディザスタリカバリ .....	29

## MetroClusterの概要

NetApp MetroClusterの構成は、世界中の数千社の企業で、高可用性 (HA) 、データ損失ゼロ、データセンタ内外のノンストップ オペレーションに使用されます。MetroClusterは、ONTAPソフトウェアのフリー機能で、別々の場所または障害ドメインにある2つのONTAPクラスタ間でデータと設定を同期的にミラーリングします。

MetroClusterは、次の2つの目的を自動的に処理することで、アプリケーションに継続的に使用できるストレージを提供：

- クラスタに書き込まれたデータを同期的にミラーリングすることで、RPO (Recovery Point Objective : 目標復旧時点) がゼロになります
- 2番目のサイトで構成をミラーリングし、データへのアクセスを自動化することで、ほぼゼロのRTO (Recovery Time Objective : 目標復旧時間) を実現します

MetroClusterは、2つのサイトにある2つの独立したクラスタ間でデータと構成を自動的にミラーリングすることで、シンプルな作業を実現します。1つのクラスタ内でストレージがプロビジョニングされると、2つ目のサイトの2つ目のクラスタに自動的にミラーリングされます。NetApp SyncMirror®は、すべてのデータの完全なコピーをゼロRPOで提供します。つまり、1つのサイトのワークロードをいつでも反対側のサイトに切り替えて、データを損失することなくデータの提供を継続できます。

MetroClusterは、2番目のサイトでNASおよびSANでプロビジョニングされたデータへのアクセスを提供するスイッチオーバープロセスを処理します。検証済みのソリューションとしてのMetroClusterの設計には、プロトコルのタイムアウト期間内またはそれよりも短い期間（通常は120秒未満）でスイッチオーバーを実行できるサイジングと構成が含まれています。これにより、ストレージのリカバリがストレージプロトコルのタイムアウト期間内に発生し、RPOがほぼゼロになります。アプリケーションは、障害を発生させることなくデータにアクセスし続けることができます。

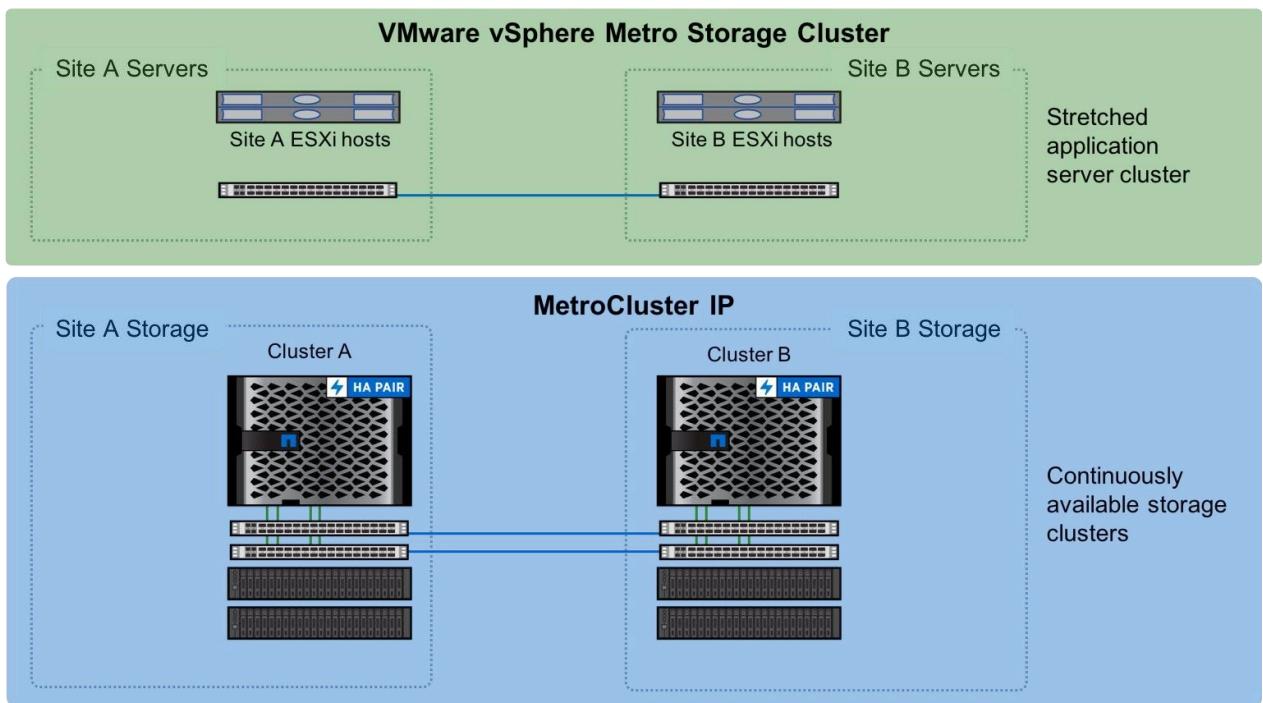
MetroClusterには、バックエンドストレージ ファブリックで定義されたいくつかのバリエーションがあります。MetroClusterには、2種類のストレージファブリック：FCとイーサネットがあります。イーサネットストレージ ファブリックは、MetroCluster IPと呼ばれます。

## 可用性の維持ソリューションの概要

MetroClusterは、ストレージの可用性の維持に対するニーズを満たします。同様のアプリケーション可用性製品と組み合わせることで、完全なソリューションが耐障害性が高いアーキテクチャを提供し、サイト全体で災害が発生した場合でも運用を継続できます。

例えば、MetroCluster IPとVMware vSphere Metro Storage Cluster (vMSC) を使用する場合です。この2つの製品を組み合わせることで、ビジネスクリティカルなアプリケーションのニーズに対応する、耐障害性が高い仮想インフラを構築できます。MetroCluster IPはストレージの可用性を提供し、vMSCはサイト全体が停止した場合でも操作可能なクロスサイト コンピューティング クラスタを提供します。

図1) MetroCluster IPおよびVMware vSphere Metroストレージクラスタ



データベースやMetroClusterと連携して動作するその他のアプリケーションにも、同様のマルチサイトアプリケーションソリューションの可用性があります。

## MetroCluster IPとMetroCluster FCの比較

次の機能は、MetroCluster IPとFCの違いをまとめています。

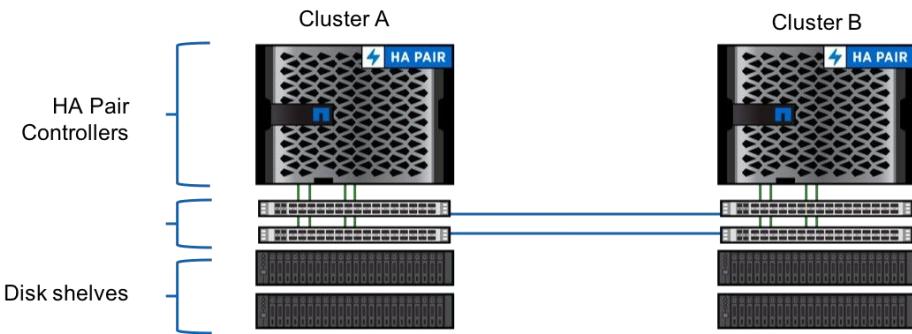
- MetroCluster IPは、FCバックエンドストレージファブリックではなくイーサネットバックエンドストレージファブリックを使用するため、専用のFCスイッチは不要です。
- MetroCluster IPは、ローカル レプリケーションとの両方でクラスタ間スイッチを縮小するため、FCスイッチは不要です。
- MetroCluster IPにはSASブリッジは不要です。
- MetroCluster IPは、リモートダイレクトアクセスメモリプロトコルであるiWARPを使用してNVRAMを複製します。
- MetroCluster IPは、iSCSIプロトコルを使用してリモートディスクにアクセスします。リモートディザスタリカバリノードはiSCSIのターゲットとして機能し、統合ストレージを備えたフラッシュシステムをサポートします。

MetroCluster FCは、MetroCluster Stretch FCと呼ばれる小規模な構成でも可用性があります。ONTAPを使用したMetroCluster Fcの詳細は、[TR-4375 : NetApp MetroCluster FC for ONTAP](#)をご参照ください。

## MetroCluster IPアーキテクチャ

MetroCluster IPはイーサネットストレージファブリックを使用します。MetroClusterストレージファブリックは、バックエンドストレージファブリックとも呼ばれ、ONTAPだけで使用されます。ONTAPクラスタインターフェース、MetroCluster SyncMirror、MetroCluster NVRAMミラー通信専用のネットワークです。

図2) MetroCluster IPアーキテクチャ



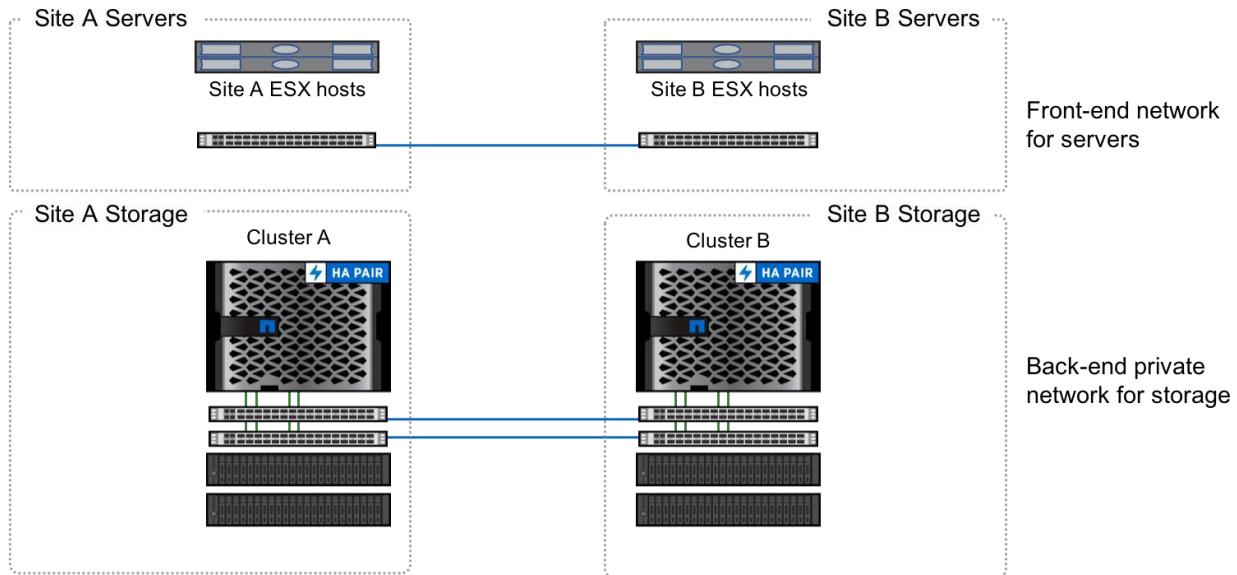
MetroCluster IPハードウェアの概要 :

- サイトごとに1台のHAペア コントローラ
- サイトごとに2台の高速イーサネット スイッチ :
  - ローカルおよびリモート レプリケーション用のコラピースト クラスタ内およびクラスタ間スイッチ
- ディスク シェルフまたは内蔵ストレージ

MetroClusterは、2つの独立したONTAPクラスタ間でデータをミラーリングすることで、ONTAPの可用性を拡張します。各クラスタはサイトまたは障害ドメイン内にあり、FASシステムまたはAFFシステムの標準HA機能を利用します。MetroClusterは、2つのONTAPクラスタ間でデータと構成の両方をミラーリングする機能を提供します。MetroClusterには、ストレージプロトコルの標準のタイムアウト期間内に1つのサイトから別のサイトへのフェイルオーバーを提供するように設計された、検証済みのシステム パラメータと制限が含まれています。

MetroClusterの機能とハードウェアは、一般的なONTAP FASシステムとAFFシステムの認定されたサブセットです。

図3) ストレージとサーバ



MetroClusterには、論理的に複数の機能エリアまたはコンポーネントに分割できるアーキテクチャがあります。レプリケーションなどのこれらのコンポーネントがどのように動作するかを理解することは、適切に設計されたソリューションを構築し、ソリューションを管理する上で重要です。

MetroClusterの主なコンポーネントは次の通りです。

- ディザスタリカバリグループ
- レプリケーション
- ネットワーク
- ストレージ

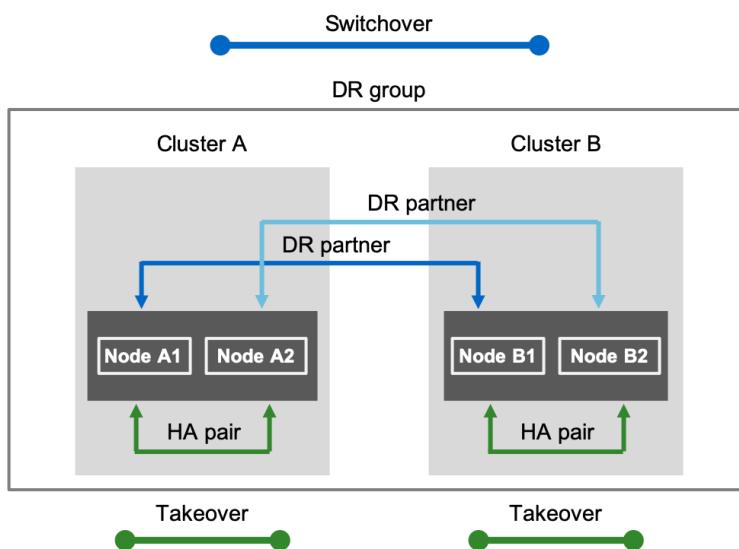
## ディザスタリカバリグループ

MetroCluster IPでは、グループとパートナーのディザスタリカバリというコンセプトを使用して、フェイルオーバーとスイッチオーバーの関係を判断します。2つのクラスタ（サイトAとサイトB）は、一緒にディザスタリカバリグループとして設定されます。グループ内では、ノードはディザスタリカバリパートナーとして関連付けられます。

HA関係は、標準クラスタと同じです。HAは、単一のコントローラ障害から保護し、ローカルでフェイルオーバーを実行します。また、HAは無停止のONTAPアップデートにも利用されます。サイト全体の障害の場合、ディザスタリカバリ関係を使用してサイトAからサイトBに切り替えます。これはスイッチオーバーと呼ばれます。

ディザスタリカバリパートナー関係は、MetroClusterの初期セットアップで設定され、変更されません。クラスタAから1つのノードを割り当て、クラスタBから1つのノードをパートナーとして割り当てるコマンドが1つあります。残りのノードは、ディザスタリカバリグループの設定を完了するために自動的に割り当てられます。

図4) MetroCluster HAとディザスタリカバリ



## MetroCluster IPでのレプリケーション

MetroCluster IPは、直接接続型ストレージを利用するため、外部SASブリッジを使用してディスクをストレージファブリックに接続する必要がありません。この構成では、ディザスタリカバリグループ内の各ノードがストレージプロキシまたはiSCSIターゲットとして機能し、グループ内の他のノードにディスクをエクスポートします。IPファブリックのストレージ転送プロトコルはiSCSI (SCSI over TCP/IP) です。これにより、iSCSIイニシエータとターゲットの間でTCP/IPファブリックを介した通信が可能になります。

リモートストレージにアクセスするために、ディザスタリカバリグループ内の各ノードはiSCSIイニシエータを利用して、パートナーノードのiSCSIターゲットとのiSCSIセッションを確立します。iSCSIおよび直接接続型ストレージを利用してことで、内蔵ディスクを搭載したシステムも使用できます。この構成により、各ノードは、外部ディスクシェルフに配置された内蔵ストレージデバイスとストレージデバイスの両方にディザスタリカバリパートナーノードからアクセスできるようになります。

MetroClusterには、次の3つのレプリケーションプレーンがあります。

- 設定レプリケーション- MetroCluster (MC) は、2つのONTAPクラスタで構成された構成です。各クラスタに専用のレプリケートされたデータベース (RDB) が格納され、それぞれ独自のメタデータが格納されます。スイッチオーバー中のクラスタ間でのメタデータオブジェクトの転送メカニズムには、クラスタピアリング、設定レプリケーションサービス (CRS) 、メタデータボリューム (MDV) の3つのコンポーネントがあります。MCを使用すると、ピアリングネットワークを使用してクラスタ間で設定オブジェクトをレプリケートできます。変更は、設定レプリケーションネットワーク経由でもう一方のクラスタのRDBにほぼ同期的に伝播されます。MDVは、クラスタピアリングネットワークが一時的に使用できない場合のフォールバックメカニズムとして使用されます。
- NVRAMレプリケーション- NVRAMレプリケーションでは、フェイルオーバーまたはスイッチオーバー時のデータ損失を防ぐために、ローカルノードのNVRAMをリモートディザスタリカバリノードのNVRAMにコピーします。NVRAMは、ローカルの高可用性パートナーとリモートのディザスタリカバリパートナーの両方にローカルでミラーリングされます。不揮発性キャッシュは、ローカル、HAパートナー、DRパートナー、およびDR補助パートナーの4つのパーティションに分割されます。4ノード構成では、各ノードのNVRAMが2回ミラーされます。この構成には、2つの4ノードDRグループを含む8ノード構成が含まれます。ミラーリングはHAパートナーとDRパートナーの両方に対して行われ、更新はiWARPプロトコルを使用してMetroCluster IP用のISL (スイッチ間リンク) 経由で転送されます。
- ストレージレプリケーション- MetroCluster IPでは、RAID SyncMirror (RSM) を使用してローカルおよびリモートのバックエンドディスクをミラーリングします。ディザスタリカバリグループ内の各ノードはリモートiSCSIターゲットとして機能し、バックエンドストレージを論理的に共有されたものとして認識します。ノードがリモートバックエンドディスクにアクセスするには、リモートディザスタリカバリパートナーノードを経由して、iSCSIターゲットを介して提供されるリモートディスクにアクセスします。ブロックは各サイトのペアノードにNVRAMとSyncMirrorの両方で書き込まれ、SyncMirrorの書き込みはRAIDレイヤで行われるため、重複排除や圧縮などのStorage Efficiency機能によって、SyncMirror処理によって書き込まれるデータを削減できます。

MetroCluster内のレプリケーションの詳細については、次のドキュメントを参照してください。

- [TR-4705 : NetApp MetroClusterソリューションアーキテクチャと設計](#)

## レプリケーションの暗号化

MetroClusterには、サイト間で送信されるデータを暗号化するメカニズムはありません。ネットワークレイヤやストレージレイヤでデータを暗号化するオプションがあります。

- ネットワーク層の暗号化は、Wavelength Division Multiplexing (WDM;波長分割多重) デバイスとスイッチベースの暗号化を使用して実現できます。ラウンドトリップレイテンシが要件の範囲内であれば、外部暗号化デバイスの使用がサポートされます。
- または、データのホスト側の暗号化を使用することもできます。欠点は、ONTAPが通常提供するストレージ効率をすべて無効にすることです。

注： NetApp Volume Encryption (NVE) を使用してボリュームに書き込まれたデータを暗号化できますが、ホストによって書き込まれた暗号化されていないブロックデータも含め、書き込みはNVRAMにレプリケートされます。

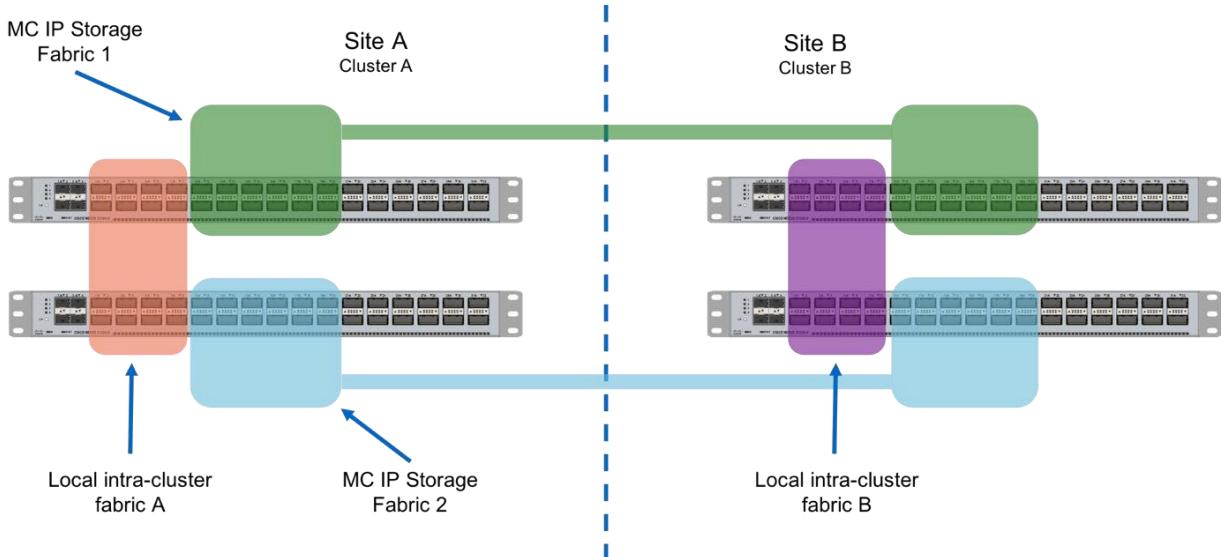
## ネットワーク

MetroClusterには、次の2つの独立したストレージファブリックがあります。

- クラスタ インターコネクト
- MetroCluster IPネットワーク

各ネットワークは特定の機能専用です。各ネットワークにマッピングされた特定の仮想LAN (VLAN) があり、個別のデータリンク レイヤ、またはOSI標準のレイヤ2が作成されます。

図5) MetroCluster IPの組み合わせファブリック。



NetAppでは、リファレンス構成ファイル (RCF) と呼ばれる標準的なスイッチ構成を提供しています。スイッチベンダー や プラットフォーム モデルに応じて、[RCFジェネレータツール](#)を使用して RCF ファイルを作成します。NetApp Support Site は、RCF ファイル ジェネレータのダウンロードをホストします。RCF ファイルを使用して、スイッチの設定を変更する必要があります。RCF ファイルは、スイッチごとに1つずつ、4つの個別のコンフィギュレーション ファイルのバンドルです。

RCF は、VLAN と チャネル グループ ID を指定します。これらはバックエンドストレージスイッチ内でのみ使用されます。IP アドレスには特定の要件があります。ワークシートおよび要件の説明については [MetroCluster IP インストールおよび設定ガイド](#) をご参照ください。

## クラスタ インターコネクト

ONTAP クラスタ インターコネクトは、NetAppストレージシステムのノード通信に不可欠な高速のローカル専用ネットワークです。すべてのクラスタ トラフィックはノードが配置されているサイトに対してローカルであるため、サイト間では接続されません。効率的で信頼性の高い通信を保証し、スケーラビリティをサポートします。

MetroCluster IP セットアップでは、専用のスイッチポートがインターフェイストラフィックを処理します。2つのISLポートでサイトのスイッチを接続し、冗長ローカルVLANを形成します。ブレークアウトケーブルは、ネイティブよりも低いポート速度で使用されます。

NetApp インターコネクトは、データネットワークや管理ネットワークとは別に、ストレージクラスタ内に専用のプライベートネットワークを形成します。スイッチ、ケーブル、NIC または HCA を使用して、冗長性、高速接続、および拡張性を提供します。これらのコンポーネントは、クラスタストレージ環境で高いパフォーマンスと信頼性を実現します。

注： AFF A150、AFF A250、AFF C250、FAS500f、および FAS2750 は、VLAN を使用して分離された MetroCluster IP ネットワークとクラスタポートを共有します。

注： ONTAP 9.7 以降では、ストレージコントローラ間でポートをクロス接続してスイッチレスインターフェイストラフィックを使用できます。これは、NetApp 準拠スイッチを使用する構成で使用されます。

## MetroCluster IP ネットワーク

MetroCluster IP構成では、各サイトが2つの独立したストレージファブリックで構成されます。これらのローカルファブリックは対応するリモートファブリックに接続されますが、クラスタ インターコネクトとは異なり、相互に接続されません。

各MetroCluster IPスイッチには、ノード接続用に指定された複数のポートがあります。サイトごとに2つのノードを含む標準の4ノードMetroCluster IP構成では、これらのポートのうち2つだけが使用されます。サイトの2つのノードはそれぞれ、別々のイーサネットインターフェイスを使用して両方のスイッチに接続します。通常、これらの接続には、標準のクラスタインターコネクトポート（オンボードポートまたは標準のONTAP HA構成で使用されるインターフェイス）が使用されます。

MetroCluster IPネットワークは、Internet Wide Area Remote Direct Memory Access（RDMA;インターネットワイドエリアリモートダイレクトメモリアクセス）プロトコル（iWARP）用に最適化された専用イーサネットアダプタを採用しています。このアダプタは、TCPオフロードエンジン（TOE）とiSCSIオフロード機能の両方を備えており、高速イーサネット経由でのRDMAを容易にします。各ノードには2つのiWARP / iSCSIアダプタポートが装備されており、1つのポートを各スイッチに接続します。スイッチは、図5に示すように、ローカルに接続されていない別々のファブリックを形成します。

NetApp AFF A150、AFF A250、AFF C250、FAS500f、FAS2750などのプラットフォームでは、ソフトウェアiWARPが使用されます。これらのプラットフォームには一定数のネットワークポートがあり、フロントエンドのホスト側データアクセス用にネットワークポートの使用を最適化するために、通常はクラスタインターフェイス用に予約されている2つのオンボードポート（e0aとe0b）を組み合わせて使用します。この機能を使用すると、クラスタ トライフィックとMetroCluster IPトライフィックで同じポートを共有でき、残りの4つのネットワークポートをホスト側のデータアクセスに使用できます。

注：iWARPは[IETF RFC 5040 - リモート ダイレクト メモリ アクセス プロトコル仕様書](#)で規定されている標準ベースのプロトコルです。

## ストレージ

MetroCluster IP用のストレージは、2つのサイト間で直接共有されることはありません。複数のサイトで一意のシェルフIDを設定する必要はありませんが、シェルフごとに一意のIDを割り当てる 것을 推奨します。これにより、ストレージシステム内のシェルフを識別して管理しやすくなります。各サイトのストレージには、ローカルHAペアだけが直接アクセスできます。リモートストレージは、「ストレージレプリケーション」セクションの説明に従って、iSCSIを使用してローカルノードの可用性を高めます。

## SyncMirror

SyncMirror、またはRAID SyncMirror（RSM）は、MetroClusterでサイト間のアグリゲートをミラーリングするために使用されるテクノロジです。各アグリゲートで2つのプレックスを設定できます。pool0とpool1と呼ばれます。Pool0にはノードのローカルストレージが含まれ、pool1にはリモートミラーコピーが含まれます。

## ADP

ONTAP 9.4以降では、MetroCluster IPでルート/データ/データ（RD2）パーティショニングを使用するAFFシステムでアドバンストディスクパーティショニング（ADPv2）がサポートされます。アドバンストドライブパーティショニング（ADP）は、ストレージ効率を高める機能です。これにより、HAペア内のアグリゲートとコントローラ間で物理ドライブの容量を共有できます。ADPは、ディスク全体のパーティショニングに比べて、使用可能容量と実効容量が増加し、ストレージ効率が最大10~40%向上します。RD2パーティショニングは、ドライブを1つのルートパーティションと2つのデータパーティションに分割し、1つのドライブの容量とIOPSをAFFシステムの両方のコントローラで使用できるようにします。ADPはMCの初期化時にデフォルトで適用されます。

このソリューションの詳細については、以下のドキュメントを参照してください。

- [TR-4705 : NetApp MetroClusterソリューションアーキテクチャと設計](#)

## ソリューションの設計

ソリューションの適切な設計は、パフォーマンス、容量、耐障害性の要件に対応するための鍵となります。ソリューションの設計の全体的な手順には、サポートされているホストとプラットフォームの構成を確認し、容量とパフォーマンスのニーズを満たすサイジングを行うことが含まれます。次の問題を考慮する必要があります：

- ホストとプロトコルのサポートの確保
- パフォーマンスのためのソリューションのサイジング
- 容量のソリューションのサイジング：容量のアクティブ-アクティブ、アクティブ-パッシブ構成
- システムの制限を確認する
- サイト間のISLのサイジング
- ケーブル要件

### サポートの確認

[NetApp Interoperability Matrix Tool \(IMT\)](#) を参照して、ホスト側のプロトコルとオペレーティングシステムのバージョンが、ONTAPの設計と同じ方法でサポートされていることを確認します。結果ページに記載されているアラートがMetroClusterに適用されているか確認します。

[Hardware Universe](#) には、システム仕様とサポートされる制限が記載されています。ONTAP 9.6以降では、Hardware UniverseにONTAP 9.6以降の相互運用性情報も含まれています。

### ハードウェア コンポーネント

次に、MetroCluster IPセットアップに含まれる主要なハードウェアコンポーネントの概要を示します。具体的なモデルとパート番号は、組織の正確な構成と要件によって異なる場合があります。これらの詳細については、[Interoperability Matrix Tool \(IMT\)](#) および [Hardware Universe](#)を参照してください。詳細については、[MetroCluster IP設定のインストール](#)に関するドキュメントを参照してください。

### プラットフォーム

次のプラットフォームは、MetroCluster IP構成でよく使用されるモデルを提供します。

- AFF Aシリーズ-フラッシュに特化して設計されたNetApp AFF Aシリーズは、業界をリードするパフォーマンス、密度、拡張性、セキュリティ、そしてネットワーク接続。このシステムはエンタープライズオールフラッシュアレイで業界最短のレイテンシを実現しており、最も要件の厳しいワークロードや、人工知能、ディープラーニングといったアプリケーションに最適な選択肢です。
- NetApp AFF Cシリーズシステムは、最新のQLCフラッシュテクノロジーで、より多くのデータをフラッシュに移動できるようにします。大容量を必要とする環境に最適で、手頃な価格でデータセンターを刷新してオールフラッシュに移行でき、さらにクラウドへの対応も実現します。
- AFF ASA - NetApp ASAシステムは、業界をリードするパフォーマンスと信頼性を提供するNetApp AFFシステム上に構築されています。AFFシステムは、複数のワークロードのストレージリソースを統合して共有したいお客様に、エンタープライズクラスのSAN解決策を提供します。
- FAS - NetApp FASストレージアレイは、フラッシュドライブとハードディスクドライブの混在を処理できるハイブリッドストレージシステムです。容量とパフォーマンスを最適なバランスで提供するFASシステムは、導入と運用が容易なだけでなく、将来の成長やクラウド統合にも対応できる柔軟性を備えています。

詳細については、次のデータシートを参照してください。

- [AFF Aシリーズ](#)
- [AFF Cシリーズ](#)
- [FASストレージアレイ](#)

## コントローラモデル

MetroCluster IPで利用可能なNetAppモデルは、さまざまなレベルのパフォーマンス、容量、拡張性を提供し、さまざまなビジネスニーズや予算要件に対応します。MetroCluster IP解決策を導入する場合は、プライマリサイトとセカンダリサイトの両方で、同じタイプのストレージシステム (AFFまたはFAS) と互換性のあるハードウェアおよびソフトウェアコンポーネントを使用する必要があります。次の表に、MetroCluster IPでサポートされるコントローラモデルを示します。

表1) MetroCluster IPコントローラモデル

プラットフォーム	エントリ	ミッドレンジ	ハイエンド
AFF Aシリーズ	A150	A250、A400	A700、A800、A900
AFF Cシリーズ		C250、C400	C800
FAS	FAS27xx	FAS500f, FAS8300, FAS8700	FAS9000、FAS9500

特定のモデルおよび機能は変更される場合があります。 詳細については、[AFFおよびFASのシステムドキュメント](#)を参照してください。

## ディスク シェルフ

ディスクシェルフは、ストレージシステムの容量を拡張するためのストレージデバイスとして機能します。 MetroCluster IPセットアップでは、物理的に離れた2つの場所間のデータレプリケーションを容易にすることで、データの高可用性を確保する上で重要な役割を果たします。両方の場所のノード（コントローラ）からディスクシェルフにアクセスできるため、2つのサイト間にミラーボリュームを作成できます。したがって、ディスクシェルフはMetroCluster IP構成の重要な要素であり、地理的に分散したサイト全体でデータとアプリケーションの可用性を維持するために必要なストレージ容量とデータレプリケーション機能を提供します。ディスクはさまざまなストレージ容量で利用でき、物理的な寸法に応じてラージフォームファクタ (LFF)（通常は3.5インチ）またはスマールフォームファクタ (SFF)（通常は2.5インチ）に分類されます。次の表に、使用可能なシェルフと、サポート対象のMetroCluster IPコントローラとの互換性を示します。

表2) NetAppディスクシェルフの機能の比較

項目	NS224	DS224C	DS212C	DS460C
フォーム ファクタ	2U	2U	2U	4U
エンクロージャあたりのドライブ数	24 SFF	24 SFF	12 LFF	60 LFF
ドライブ タイプ：	SSD	SSDs HDD	SSDs HDD	SSDs HDD
I/O モジュール	デュアルNSM	デュアルIOM12	デュアルIOM12	デュアルIOM12
接続	100Gb/秒イーサネット	12Gb/秒SAS	12Gb/秒SAS	12Gb/秒SAS

表3) NetApp AFFコントローラとディスクシェルフの互換性

	AFF A150	AFF A250	AFF A400	AFF A700	AFF A800	AFF A900	AFF C250	AFF C400	AFF C800
DS212C									
DS224C	X	X	X	X	X	X			
DS460C									
NS224	X	X	X	X	X	X	X	X	X

表4) NetApp FASコントローラとディスクシェルフの互換性

	FAS2750	FAS500f	FAS8300	FAS8700	FAS9000	FAS9500
DS212C	X		X	X	X	X
DS224C	X		X	X	X	X
DS460C	X		X	X	X	X
NS224		X				

[NS224シェルフ](#)の詳細については、製品ドキュメントを参照してください。DS212C、DS224C、およびDS460Cシェルフの詳細については、[SASシェルフセクション](#)を参照してください。

## スイッチ

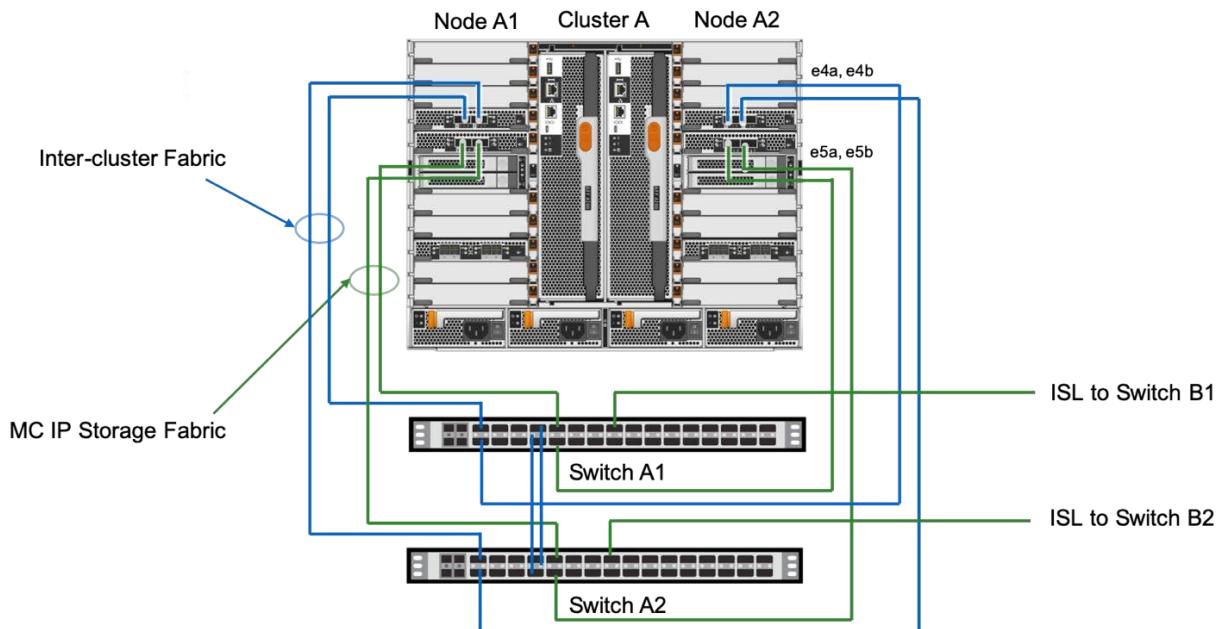
これらのレイヤ3イーサネットスイッチは、データセンター間の堅牢なIPベースの相互接続を作成するために不可欠であり、データの整合性と中断のない運用を確保します。10 / 25 / 40 / 50 / 100ギガビットイーサネット接続をサポートし、迅速なデータ転送と最小限のレイテンシを実現します。

これらのスイッチは、ジャンボフレーム、VLANタギング、リンクアグリゲーションなどの機能を提供し、MetroCluster環境のパフォーマンスと信頼性を強化します。ストレージコントローラ間のシームレスな通信を維持するために、スイッチはOSPFやBGPなどのルーティングプロトコルをサポートし、最適なパス選択とネットワーククロードバランシングを実現します。

MetroCluster IP構成ごとに4つのスイッチ（冗長性を確保するためにサイトごとに2つ）が必要で、異なるスイッチモデルを混在させることはできません。特定のプラットフォームおよびONTAPバージョンでサポートされるスイッチモデルについては、NetApp Support Siteの[Interoperability Matrix Tool](#)および[Hardware Universe](#)を参照してください。

ブレークアウトケーブルは、10Gb ISLリンクで光モジュールを接続する場合など、ネイティブポート速度よりも低いポート速度に使用されます。

図6) AFF A700ネットワークの例（1サイト）

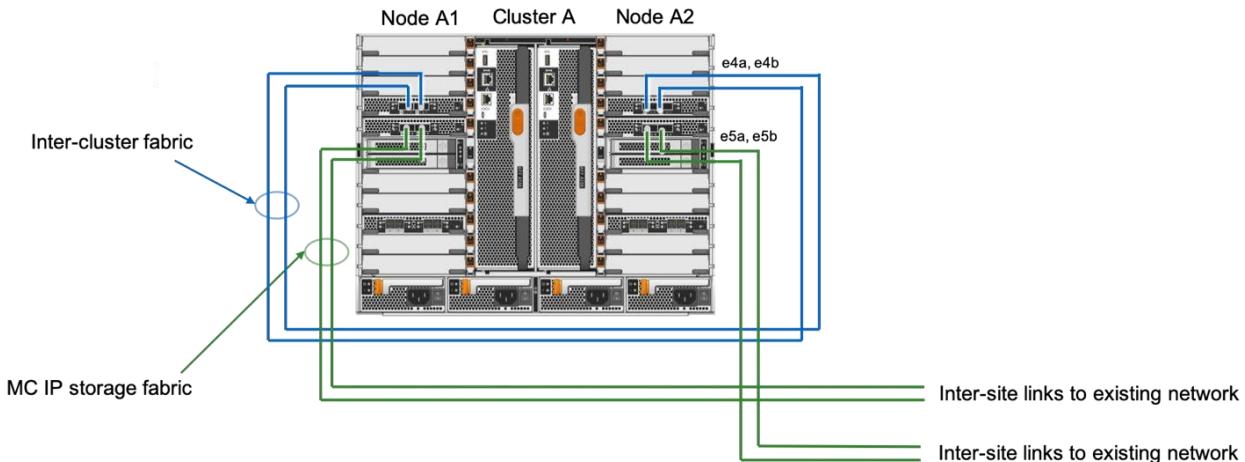


## NetApp準拠スイッチ-既存のスイッチを使用したMetroCluster IP

ONTAP 9.7では、特定のプラットフォームでMetroCluster IPがサポートされるようになりました。NetApp検証済みスイッチは必要ありません。これにより、MetroCluster IPバックエンドストレージアブリック以外の目的でも既存のスイッチを使用できます。解決策は、2ノードスイッチレスクラスタで構成されるスイッチレスクラスタ構成を各サイトで採用しています。このセットアップでは、クラスタインターコネクトインターフェイスをクロス接続し、MetroCluster IPインターフェイスを既存のMetroCluster準拠スイッチに接続します。

詳細については、[MetroCluster準拠スイッチを使用する場合の考慮事項](#)を確認してください。

図7) ネットワークスイッチを使用しないAFF A700単一サイトの例



次の表は、使用可能なスイッチモデルと、サポートされるMetroClusterコントローラモデルとの互換性を比較したものです。

表5) MetroCluster IPスイッチのモデル

モデル説明	NetApp PN	NetAppのポート速度
Broadcom BES-53248	X190005	10 / 25 / 100GB
Cisco Nexus 9336C	X190200-CS-PE	40GB / 100GB
Cisco Nexus 3232C	X190100	40GB / 100GB
NVIDIA SN2100	X190006	40GB / 100GB

表6) NetApp AFFコントローラとスイッチの互換性

	AFF A150	AFF A250	AFF A400	AFF A700	AFF A800	AFF A900	AFF C250	AFF C400	AFF C800
Broadcom BES-53248	X	X	X					X	
Cisco Nexus 9336C	X	X	X	X	X	X	X	X	X
Cisco Nexus 3232C	X	X	X	X	X	X	X	X	X
NVIDIA SN2100			X	X	X	X	X	X	X
NetAppへの準拠				X	X	X		X	X

表7) NetApp FASコントローラとスイッチの互換性

	FAS2750	FAS500f	FAS8300	FAS8700	FAS9000	FAS9500
Broadcom BES-53248	X	X	X	X		
Cisco Nexus 9336C	X	X	X	X	X	X
Cisco Nexus 3232C	X	X	X	X	X	X
NVIDIA SN2100			X	X	X	X
NetAppへの準拠			X	X	X	X

## ネットワーク アダプタ

MetroCluster IP構成では、専用のネットワークアダプタを使用して、IPネットワークを介した効率的なデータレプリケーションと通信を実現します。これらのアダプタはプラットフォームに依存しており、デュアルポートを備えたファストイーサネットを提供し、2つの異なるレイヤ2またはレイヤ3ネットワークへの接続を可能にします。さらに、アダプタはiWARPオフロード機能を提供し、イーサネットネットワーク経由でのリモートダイレクトメモリアクセス (RDMA) を可能にします。この機能により、サーバとストレージシステム間でメモリ間データを直接転送し、レイテンシとオーバーヘッドを最小限に抑えることができます。

これらのアダプタは、ストレージおよびNVRAMレプリケーションに使用されるノードからスイッチへの接続に使用されます。クラスタインターコネクトには、別のネットワークアダプタまたはネットワークポートが使用されます。

NetAppモデルによって、iWARPとネットワーク接続がさまざまな方法で実装されます。また、プラットフォームモデルでは、ネットワークアダプタの要件と実装が異なります。たとえば次のように指定できます。

- AFF A150、AFF A250、AFF C250、FAS500f、およびFAS2750はオンボードポートを使用します。
- AFF A400、AFF A700、FAS8300、FAS9000では、単一のネットワークアダプタを使用します。
- AFF A800は、オンボードとアダプタの2つのアダプタを使用します。
- AFF A900およびFAS9500は2つのネットワークアダプタを使用します。

AFF A150、AFF A250、AFF C250、FAS500f、およびFAS2750では、ソフトウェアiWARPをクラスタインターフェイスで組み合わせて使用するため、オンボードのe0aインターフェイスとe0bインターフェイスでトラフィックを共有できます。これにより、バックエンドストレージに必要なポート数が削減され、ホスト側のデータインターフェイスに使用できるポート数が最大になります。

## ソリューションのサイジング

ソリューションは、特定のストレージ容量やパフォーマンス要件に合わせてサイジングできます。

MetroClusterのサイジングは、容量に関するHAペアのサイジングと同様です。MetroClusterでは、ストレージデバイスはHAペアに使用される容量の2倍になり、反対側のサイトにデータのミラーコピーを提供します。

パフォーマンスのサイジングに関しては、ISLサイジング スプレッドシートを使用して考慮できる要素がISLです。

## アクティブ-パッシブ設定

NetApp環境のアクティブ/パッシブ構成では、2つのストレージノードまたはクラスタを使用して、高可用性、最小限のダウンタイム、シームレスなフェイルオーバー/フェイルバックの移行を実現します。アクティブノードはクライアントI/O要求とストレージリソースを管理し、パッシブノードはアクティブノードの健常性を監視して、障害やメンテナンスの際にテイクオーバーできるようにします。主なメリットには、高可用性、無停止のフェイルオーバーとフェイルバック、継続的監視などがあります。

アクティブ-パッシブ設定には、次の2種類があります。

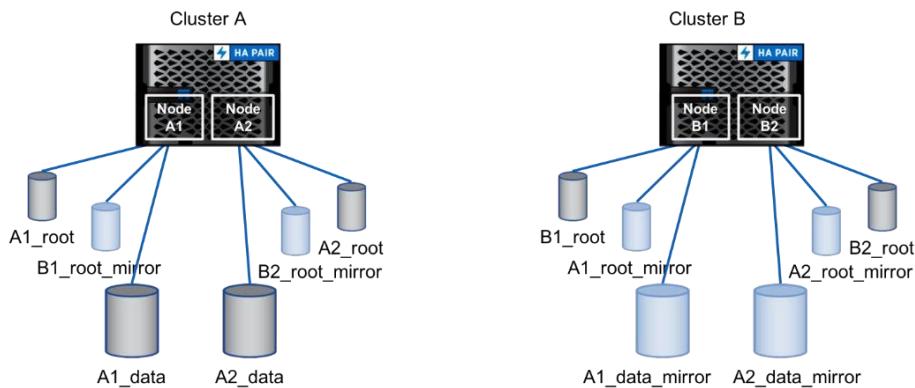
- アクティブ-パッシブ クラスタ
- クラスタ内内のアクティブ-パッシブHA

### アクティブ-パッシブクラスタ

アクティブ-パッシブ クラスタまたはサイトとは、1つのサイトが本番ワークロード、アクティブ サイト、および2番目のサイトのフェイルオーバーに使用される容量が最小の場です。これにより、アクティブ サイトにアクティブなストレージとワークロードが含まれる小規模なストレージ設定が可能になります。また、スイッチオーバーやサイト障害が発生した場合にのみ2番目のサイトを使用する場合は、ディスク容量を必要としないため、コスト効率も高くなります。

アクティブ-パッシブクラスタ設定の場合、1つのクラスタにはすべてのpool0ディスクがあり、もう1つのクラスタにはすべてのpool1ディスクがあります。アクティブ-パッシブ クラスタまたはサイトでは、MetroCluster用のMDVをホストするために、小さなデータアグリゲートを作成する必要があります。ルートボリュームとMDV用の小さなデータボリュームを除き、パッシブ サイトにはデータのミラー コピーのみが含まれます。

図8) アクティブ-パッシブ クラスタまたはサイト



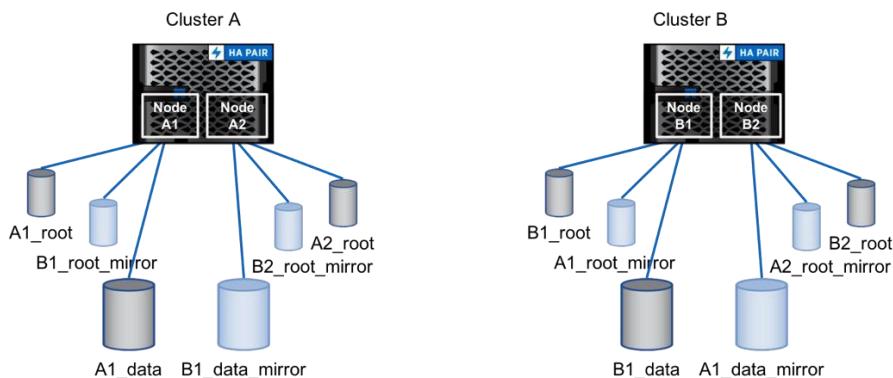
### アクティブ/パッシブHA構成

アクティブ-パッシブHA設定は、ストレージがHAペアのノードの1つに割り当てられる場合です。この設定は通常、小規模な設定で容量を最大化するために行われます。

AFFでは、ノード間でストレージを均等に分散させる必要があります。

この例では、各ノードがルートボリュームを持つディスクを所有しています。ローカルのアクティブなデータボリュームはホストされたノード1で、リモートミラーコピーはノード2でホストされます。

図9) アクティブ-パッシブHA

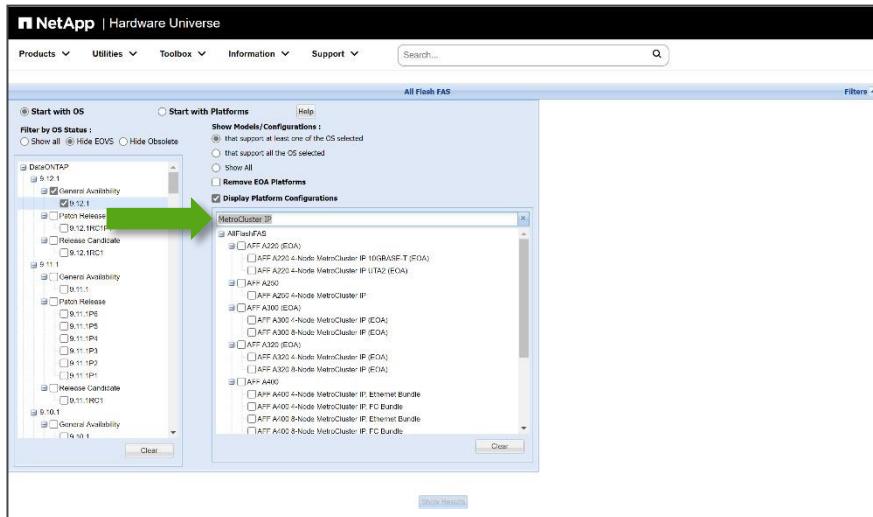


アクティブ/パッシブHAペアの詳細については、ONTAP 9ドキュメントセンターで[HAペアの管理の概要](#)を参照してください。

## プラットフォームの制限事項

[Hardware Universe](#)にはONTAPの制限があり、指定されたプラットフォームとONTAPのバージョンで確認できます。結果ページの下部に、プラットフォームの制限が表示されます。HAペアの制限を表示するには、[Platform Configurations]メニューから[MetroCluster IP]を選択します。

図10) Hardware Universeのプラットフォームの制限



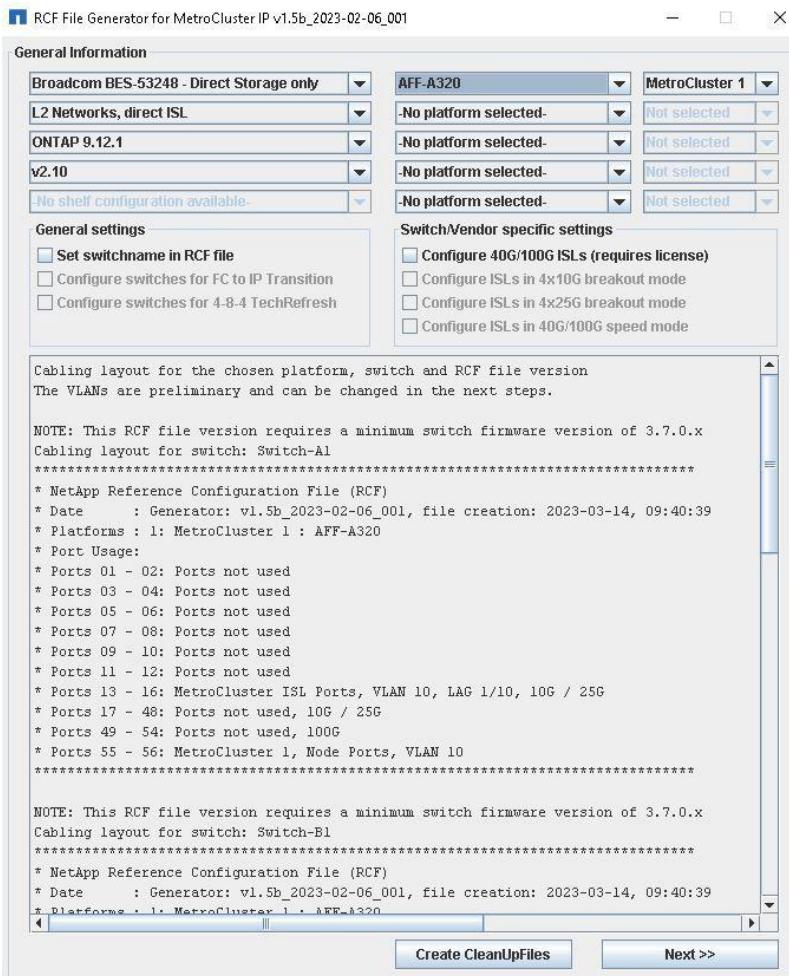
## ネットワークの設定

ネットワーク設定では、VLANとIPアドレスが他のネットワークと重複しないようにする必要があります。VLANはプライベートバックエンドアプリック用で、スイッチRCFファイルに自動的に割り当てられます。

ONTAP 9.6以前のリリースでは、RCFファイルはスイッチモデル、プラットフォームおよびONTAPリリースで作成されていました。ただし、ONTAP 9.6以降では、すべてのMetroCluster IPスイッチモデルとサポート対象プラットフォーム用のRCFファイルを作成できる新しいRCFファイルジェネレータユーティリティを使用できます。RCFファイルジェネレータユーティリティでは、共有レイヤ2のサイト間ネットワークをサポートするために、お客様が用意したVLAN IDを使用してRCFファイルを生成できます。また、RCFファイルジェネレータでは、レイヤ3ネットワーク構成をサポートしています。RCFファイルジェネレータからの出力に、選択したプラットフォーム、スイッチ、RCFファイルバージョンのケーブルレイアウトが表示されます。

**注：**ONTAP 9.7以前では、AFF A220コントローラおよびFAS 2750コントローラでVLAN IDの変更が有効になっていません。ただし、ONTAP 9.8では、AFF A220、AFF A250、FAS 2750、およびFAS500fにVLANを指定できます。デフォルトは10と20で、ユーザ指定のVLANは100より大きく4096未満である必要があります。

図11) RCFファイルジェネレータ



一般的なスイッチの構成では、ほとんどのポートは将来の拡張に備えてオープンです。スイッチは2つの冗長ファブリックを作成し、各ノードにはクラスタ接続とMetroCluster IPノード接続があります。スイッチへのノードの距離に応じて、ケーブル接続にはいくつかの選択肢があります。可能であれば、ノードとスイッチを同じ装置ラックに配置して、光ケーブルではなくTwinaxial銅線ケーブルを使用できるようにすることを推奨します。

スイッチは、モデルに応じて、10/25Gbps、40Gbpsまたは100GbpsのいずれかのNetAppのポート速度に対応しています。ネイティブポートはブレークアウトモードでも動作できます。ブレークアウトモードでは、ポートは4つのレーンに分割され、個々のインターフェイスとして使用されます。ブレークアウトモードで動作している場合、40GBスイッチのポートは、ケーブルまたは光モジュールに依存する4つの10Gbpsインターフェイスとして動作できます。すべての光モジュールがブレークアウトモードをサポートしているわけではありません。

100Gbスイッチは、ブレークアウトモードもサポートしています。ネイティブポート速度で動作する場合、1つの物理インターフェイスは4つの25GBPインターフェイスとして動作します。特定のケーブルと光モジュールは、ブレークアウトモードでの動作をサポートしています。RCFファイルは、各プラットフォームに必要な速度に合わせてブレークアウトポートを事前に設定します。

**注：** RCFファイルジェネレータはコマンドラインから起動できます。JAVA\_HOME環境変数を設定する必要がありますRCFファイルジェネレータを実行するには、次のコマンドを入力します。

```
java -jar RcfFileGenerator.jar
```

## クラスタファブリックのスイッチ間リンク (ISL)

クラスタインターコネクトファブリックは、MetroCluster IP構成のノード間の通信に使用される専用ネットワークです。このネットワークでは、信頼性と最適なパフォーマンスを確保するために、特定のスイッチ要件とケーブルタイプが必要です。

ISLネットワークに使用するスイッチにはネイティブポートが搭載され、高速なデータ転送をサポートするためにジャンボフレームが設定されている必要があります。ファブリックには、少なくとも2つのスイッチが必要です。

Twinaxial銅ケーブルまたは光ファイバーケーブルのいずれかを使用することをお勧めします。通常、2軸銅ケーブルは最大7メートルの距離に使用されますが、光ファイバーケーブルは長距離に使用されます。ケーブルは、干渉を減らし、信頼性の高いデータ転送を保証するために、高品質で適切にシールドされている必要があります。

## MetroCluster IP Inter-Switch リンクの設計

MetroCluster IP構成で高可用性と耐障害性を確保するには、2つのスイッチを高速かつ低レイテンシのリンクで接続するためにISL設計が必要です。ISLリンクの設計には、レプリケーショントラフィックをサポートするための十分な帯域幅、迅速なデータ転送のための低レイテンシ、高可用性のための冗長性、レプリケーショントラフィックを優先するための適切なQoS設定、不正アクセスを防止するためのセキュアな認証など、いくつかの重要な要素が含まれます。

MetroCluster IPのISLリンク設計プロセスは、サイト間の距離やネットワーク機能など、サイト固有の要因によって複雑になる可能性があります。ストレージパフォーマンス要件、直接ファイバの可用性、多重化デバイス、既存のファイバインフラストラクチャなど、さまざまな要素を考慮して、各リンクに必要なコンポーネントとサポートされる距離を決定する必要があります。必要な計算を実行するためにNetAppに相談することをお勧めします。

現在、ISLネットワークのMetroCluster IPのRound Trip Time (RTT ; ラウンドトリップ時間) 要件は次のとおりです。レイテンシは7ミリ秒以下、ジッタは3ミリ秒以下、パケット損失は0.01%未満です。これらの要件により最大10ミリ秒のレイテンシが実現されます。これは、データ転送の時間制限が厳しい同期レプリケーションに不可欠な要件です。

MetroCluster IP用のISL構成を設計する場合は、適切な光モジュールを選択し、サポートされる最大距離を提供する対応する光ケーブル構成を選択することが重要です。短距離の場合は、マルチモード光ファイバとケーブルが適しており、コスト効率が高くなります。長距離の場合は、長距離光ファイバとシングルモードファイバが必要です。最大距離については、通信プロバイダーに相談する必要があります。

場合によっては、単一のデータセンター内の距離を別々のアベイラビリティゾーンのラックで区切った距離に標準イーサネットケーブルを使用できます。スイッチとモジュールに応じて、ネイティブスイッチポートのISLの速度が25GB、40GB、または100GBであれば、標準イーサネットケーブルを使用できます。

要約すると、サイト固有の要因を考慮して、MetroCluster IP用のISLリンク設計を慎重に計画および実装することで、スイッチ間の高速で信頼性の高いセキュアなレプリケーショントラフィックを確保できます。

### ラック間：短距離

別々の場所が近接しているMetroCluster IP構成（ラック内や並列ラック内など）では、ISLリンクに銅線ケーブルを使用できます。多くの場合、これはラボまたはテスト構成の場合に該当します。とに、40GBスイッチと100GBスイッチの銅ケーブルをスイッチ接続するスイッチのポート番号を示します。

表8) スイッチ間の距離 40GB 3~5m (ケーブルのおおよその長さ)

距離	ケーブルPN	説明
1m未満	X66100-3	ケーブル、銅線、QSFP+-QSFP+、40GbE、1m
3m未満	X66100-3	ケーブル、銅線、QSFP+-QSFP+、40GbE、3m
5m未満	X66100-5	ケーブル、銅線、QSFP+-QSFP+、40GbE、5m

表9) スイッチ間の距離 100GB 3m~5m (ケーブルのおおよその長さ)

距離	ケーブルPN	説明
1m未満	X66211A-1	ケーブル、銅線、QSFP28-QSFP28、100GbE、1m
2m未満	X66211A-2	ケーブル、銅線、QSFP28-QSFP28、100GbE、2m
5m未満	X66211A-5	ケーブル、銅線、QSFP28-QSFP28、100GbE、5m

同様に、ラック間の光ケーブル配線も可能です。これにより、距離が光モジュールの仕様範囲内にある場合に、単純なISL構成が可能になります。

表10) 40GbEスイッチ用の短距離光モジュール

距離	モジュールPN	説明
OM4は400mまで	X65401	XCVR、QSFP+、光、40GbE、短波

表11) 100GbEスイッチ用の短距離光モジュール

距離	モジュールPN	説明
OM4は100mまで	X65405	XCVR、QSFP28、光、100GbE、短波

表12) 40GBおよび100GBの光ケーブル

長さ	モジュールPN	説明
2M	X66200-2	ケーブル、光、OM4、MPO/MPOタイプB
5M	X66200-5	ケーブル、光、OM4、MPO/MPOタイプB
15M	X66200-15	ケーブル、光、OM4、MPO/MPOタイプB
30M	X66200-30	ケーブル、光、OM4、MPO/MPOタイプB

## キャンパス リンク

短距離間で直接ファイバ接続を使用するキャンパス リンクは、ラック間ISLを使用する場合と同様です。考えられる違いの1つは、長距離光ファイバとシングルモードケーブルを使用して、マルチモードケーブルや短距離光ファイバと比べて長距離を実現することです。

現在、NetAppでは、40GbEまたは100GbEスイッチ用の長距離光モジュールを提供していません。長距離光ファイバを必要とするリンクの設計については、特定のスイッチ モデルの[Ciscoサポートマトリックス](#)および[Cisco光モジュールデータシート](#)を参照して、距離と接続の詳細をご確認ください。

## 専用ファイバ リンク

隣接する建物を接続するキャンパス ネットワークでは、専用ファイバ リンクがより一般的です。専用ファイバ リンクを使用すると、多数のファイバ接続からの信号をより少ないファイバ リンクに多重化できます。これにより、使用率を最大化し、サイト間のファイバの数を削減できます。光信号の多重化は波長分割多重化 (WDM) と呼ばれ、粗波長分割多重化装置 (CWDM) と高密度波長分割多重化装置 (DWDM) の2種類があります。

CWDMでは、DWDMよりも少ない数の波長を多重化できます。

**CWDM**は、一般的に、光モジュールからの光を光多重化および逆多重化し、单一のファイバペアで伝送できる单一の信号にする受動デバイスです。光モジュールは波長固有で、チャネルと呼ばれることがあります。2つの異なるファイバ信号を多重化するには、それぞれの送信元信号を異なる波長を使用する光モジュールから生成する必要があります。**CWDM**光モジュールはCiscoから入手でき、8種類の波長をサポートしています。これにより、8つのファイバ信号リンクを1つのファイバリンクに多重化できます。**CWDM**多重化装置はパッシブであり、信号を多重化および逆多重化する光ファイバのみを含みます。これは、一般敵に、**DWDM**と比較して、多重化デバイスと関連する光モジュールのコストを削減できます。

**DWDM**は、信号のマージに**CWDM**デバイスと同様の方法を使用します。主な違いは、光モジュールが生成する信号の精度が高く、狭い信号のスペクトル幅が狭くなり、信号間の間隔が狭くなることです。これにより、サイト間のファイバリンクでの伝送により多くの信号を組み合わせることができます。**DWDM**デバイスは、アクティブまたはパッシブでできます。受動的デバイスは、**CWDM**と同じアプローチを使用します。**CWDM**では、光モジュールが**DWDM**デバイスにマージされた特定の波長またはチャネルを送信して、单一の信号を生成します。この信号は、サイト間の長距離ファイバケーブルで伝送されます。

**DWDM**デバイスはアクティブデバイスとしても使用できます。この場合、スイッチと**DWDM**デバイス間の信号は標準光ファイバを使用し、**DWDM**デバイスを使用して、サイト間ファイバリンクにマージできる波長で信号を生成します。

距離に関して、光モジュールは、許容距離と仕様を満たすために必要なリンク特性を詳細に示します。長距離での伝送には、信号増幅器が必要になる場合があります。**DWDM**に適用される増幅器には、光増幅器など、いくつかの種類があります。NetAppは、最適な構成を設計するために、通信専門家とのご相談を推奨します。

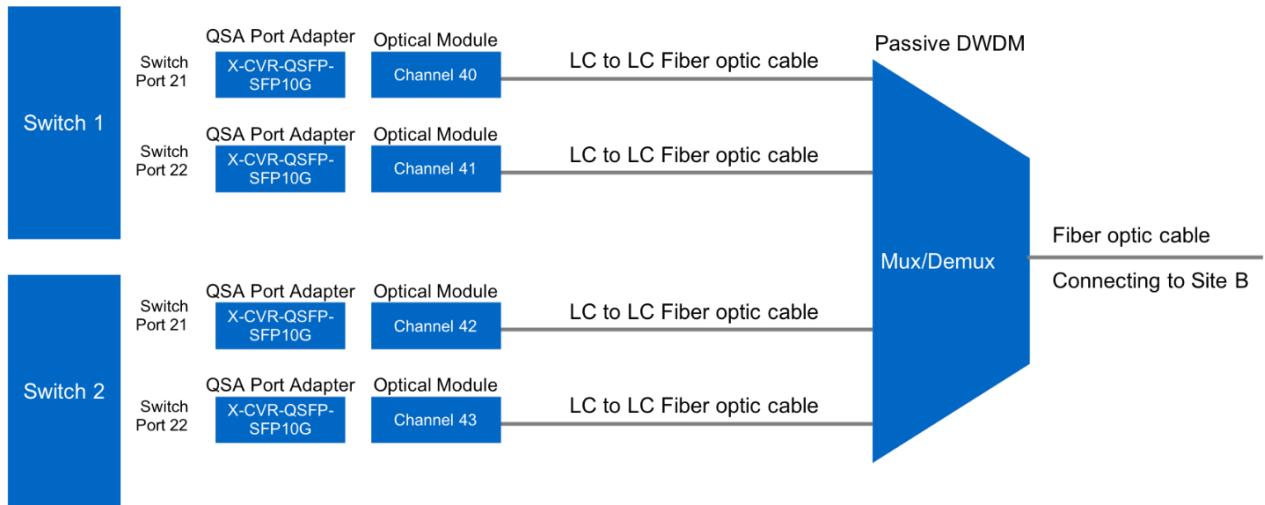
Ciscoでは、粗波長または高密度波長の多重化に使用できる10GB SFP+モジュールを提供しています。**xWDM**を使用すると、複数の光信号をサイト間の单一のファイバペアで結合または多重化し、スイッチまたはデバイスに信号をルーティングする前に逆多重化できます。

はパッシブ**DWDM**の使用例であり、特定のチャネル上の光モジュールのマッピングを示しています。各サイトのすべての光モジュールは、固有のチャネルを使用します。反対側のサイトB光と同じチャネルを一致させます。

#### **DWDM**モジュールの使用例

- チャネル40の2個の光モジュールを使用するサイトA スイッチ1 ポート21 - サイトB スイッチ1 ポート21
- チャネル41の2個の光モジュールを使用するサイトA スイッチ2 ポート21 - サイトB スイッチ2 ポート21
- チャネル42の1個の光モジュールを使用するサイトA スイッチ1 ポート22 - サイトB スイッチ2 ポート22
- チャネル43の2個の光モジュールを使用するサイトA スイッチ2 ポート22 - サイトB スイッチ2 ポート22

図12) サイトA (10Gb光モジュールとQuad Small Form-factor Pluggable Adapter (QSA) を使用したパッシブDWDMの例)



この例の最初のモジュールは、Cisco パーツ番号 DWDM-SFP 10G-45.32です。これは、ITUチャネル40の 1545.32nm波長 (100 GHz ITUグリッド) で動作する10GBASE-DWDM SFP +モジュールです。このサイトの設定を完了するには、チャネル41、42、43に対応する3個のモジュールを追加する必要があります。サイトBには、光モジュール、ポートアダプタ、およびパッシブDWDMとまったく同じ設定が含まれています。

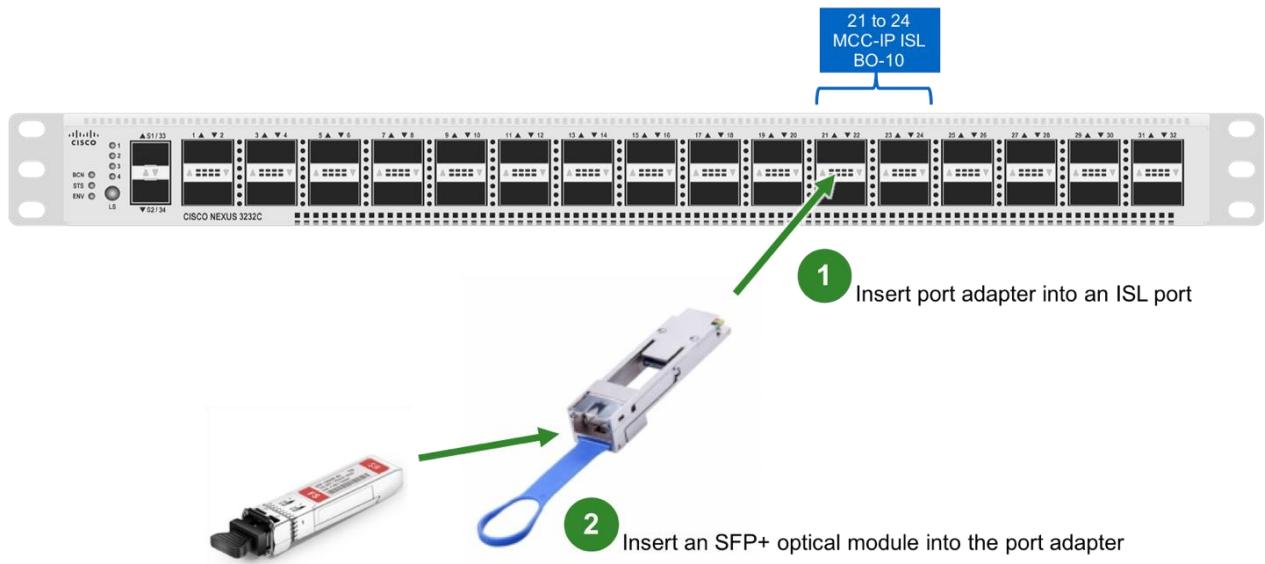
## 都市間リンク

長距離リンクでは、アクティブDWDM回線またはTelco回線がよく使用されます。スイッチからほとんどの通信機器またはアクティブDWDMデバイスへの接続は、データセンターまたはラック間構成で使用されるのと同じ光モジュールを使用して行われます。ラック間のケーブル配線とモジュールについては、「ラック間：短距離」の項を参照してください。

注：機器によっては、アクティブDWDMがISL トラフィックの暗号化を提供できる場合もあります。

NetAppは、都市間リンクの最適な構成を設計するために通信の専門家へのご相談を推奨します。

図13) 10GBポートアダプタを搭載したISL



ケーブル接続と光モジュールの詳細は[追加情報の入手先](#)のMetroCluster IPスイッチの技術資料をご参照ください。

Cisco DWDM光モジュールには、次の3つの構成があります。

- Cisco DWDM-SFP10G-XX.XXモジュール
  - DWDM固定モジュールは、40種類のノンチューナブルITU100 GHz波長をサポートします
- Cisco DWDM-SFP10G-Cモジュール
  - DWDMチューナブルモジュールは、最大80kmまで96のチューナブルITU 50GHz波長をサポートします。
- Cisco DWDM-SFP10G-C-Sモジュール
  - チューナブル トランシーバ モジュールはイーサネットのみです
  - DWDMチューナブルモジュールは、最大70 kmまで96のチューナブルITU 50GHz波長をサポートします。

サポートされている光モジュールおよび PN をCiscoパートナーから発注する方法の詳細は、「[詳細情報の入手先](#)」セクションをご参照ください。

## ISLネットワーク：専用リンク、共有レイヤ2、またはレイヤ3

MetroCluster IP構成では、サイト間のレプリケーションに専用または共有のISLを使用できます。専用ISL構成では、ISLスイッチに、MetroCluster IPスイッチのISLポートと同じ速度のネイティブポートが必要です。

MetroCluster IPスイッチとお客様のネットワークスイッチ間でISLの数と速度が一致している必要はありませんが、各MetroClusterスイッチを中間スイッチに接続するISLの数と速度が両方のMetroClusterサイトで同じである必要があります。

共有レイヤ2ネットワーク構成では、複数のMetroCluster IP構成で同じ中間ネットワークをISLで共有できます。ただし、これには、データレプリケーション用の低レイテンシと高スループットを確保するために、十分な容量と適切なISLサイジングが必要です。MetroClusterトラフィックが必要なサービスレベルを満たすように、MetroCluster IPスイッチ間のパスに沿ってクラスマップ、ポリシーマップ、およびQoSアクセスマップを設定する必要があります。

パケット損失は0.01%以下でなければなりません。サポートされるジッタ値はラウンドトリップで3ミリ秒です。共有ISLの理論上の最大スループットは、スイッチのモデルとポートタイプによって異なります。

MetroCluster IPスイッチを共有ネットワークに接続するISLの数は、スイッチのモデルとポートタイプによって異なります。

ONTAP 9.9.1以降では、MetroCluster IP設定をIPルーティング（レイヤ3）バックエンド接続で実装することができます。MetroClusterバックエンドスイッチは、ルータに直接接続するか、または他の介在するスイッチを介して、ルーティングIPネットワークに接続されます。4ノードMetroCluster構成のみがサポートされ、MetroClusterトライフィックではダイナミックルーティングはサポートされません。各MetroClusterサイトに2つのサブネット（各ネットワークに1つ）が必要です。

全体として、お客様は、MetroCluster IPを導入するための適切なネットワークを確保するために、特定の構成の要件と推奨事項を慎重に確認する必要があります。

詳細については、次のドキュメントを参照してください。

- ISLsに関する考慮事項  
[https://docs.netapp.com/us-en/ontap-metrocluster/install-ip/concept\\_considerations\\_isls.html](https://docs.netapp.com/us-en/ontap-metrocluster/install-ip/concept_considerations_isls.html)
- プライベートレイヤ2ネットワークの共有に関する考慮事項  
[https://docs.netapp.com/us-en/ontap-metrocluster/install-ip/concept\\_considerations\\_layer\\_2.html](https://docs.netapp.com/us-en/ontap-metrocluster/install-ip/concept_considerations_layer_2.html)
- レイヤ3ワイドエリアネットワークに関する考慮事項  
[https://docs.netapp.com/us-en/ontap-metrocluster/install-ip/concept\\_considerations\\_layer\\_3.html](https://docs.netapp.com/us-en/ontap-metrocluster/install-ip/concept_considerations_layer_3.html)

## 運用と管理

MetroClusterの運用と管理には、MetroClusterのヘルスと監視のチェックまたは検証が含まれます。ほとんどの運用では、MetroClusterを含むストレージを管理する手順は、ONTAPのマニュアルをご参照ください。

MetroCluster機能の管理および操作の詳細については [MetroCluster管理およびディザスタリカバリガイド](#) をご参照ください。

## 高可用性（HA）とディザスタリカバリ（DR）

MetroClusterには、主に2種類のリカバリメカニズムがあり、ローカルレベルやクラスタ（サイト）レベルの障害が発生した場合でも、継続的なデータ可用性と効率的なシステム機能を確保します。具体的には、テイクオーバーとギブバックと呼ばれるローカルレベルのリカバリと、スイッチオーバーとスイッチバックと呼ばれるクラスタレベルのリカバリがあります。

ONTAPのアップグレードなどのローカル障害やノンストップオペレーションは、HAパートナーによって処理されます。MetroClusterでは、標準のONTAP用語を使用してHAの運用を行います。

### ローカルレベルのリカバリ（HA）

テイクオーバーとギブバックのプロセスはMetroCluster IPのハイアビラビリティ（HA）ペアに関係するため、フォールトトレランスとノンストップオペレーションが確保されます。

- テイクオーバーとは、データ処理でエラーが発生したり停止したりしたときに、HAペアのノード（ローカルノード）がパートナー（パートナーノード）のストレージの制御を引き継ぐプロセスです。ソフトウェアやシステムの障害、停電、ハートビートメッセージの中止、手動による開始など、さまざまな状況でテイクオーバーがトリガーされる可能性があります。ローカルノードはパートナーの障害ステータスを検出し、データ処理をテイクオーバーすることでデータの可用性とシステム機能を維持します。
- ギブバックとは、問題からのリカバリまたはメンテナンスの完了後に、パートナーノードがストレージの制御を再開するプロセスのことです。ローカルノードはストレージをパートナーノードに戻し、通常の運用を再開します。テイクオーバーは、さまざまな状況で自動的に実行することも、storage failover takeoverコマンドを使用して手動で開始することもできます。

テイクオーバーとギブバックの詳細については、[HAペアの管理の概要](#)を参照してください。

## サイトレベルのリカバリ (DR)

スイッチオーバーとスイッチバックはMetroCluster IPのディザスタリカバリ (DR) プロセスに関連しており、あるクラスタサイトが別のクラスタサイトのタスクをテイクオーバーできるため、メンテナンスや災害からのリカバリが容易になります。ONTAP System Manager 9.6以降では、これらの処理がサポートされます。

- スイッチオーバーとは、クラスタサイト (サイトA) がストレージ制御とクライアントアクセスを別のクラスタサイト (サイトB) からテイクオーバーできるようにするプロセスです。これにより、テスト、メンテナンス、サイト障害時のノンストップオペレーションが実現します。計画的なネゴシエートスイッチオーバー (NSO) は、ディザスタリカバリのテストや計画的メンテナンスに使用できます。計画外スイッチオーバー (USO) は、どちらかのサイトに影響する災害が発生した場合に発生します。System Managerでスイッチオーバーが可能かどうかが判断され、ワーカロードが適宜調整されます。

スイッチオーバー後、MetroCluster IP構成の修復プロセスが2つのフェーズで実行されます。最初のフェーズでは、ミラーされたプレックスを再同期し、ルートアグリゲートをディザスタサイトにスイッチバックします。第2フェーズでは、スイッチバックプロセスに備えてサイトを準備します。

- スイッチバックは、サイトBでのメンテナンスや修理の完了後に、ストレージおよびクライアントアクセスの制御をサイトAからサイトBに戻す処理です。スイッチバックが成功するためには、サイトBのホームノードとストレージシェルフの電源をオンにした状態、修復フェーズが正常に完了した状態、サイトAのすべてのアグリゲートのミラーステータス、スイッチバック処理を実行する前に以前の設定変更がすべて完了した状態など、一定の条件を満たす必要があります。

ONTAP 9.5以降では、MetroCluster IP構成でのネゴシエートスイッチオーバー処理の実行中に自動的に修復が実行されます。ONTAP 9.6以降では、計画外スイッチオーバー後の自動修復がサポートされます。これにより、`metrocluster heal` コマンドを問題する必要がなくなります。

スイッチオーバーとスイッチバックの詳細については、[MetroClusterスイッチオーバーとスイッチバック](#)に関するドキュメントを参照してください。

## クオーラム監視

ストレージクラスタでは、高可用性を維持し、データ破損を防止し、フェイルオーバーとリカバリを容易にするために、クオーラム監視が不可欠です。クオーラム監視はTiebreakerとして機能し、大多数のノードが相互に通信できるようにすることで、クラスタノードが独立して動作してデータの不整合を引き起こすスプリットブレインシナリオを回避できます。MetroCluster IPは、クオーラム監視としてMetroCluster TiebreakerまたはONTAPメディエーターのいずれかをサポートします。

## ONTAPメディエーターソフトウェア

ONTAP 9.7には、新しいONTAPメディエーターソフトウェア解決策for MetroCluster IPがリリースされています。ソフトウェアは第3の障害ゾーンに配置され、MetroCluster IPでAutomated Unplanned Switchover (AUSO；自動計画外スイッチオーバー) を実行できます。さらに、2つのサイト間でデータのミラーリングに障害が発生した場合にAUSOを無効にする機能もあります。これにより、サイト間リンクが停止した場合の自動スイッチオーバーが回避され、管理者は手動でスイッチオーバーすることが適切かどうかを判断できます。

新しいONTAPメディエーターサービスは、一方のサイトのいずれかのMetroCluster IPノードから設定します。ONTAPは、すべてのノードと2つ目のクラスタに対して設定を自動的に実行します。初期リリースには、ONTAP 9.7以降を実行するMetroCluster IPおよびONTAPメディエーターソフトウェアリリースが必要です。Windows 1.0 以降

ONTAPでは、ONTAPメディエーターを設定するための新しいコマンドが追加されています。

`MetroCluster configuration - settings mediator add - mediator-address <mediator-ip>`

MetroClusterの設定-設定メディエーターの削除

新しいメディエーターを追加すると、有効なメディエーターのクレデンシャルを入力するように求められます。クレデンシャルはインストール時に設定されます。また、メディエーターシステムにログインし、次のコマンドを使用して変更することもできます。

- メディエーターのアカウント名を変更します。

```
/opt/netapp/lib/ontap mediator/tools/mediator_change_user
```

- メディエーターのパスワードを変更します。

```
/opt/netapp/lib/ontap mediator/tools/mediator_change_password
```

- メディエーターのステータスを確認します。

```
systemctl status ontap_mediator
```

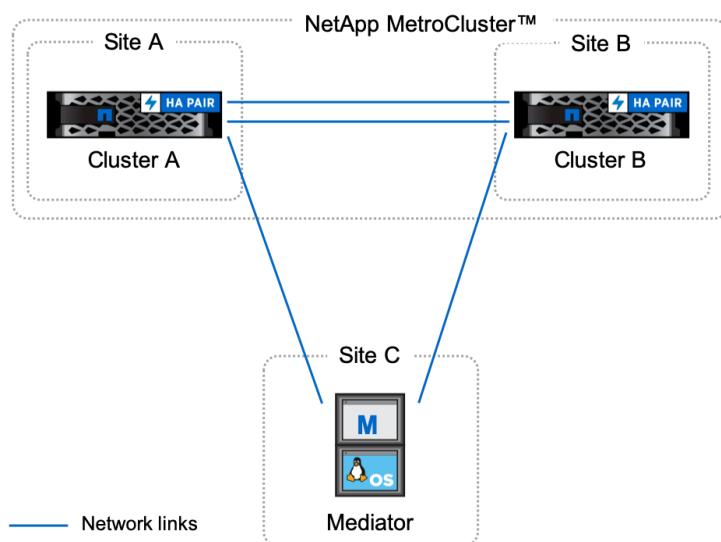
- ログ収集ツールを実行します。

```
/opt/netapp/lib/ontap メディエーター/tools/mediator_generate_support_bundle
```

ONTAP メディエーターの要件および障害の詳細については、[MetroCluster IP インストールおよび設定ガイド](#)を参照してください。

注： TiebreakerとONTAP メディエーターの両方を使用した同じ MetroCluster構成の管理はサポートされません。 MetroCluster構成の管理に使用できる製品は1つだけです。

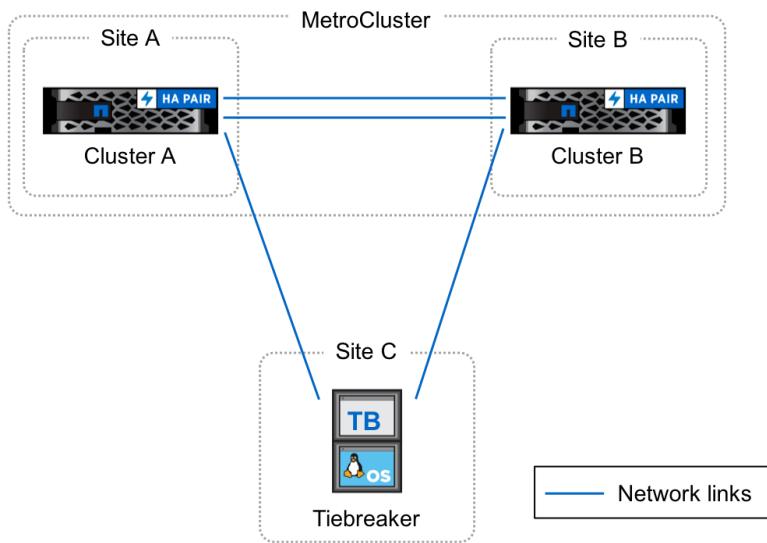
図14) MetroCluster メディエーター サイト



## Tiebreaker ソフトウェア

MetroCluster Tiebreaker Management Packは、MetroClusterシステムを監視し、サイトのディザスタやISL障害を検出する機能を提供します。Tiebreaker ソフトウェアは、通常は仮想マシン上のLinuxホストにインストールされます。この仮想マシンは、MetroClusterソリューションのいずれかのクラスタの障害ドメインとは別の第3の障害ドメインに配置されます。

図15) MetroCluster Tiebreakerサイト



Tiebreakerソフトウェアは、ノード管理LIFへの複数のパスとクラスタ管理LIFへの冗長接続を確立することによって、各コントローラを監視します。

注：TiebreakerとONTAPメディエーターの両方を使用した同じMetroCluster構成の管理はサポートされていません。MetroCluster構成の管理に使用できる製品は1つだけです。

### Tiebreakerサイト障害の症状

サイト障害時に、一方のクラスタがTiebreakerソフトウェアから到達不能で、もう一方のクラスタが到達可能である場合、Tiebreakerソフトウェアがアラートをトリガーする前に、到達可能なクラスタがパートナークラスタとの通信を失ったことも示す必要があります。2つのクラスタが引き続き通信できる場合、TiebreakerはTiebreakerソフトウェアと到達不能なクラスタとの間のネットワークでの接続の切断を識別します。

図16) Tiebreakerサイトリンク障害

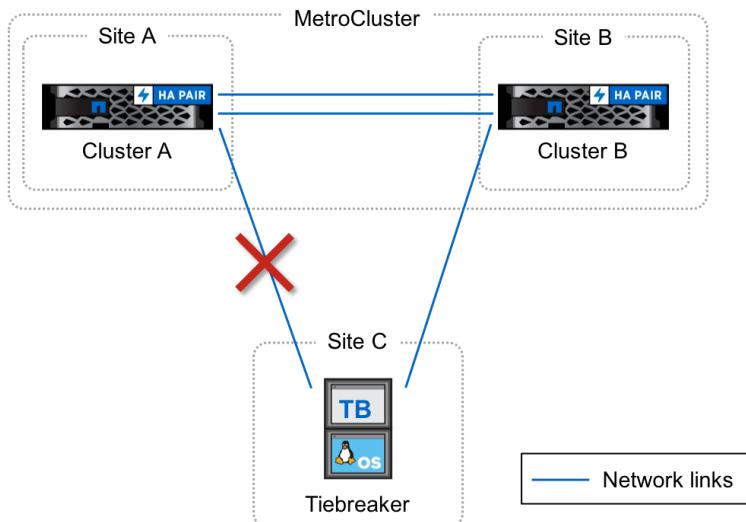
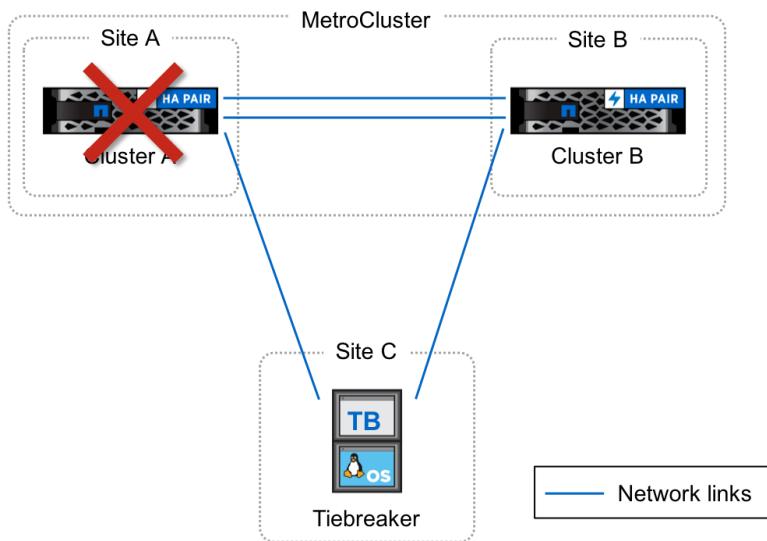


図17) Tiebreakerサイト障害



詳細は[Tiebreakerソフトウェアのインストールおよび設定ガイド](#)をご参照ください。

## 相互運用性

MetroClusterは、最も一般的なONTAP機能をサポートしています。ただし、SnapMirrorなどのONTAPの一部の機能は、現在MetroClusterではサポートされていません。MetroClusterのサポートに関するガイダンスは、ONTAPの機能のマニュアルをご参照ください。

### SnapMirror Asynchronous

SnapMirrorは、MetroClusterではサポートされていません。SnapMirror非同期保護は、ソースボリュームとデスティネーションボリュームの間で、スケジュールされたブロックレプリケーション、ミラー保護を3番目のクラスタに提供します。MetroClusterシステムは、SnapMirrorレプリケーション関係のソースおよびデスティネーションとして設定できます。詳細については、次のリンクを参照してください。

- [非同期SnapMirrorディザスター カバリの基本](#)
- [TR-4015 : SnapMirror Configuration and Best Practices Guide](#)

### NetApp ONTAP FlexGroupボリューム

NetApp FlexGroupボリュームは、ONTAP 9.6以降のMetroClusterでサポートされます。FlexGroupボリュームはスケールアウトNASコンテナであり、自動負荷分散と拡張性に加えて、高いパフォーマンスを提供します。詳細については、次のリンクを参照してください。

- [FlexGroupボリューム管理](#)
- [TR-4571 : 『NetApp ONTAP FlexGroup volumes-Best Practices and Implementation Guide』](#)

### NetApp FlexCache

NetApp FlexCache®テクノロジは、ONTAP 9.7以降のMetroCluster IPでサポートされています。FlexCacheは、ファイル配信を簡易化してWANのレイテンシを軽減し、WAN帯域幅のコストを削減できるリモート キャッシュング機能です。ファイル全体やボリューム全体ではなく、アクティブに読み取りが行われるデータのみをデータセンター内でローカルにキャッシュするか、リモートサイトに地理的に分散しているデータをキャッシュします。詳細については、次のリンクを参照してください。

- [FlexCacheボリューム管理](#)
- [TR-4743 : 『FlexCache in NetApp ONTAP』](#)

## NetApp FabricPool

NetApp FabricPoolは、ONTAP 9.7以降のMetroCluster IPでサポートされます。FabricPoolはONTAPに搭載されたハイブリッドストレージ解決策で、オールフラッシュ (SSD) アグリゲートをパフォーマンス階層として使用し、パブリッククラウドサービスのオブジェクトストアをクラウド階層として使用します。詳細については、次のリンクを参照してください。

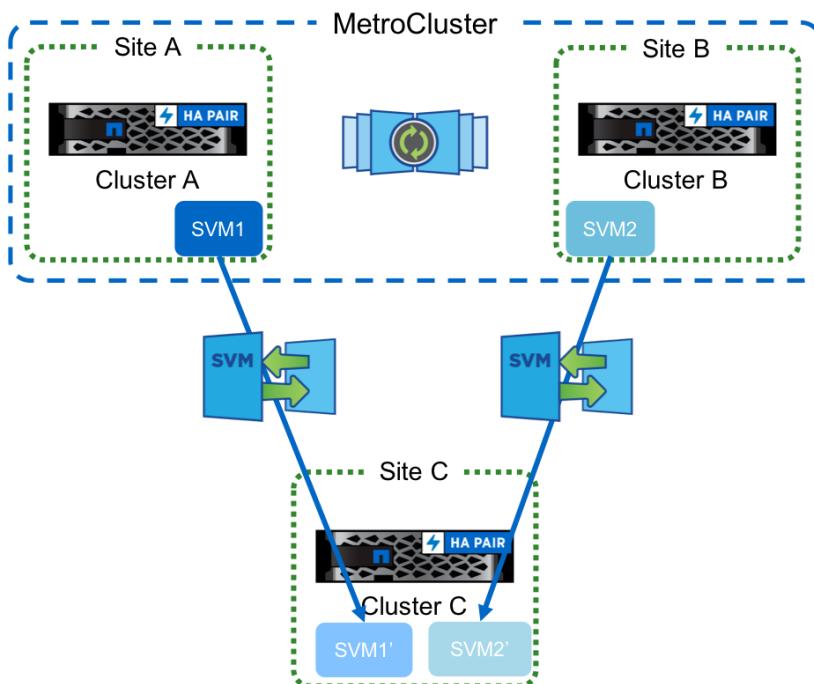
- [FabricPoolの製品 ドキュメント](#)
- [TR-4598 : 『FabricPool Best Practices』](#)

## SVMディザスタ リカバリ (SVM DR)

ONTAP 9.5以降では、SVM DRがMetroCluster IPと互換性を持つようになりました。SVM DRは、SVMにデータ保護機能とディザスタリカバリ機能を提供するように設計された解決策です。管理者はこの機能を使用してプライマリSVMのレプリカを作成し、リモートサイトでデータと設定の最新のコピーを使用できるようにすることで、災害を軽減できます。これは、SnapMirror非同期テクノロジを利用して、プライマリサイトのソースSVMとDRサイトのデスティネーションSVMの間の関係を確立することによって実現されます。災害が発生した場合は、セカンダリSVMをアクティブ化してデータへのアクセスを提供することで、ダウンタイムを最小限に抑えることができます。

注：ONTAP 9.5以降では、SVMを保護できるのはMetroCluster構成の1つのサイトだけになりました。  
ONTAP 9.11.1では、MetroClusterの両方のサイトからSVMを保護する機能が導入されました。

図18) SVMディザスタ リカバリ



## 詳細情報の入手方法

このドキュメントに記載されている情報の詳細は、以下のドキュメントやウェブサイトをご参照ください。

- MetroCluster IP Installation and Configuration Guide  
<https://docs.netapp.com/us-en/ontap-metrocluster/install-ip/index.html>
- MetroClusterディザスタリカバリ ガイド  
[https://docs.netapp.com/us-en/ontap-metrocluster/disaster-recovery/concept\\_dr\\_workflow.html](https://docs.netapp.com/us-en/ontap-metrocluster/disaster-recovery/concept_dr_workflow.html)
- ONTAPの製品ドキュメント  
<https://docs.netapp.com/us-en/ontap-family/>
- TR-4705 : NetApp MetroClusterソリューションのアーキテクチャと設計  
<https://www.netapp.com/pdf.html?item=/media/13480-tr4705pdf.pdf>
- TR-4375 : NetApp MetroCluster FC  
<https://www.netapp.com/pdf.html?item=/media/13482-tr4375pdf.pdf>
- MetroCluster IP Technical FAQ (NetApp Field Portal、ログインが必要)  
<https://fieldportal.netapp.com/content/748972>
- NetApp Interoperability Matrix Tool  
<https://imt.netapp.com/matrix/#welcome>
- NetApp MetroClusterに関するドキュメントリソース  
<https://www.netapp.com/support-and-training/documentation/metrocluster/>
- TR-4592 : 『Oracle on MetroCluster』  
<https://www.netapp.com/pdf.html?item=/media/8583-tr4592pdf.pdf>
- NetApp MetroClusterでのVMware vSphere 5.x、6.x、7.xのサポート (2031038)  
<https://kb.vmware.com/s/article/2031038>
- TR-4128 : NetApp MetroCluster 8.3のvSphere 6  
<https://fieldportal.netapp.com/content/252106> (ログインが必要です)

ケーブル接続と光モジュールの詳細は、以下のMetroCluster IPスイッチの技術資料をご参照ください。

- MetroCluster IP 10 / 25GbE Broadcom BES-53248スイッチ (NetApp Field Portal、ログインが必要)  
<https://fieldportal.netapp.com/content/886413>
- MetroCluster IP 40GbE Cisco N3132Q-Vスイッチ (NetApp Field Portal、ログインが必要)  
<https://fieldportal.netapp.com/content/729700>
- MetroCluster IP 100GbE Cisco N3232CおよびN9336Cスイッチ (NetApp Field Portal、ログインが必要)  
<https://fieldportal.netapp.com/content/757495>

サポートされている光モジュールおよびパート番号の詳細は、特定のスイッチ モデルのCisco光モジュール サポートマトリックスをご参照ください。

- Ciscoモジュールおよびスイッチ サポートマトリックス  
<https://tmgmatrix.cisco.com/home>
- Cisco CWDM SFP 10 ギガビット イーサネット ソリューションデータシート  
<https://www.cisco.com/c/en/us/products/collateral/interfaces-modules/transceiver-modules/datasheet-c78-734047.html>
- Cisco 10GBASE SFP+ モジュールデータシート  
[https://www.cisco.com/c/en/us/products/collateral/interfaces-modules/transceiver-modules/data\\_sheet\\_c78-455693.html](https://www.cisco.com/c/en/us/products/collateral/interfaces-modules/transceiver-modules/data_sheet_c78-455693.html)
- Cisco 10GBASE 高密度波長分割多重化 SFP+ モジュールデータシート  
[https://www.cisco.com/c/en/us/products/collateral/interfaces-modules/dwdm-transceiver-modules/data\\_sheet\\_c78-711186.html](https://www.cisco.com/c/en/us/products/collateral/interfaces-modules/dwdm-transceiver-modules/data_sheet_c78-711186.html)

## バージョン履歴

バージョン	日付	ドキュメントの改訂履歴
バージョン1.1	2019年2月	ONTAP 9.5のアップデートが含まれています
バージョン1.2	2019年5月	ONTAP 9.6に関する更新
バージョン1.3	2019年11月	ONTAP 9.6に関する更新
バージョン1.4	2020年11月	ONTAP 9.8のサポート
バージョン1.5	2023年5月	ONTAP 9.12.1の更新

本ドキュメントに記載されている製品や機能のバージョンがお客様の環境でサポートされるかどうかについては、NetApp サポートサイトで [Interoperability Matrix Tool \(IMT\)](#) を参照してください。NetApp IMT には、NetApp がサポートする構成を構築するために使用できる製品コンポーネントやバージョンが定義されています。サポートの可否は、お客様の実際のインストール環境が公表されている仕様に従っているかどうかによって異なります。

### 機械翻訳に関する免責事項

原文は英語で作成されました。英語と日本語訳の間に不一致がある場合には、英語の内容が優先されます。公式な情報については、本資料の英語版を参照してください。翻訳によって生じた矛盾や不一致は、法令の順守や施行に対していかなる拘束力も法的な効力も持ちません。

### 著作権に関する情報

Copyright © 2023 NetApp, Inc. All Rights Reserved. Printed in the U.S.このドキュメントは著作権によって保護されています。著作権所有者の書面による事前承諾がある場合を除き、画像媒体、電子媒体、および写真複写、記録媒体、テープ媒体、電子検索システムへの組み込みを含む機械媒体など、いかなる形式および方法による複製も禁止します。

NetApp の著作物から派生したソフトウェアは、次に示す使用許諾条項および免責条項の対象となります。

このソフトウェアは、NetApp によって「現状のまま」提供されています。NetApp は明示的な保証、または商品性および特定目的に対する適合性の暗示的保証を含み、かつこれに限定されないいかなる暗示的な保証も行いません。NetApp は、代替品または代替サービスの調達、使用不能、データ損失、利益損失、業務中断を含み、かつこれに限定されない、このソフトウェアの使用により生じたすべての直接的損害、間接的損害、偶発的損害、特別損害、懲罰的損害、必然的損害の発生に対して、損失の発生の可能性が通知されていたとしても、その発生理由、根拠とする責任論、契約の有無、厳格責任、不法行為（過失またはそうでない場合を含む）にかかわらず、一切の責任を負いません。

NetApp は、ここに記載されているすべての製品に対する変更を隨時、予告なく行う権利を保有します。NetApp による明示的な書面による合意がある場合を除き、ここに記載されている製品の使用により生じる責任および義務に対して、NetApp は責任を負いません。この製品の使用または購入は、NetApp の特許権、商標権、または他の知的所有権に基づくライセンスの供与とはみなされません。

このマニュアルに記載されている製品は、1つ以上の米国特許、その他の国の特許、および出願中の特許により保護されている場合があります。

本書に含まれるデータは市販の製品および/またはサービス（FAR 2.101 の定義に基づく）に関係し、データの所有権は NetApp, Inc. にあります。米国政府は本データに対し、非独占的かつ移転およびサブライセンス不可で、全世界を対象とする取り消し不能の制限付き使用権を有し、本データの提供の根拠となった米国政府契約に関連し、当該契約の裏付けとする場合にのみ本データを使用できます。前述の場合を除き、NetApp, Inc. の書面による許可を事前に得ることなく、本データを使用、開示、転載、改変するほか、上演または展示することはできません。国防総省にかかる米国政府のデータ使用権については、DFARS 252.227-7015(b)項で定められた権利のみが認められます。

### 商標に関する情報

NetApp、NetAppのロゴ、<https://www.netapp.com/company/legal/trademarks/>に記載されているマークは、NetApp, Inc.の商標です。その他の会社名と製品名は、それを所有する各社の商標である場合があります。

TR-4689-0523-JP