Oracleホワイト・ペーパー
2012年4月

Sun ZFS Storage ApplianceとOracle Exadata Database Machineを使用したバックアップおよびリカバリのパフォーマンスとベスト・プラクティス

ORACLE

概要

このホワイト・ペーパーでは、Oracle Exadata Database Machine向けのバックアップおよびリカバリ・ターゲットとしてSun ZFS Storage Applianceを構成および使用し、最適なパフォーマンスと可用性を達成する方法についてのMaximum Availability Architecture（MAA）ベスト・プラクティスを提供します。これらのMAA検証済み手順に従うことで、次の成果がもたらされます。

- Sun ZFS Storage Applianceを使用した、可用性と信頼性に優れたExadata Database Machineのバックアップ/リストア・アーキテクチャによって、ディスク（Exadata Database Machine）からディスク（Sun ZFS Storage Appliance）へのバックアップ速度として最大9TB/時が達成されます。

- Sun ZFS Storage Applianceからのリストア速度として最大7TB/時が達成されます。

上記のリストア速度が、アプリケーションのリカバリ・ポイント目標（RPO）とリカバリ時間目標（RTO）の要件を満たしており、上記のバックアップ速度でデータベース・バックアップが、要求されたバックアップ時間枠を満たす場合、Exadata Database MachineおよびSun ZFS Storage Appliance向けのMAAベスト・プラクティスがもっとも有効です。より高速なバックアップ速度またはリストア速度を必要とする場合、Exadata Storage Expansion Rackの使用、またはターゲット・ディスク・ストレージとして使用するExadata Database Machineの内部Exadataストレージの追加、もしくはSun ZFS Storage Appliance構成の拡張を検討する必要があります。Exadataストレージへのバックアップについて、詳しくはOracle MAAホワイト・ペーパー『Backup and Recovery Performance and Best Practices for Exadata Database Machine - Oracle Database 11.2.0.2』を参照してください。ゼロに近いRTOおよびRPOが要求されている場合は、Oracle Data Guardおよび統合クライアント・フェイルオーバーを使用してください。詳しくは、『Oracle Database 11g Release 2 High Availability Best Practices』と、Oracle MAAホワイト・ペーパー『Oracle Data Guard: Disaster Recovery Best Practices for Exadata Database Machine』および『Client Failover Best Practices for Highly Available Oracle Databases: Oracle Database 11g Release 2』を参照してください。
Exadata Database MachineとSun ZFS Storage Applianceを使用したこのバックアップ・ソリューションは、高可用性とパフォーマンスのバランスが取れた構成であり、MAAおよびSun ZFS Storage Applianceの開発チームによる検証を受けています。リカバリ時間は次の要素に応じて異なります。

- データベース・サイズをリストア速度で割った数値

- 増分バックアップ・サイズをリストア/マージ速度で割った数値

- REDO（アーカイブおよびオンラインREDO）の適用

おもな利点は次のとおりです。

- パフォーマンスに優れた高可用性のバックアップおよびリストア・ソリューション
- 高い費用効果
- 構成と運用に関する当て推量を排除することで、上記のバックアップ速度とリストア速度を実現
- Exadata Database Machineからバックアップ・データを取り除くことで、データベースの成長やExadataシステムへの追加データベースの統合に対してより多くのデータ領域を確保
- ほとんどのRPO要件およびRTO要件を満たす、迅速なバックアップ時間とリストア時間

Sun ZFS Storage Applianceでは、バックアップ・セットとイメージ・コピーの読取り専用スナップショットと読取り/書込みクローンを共有できます。スナップショットは、Oracle Recovery Manager（Oracle RMAN）カタログを使用して追跡するか、またはクローン・データベースを使用してアクセスできます。クローンは読取り/書込みモードでオープンできるため、開発システムやレポーティング・システムなどの別の処理要件に対応できます[1]。Oracle Database 11.2.0.3と同時に使用する場合、Sun ZFS Storage ApplianceはHybrid Columnar Compression（HCC）データを完全サポートするため、二次的な処理要件に対してさらなる機能性と柔軟性を提供できます。Sun ZFS Storage Applianceのリモート・レプリケーション機能を使用すると、リモートのSun ZFS Storage Applianceシステムに対してバックアップ・コピーや非構造化データをレプリケートできます。これらのコピーは、テスト・データベースのクローン作成といったリモート・サイトでの追加処理や障害時リカバリを目的として使用できます。また、Sun ZFS Storage ApplianceはDisk to Disk to Tape（D2D2T）ソリューションの一部として使用するでき、圧縮および暗号化を任意で有効化できます。ただし、このホワイト・ペーパーでは、高速かつ信頼できるDisk to Disk（D2D）のバックアップおよびリストア・ソリューションを実現することのみに焦点を合わせて説明します。次に示す構成と運用の手法によって、Sun ZFS Storage Applianceで使用可能な帯域幅、ネットワーク帯域幅、かなりの割合のExadata I/O帯域幅が最大化されます。Oracle RMANの"minimize load"オプションやリソース管理を

[1] オラクルのSun ZFS Storageを使用したデータベース・クローンについて、詳しくはMAAホワイト・ペーパー（1）『Sun ZFS Storage ApplianceとOracle Data Guardを使用したデータベース・クローニング』および（2）『Database Cloning using Oracle Sun ZFS Storage Appliance and Oracle Recovery Manager』を参照してください。

使用することで、バックアップおよびリストア操作中に重要なアプリケーション・パフォーマンスへの影響を軽減できます（詳しくは後述します）。.

はじめに

Exadata Database MachineにおいてOracle Maximum Availability Architectureのベスト・プラクティスを使用すると、Oracle Database向けとしてはもっとも包括的かつ効果的なHA（高可用性）ソリューションが実現されます。

Exadata Database MachineとSun ZFS Storage Applianceを配置する際に重要となる運用側面は、データベース・バックアップが正しく実行されており、障害発生時にOracle Databaseを素早くリストアできることを確認しておくことです。このホワイト・ペーパーはOracle Database Release 11.2.0.3以上に基づいており、ミッション・クリティカルなデータの保護に最適なバックアップおよびリカバリ戦略を実現する設定のベスト・プラクティスについて説明します。

このホワイト・ペーパーでは、次のトピックについて説明します。

- バックアップおよびリカバリのおもなパフォーマンス測定結果とメトリック（特定の構成での性能をDBAがより良く理解できるようにするため）
- Exadata Database MachineとSun ZFS Storage Applianceに対して推奨されるバックアップおよびリストア（B&R）のアーキテクチャ
- バックアップおよびリストアの戦略と構成手法
- パフォーマンスが期待を下回った場合のトラブルシューティング

おもなパフォーマンスの測定結果とメトリック

このホワイト・ペーパーでは、Exadata Database Machine X2-2のフルラック、ハーフラック、クォーターラックの各種構成に対するバックアップおよびリカバリ（B&R）のパフォーマンス・テストについて説明します。このテストは、各種のB&R構成で実現されるパフォーマンスを特定するために実施されました。顧客がZFSおよびExadataシステムのB&Rソリューションを最大限に利用できるようにするため、オラクルはこの結果からいくつかの"目安"を提供します。

ディスク・バックアップとリストアのテストは、InfiniBandおよび10GigEのインフラストラクチャを介してExadata Database Machineに接続されたSun ZFS Storage 7420cアプライアンスを使用して、バックアップ・セット形式を用いて実施されました。

図1に、InfiniBandに接続されたSun ZFS Storage Applianceを使用した場合の各アーキテクチャ・コンポーネントの最大パフォーマンスを示します。

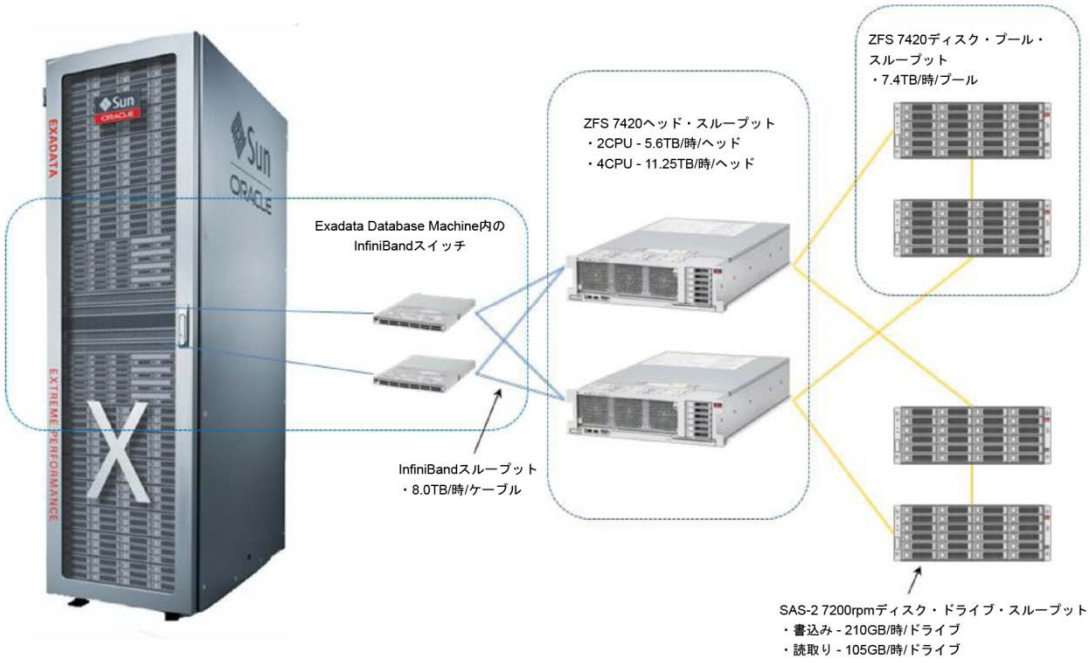


図1：InfiniBand使用時の各コンポーネントのパフォーマンス

図2に、10GigEに接続されたSun ZFS Storage Applianceを使用した場合の各アーキテクチャ・コンポーネントの最大パフォーマンスを示します。

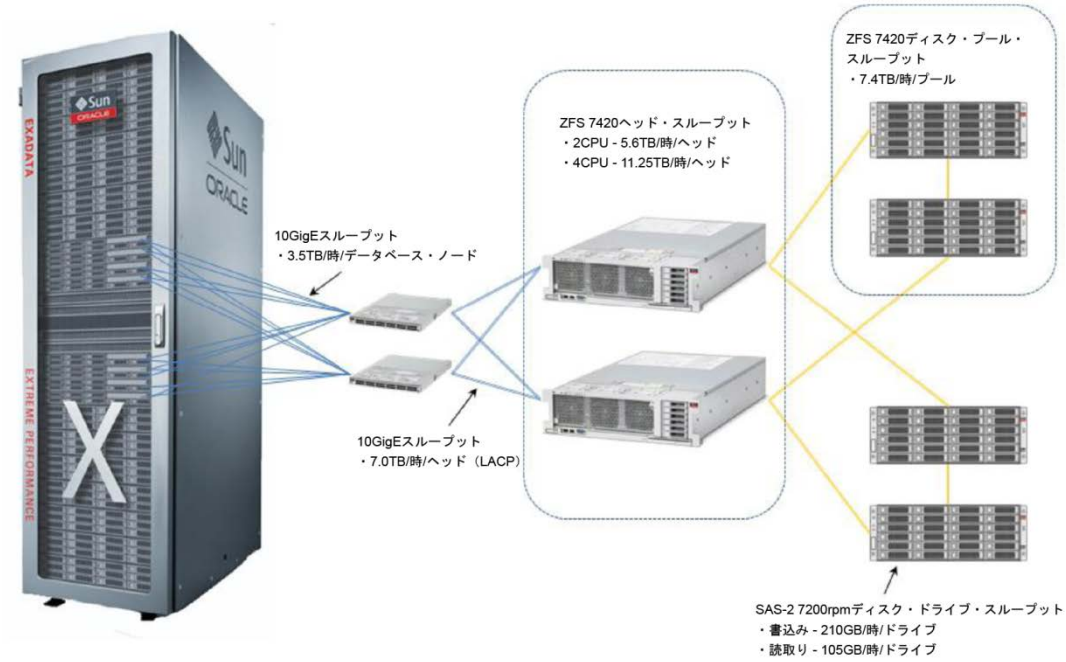


図2：10GigE（LACP）使用時の各コンポーネントのパフォーマンス

表1に、Sun ZFS Storage Applianceベースのバックアップおよびリストアのパフォーマンス結果をまとめたものを示します。クォーターラックを除くすべてのケースで、ZFSアプライアンスのハードウェア構成がB&Rの最大パフォーマンスを左右しています。可用性とパフォーマンスを最大化するには、2個のSun ZFS Storage Applianceヘッドと、ヘッドあたり最低2個のトレイを配置する必要があり、これを使用した結果が次の表に示されています。

表1：ZFSベースのバックアップおよびリストアに対するパフォーマンス測定サマリー

バックアップ・セットを使用した、ディスクへのフル・データベース・バックアップ[2]			
インスタンスとチャネル	クォーターラック	ハーフラック	フルラック
X2-2（11.2.0.3） インスタンス間で均等に分散された 16個のOracle RMANチャネル	4TB/時（10GigE） 4TB/時（InfiniBand）	8TB/時（10GigE） 8TB/時（InfiniBand）	9TB/時（10GigE） 9TB/時（InfiniBand）
ディスクからのフル・データベース・リストア			
X2-2（11.2.0.3） インスタンス間で均等に分散された 16個のOracle RMANチャネル	4TB/時（10GigE） 4TB/時（InfiniBand）	7TB/時（10GigE） 8TB/時（InfiniBand）	8TB/時（10GigE） 9TB/時（InfiniBand）

最適なバックアップおよびリカバリのパフォーマンスは、Exadata Database Machine X2-2上でインスタンスごとに2個または4個のOracle RMANチャネルを割り当て、Exadataに含まれるすべてのデータベース・インスタンス（サーバー）に対してバックアップとリストアを実行した場合に達成されました。Exadata Database Machine X2-8でのB&Rパフォーマンスについては、より詳しい分析を実施しています。最大のB&Rパフォーマンスを達成するために必要なCPUは5%未満でした。

上記の速度は一貫して測定された最大速度です。アプリケーションへの影響を最小化するため、Oracle RMANの"duration minimize load"構文を使用するなどしてバックアップ速度を実質的に下げる方法については、本書の後半で説明します。

アーキテクチャとその説明

Sun ZFS Storage Applianceは、InfiniBandまたは10GigEインフラストラクチャを使用してExadata Database Machineに接続できます。

[2] Exadata Database Machine X2-2のクォーターラックおよびハーフラックの制限要因は、高冗長性のデータ・ディスク・グループに対する読取りと書込みです。

Exadata Database Machine内のInfiniBandスイッチを使用してSun ZFS Storage Applianceに接続する場合、Sun ZFS Storage Applianceの各ヘッドとExadata Database MachineのInfiniBandリーフ・スイッチの間に4本のケーブルが接続されます。

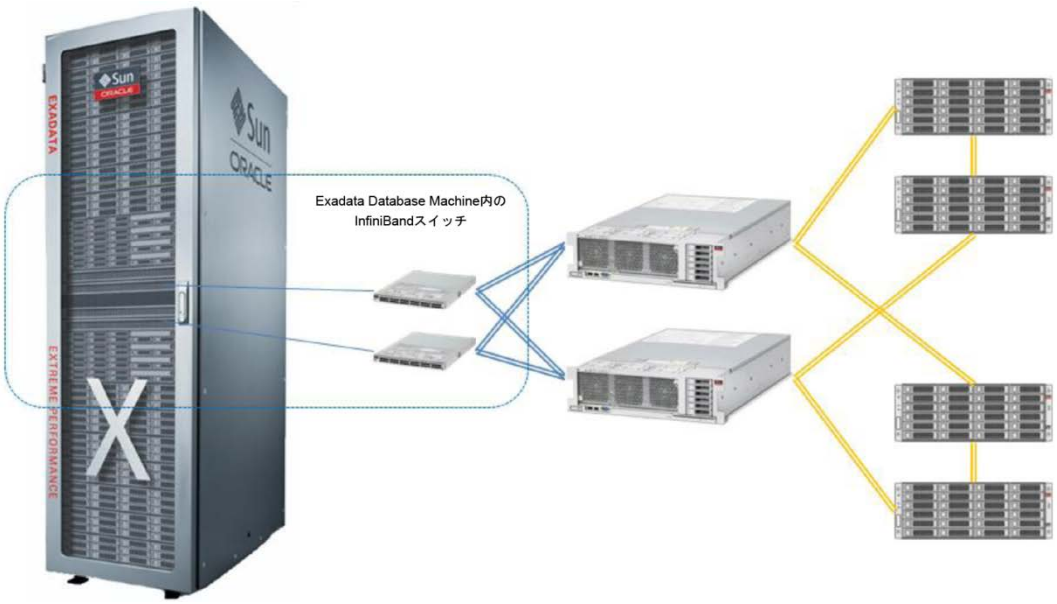


図3：InfiniBandを使用してExadata Database Machineに接続されたSun ZFS Storage Appliance

もう1つの方法として、Sun ZFS Storage Applianceラックのローカルの専用InfiniBandスイッチを使用してSun ZFS Storage Applianceに接続することもできます。この実装ではInfiniBand FatTreeトポロジが作成され、InfiniBandリーフ・スイッチが2個のInfiniBandスパイン・スイッチに接続されます。各ヘッドとSun ZFS Storage ApplianceのローカルInfiniBandリーフ・スイッチの間に4本のケーブルが接続され、4個のInfiniBandリーフ・スイッチと2個のInfiniBandスパイン・スイッチの間にFatTreeトポロジが作成されます。また、各InfiniBandリーフ・スイッチから各InfiniBandスパイン・スイッチ間に4本のケーブルが接続されます。

マルチラック構成での正しいケーブル配線についてはExadata Owners Guideを参照してください。

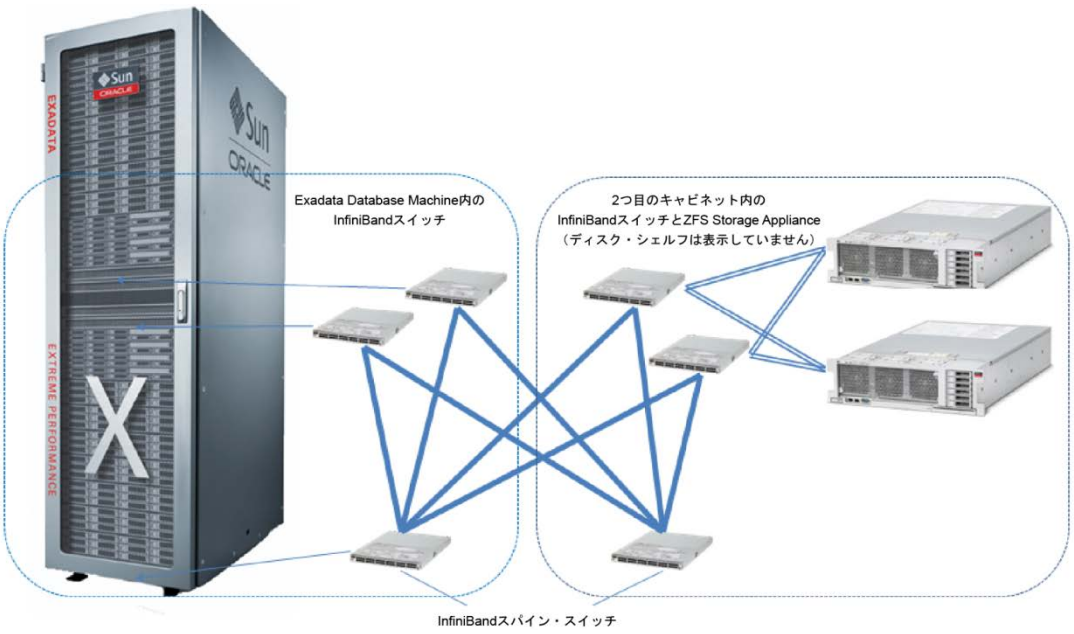


図4：専用InfiniBandスイッチを使用してExadata Database Machineに接続されたSun ZFS Storage Appliance

10GigEを使用してSun ZFS Storage Applianceに接続する場合、Sun ZFS Storage Applianceの各ヘッドと顧客の既存の10GigEネットワーク・インフラストラクチャの間に4本のケーブルが接続されます。同様に、Exadata Database Machine内の各データベース・ノードから2本のケーブルを使用して、顧客の既存の10GigEネットワーク・インフラストラクチャに対してExadata Database Machineが接続されます。顧客の既存10GigEネットワーク・インフラストラクチャを使用すると、複数のExadata Database MachineをSun ZFS Storage Applianceに接続できるため、ハブ・アンド・スポーク設計を作成できます。

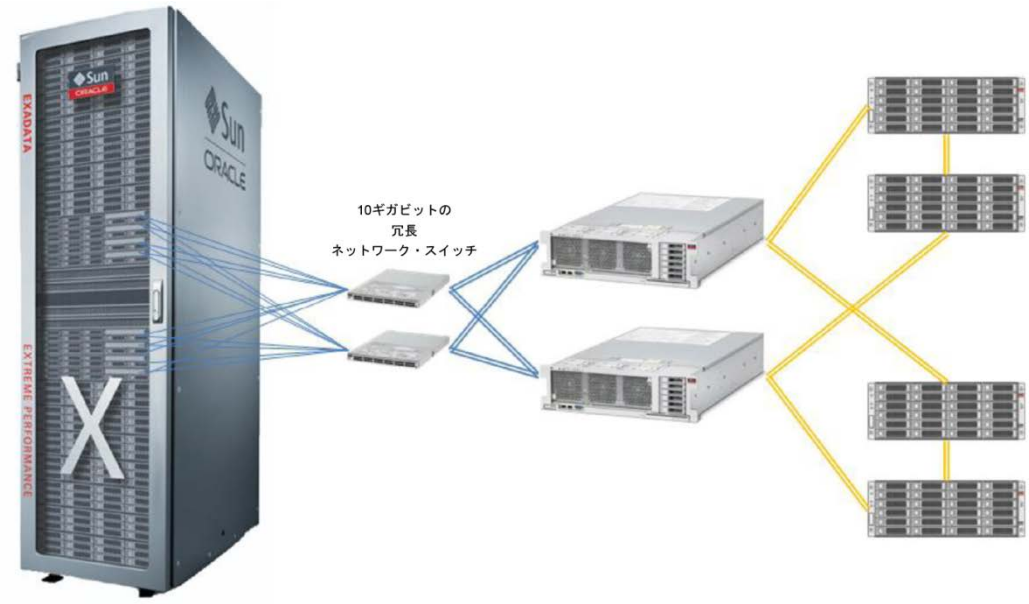


図5：10GigEイーサネットを介してExadata Database Machineに接続されたSun ZFS Storage Appliance

MAAテスト環境

このテストで使用されたSun ZFS Storage Applianceの構成は、次のとおりです。

- Sun ZFS Storage 7420アプライアンス・クラスタ（2つのコントローラ）
- 各ヘッドの構成は次のとおり
  - プロセッサ：2x2GhzのインテルR XeonR CPU X7550（2.00GHz）
  - メモリ：128GB
  - システム・データ：2x500GB
  - キャッシュ・ログ：4x500GB

- SAS-2 HBAコントローラ:2xデュアル4x6Gbの外部SAS-2 HBA
- InfiniBand：2xデュアル・ポートQDR IBのHCA M2
- 10ギガビット・イーサネット：2xデュアル・ポート10Gbの光イーサネット
- クラスタ・カード：1xFishworks CLUSTRON 200

- 4xSun Disk Shelf（SAS-2）、各トレイの構成は次のとおり
  - 2xSTEC/Zeus IOPS 18GBのログ・デバイス
  - 20xSeagate 1.0TBの7200ディスク・ドライブ

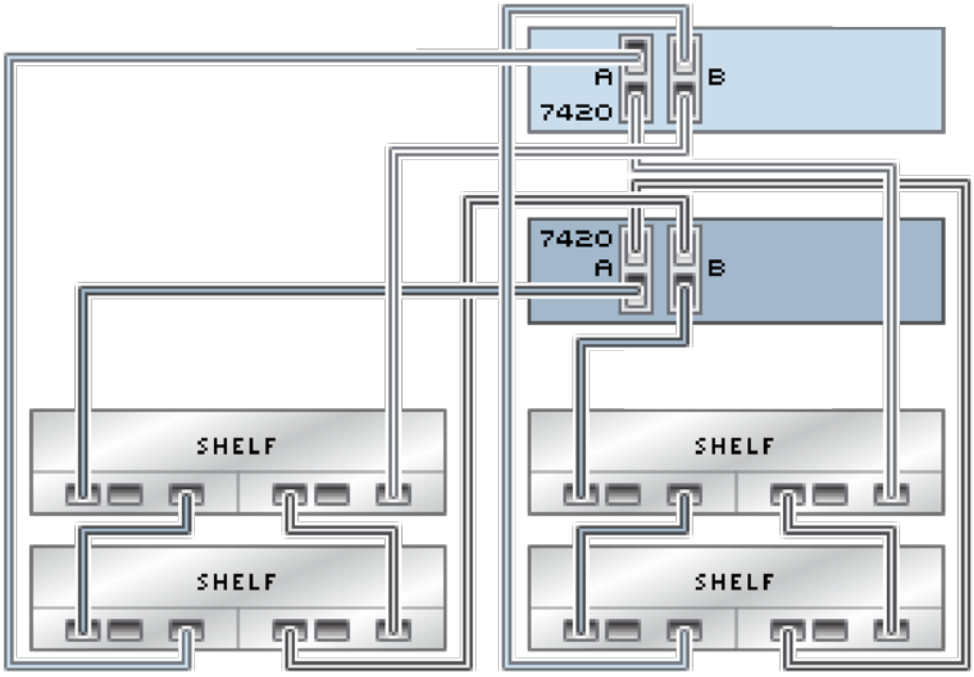


図6：4つのストレージ・トレイを備えたSun ZFS Storage 7420アプライアンス・クラスタ

図6に、クラスタ化されたSun ZFS Storage 7420および2つのSAS HBAコントローラと、各コントローラ・ポートと4つのディスク・トレイ間の接続を示します。詳しくは、『Sun ZFS Storage 7x20 Appliance Installation Guide』を参照してください。

構成のベスト・プラクティス

Sun ZFS Storage Applianceを使用したバックアップおよびリストアにおけるベスト・プラクティスは、次のカテゴリに分類されます。

- Sun ZFS Storage Applianceソフトウェア・スタック
- ネットワーク構成
- ストレージ構成
- バックアップの重要な手法
- バックアップ戦略
- リストアの重要な手法
- リストア戦略
- データの保存
- テスト手法

Sun ZFS Storage Applianceソフトウェア・スタック

Exadata Database Machineと組み合わせて使用されるSun ZFS Storage Applianceに対して推奨されるソフトウェア・スタックの最新情報については、My Oracle Support Note 1354980.1を参照してください。本書の執筆時点では、2011.1.1.1リリースのSun ZFS Storage Applianceソフトウェア・スタックを使用することが推奨されています。

ネットワーク構成

Sun ZFS Storage Applianceのネットワーク構成は、Sun ZFS Storage Applianceを配置する環境とバックアップが必要なExadata Database Machineの数によって異なります。

- 1つまたは2つのExadata Database Machineシステムのバックアップに Sun ZFS Storage Applianceが使用される予定であり、Exadata Database Machineから100メートル以内にSun ZFS Storage Applianceを配置できる場合、InfiniBandを使用してすべてのシステムを同時に接続することが可能です。オラクルの推奨構成では、それぞれのExadata Database Machineシステムを固有のInfiniBand（IB）ファブリックと異なるIBサブネット上に配置します。
- 3つ以上のExadata Database Machineシステムのバックアップに Sun ZFS Storage Applianceが使用される予定であるか、またはExadata Database Machineから100メートル以内にSun ZFS Storage Applianceを配置できない場合、10ギガビット・イーサネットを使用してExadata Database MachineからSun ZFS Storage Applianceへと接続します。

構成を簡素化するため、1つのヘッドを最初に構成してから、Sun ZFS Storage Applianceクラスタに2番目のヘッドを追加します。

InfiniBand接続のためのネットワーク構成

最善の可用性とパフォーマンスを提供するため、各Sun ZFS Storage Applianceコントローラに2枚のデュアル・ポートQDR InfiniBand HCAカードを構成し、Sun ZFS Storage Applianceコントローラごとに合計4個のポートを提供することを推奨します。

Sun ZFS Storage Applianceコントローラは、次のいずれかに影響を与える停止の発生時にも継続的な接続を提供できるように、Exadata Database Machineラック内にあるInfiniBandリーフ・スイッチに接続されています。

- Sun ZFS Storage Appliance内のInfiniBand HCAカード
- Sun ZFS Storage Applianceコントローラ（ヘッド）の障害
- InfiniBandスイッチの障害

これを実現するため、Sun ZFS Storage Applianceネットワーク構成はアクティブ/スタンバイのIPMPグループを使用して構成されます。付録にあるInfiniBand接続のためのネットワーク構成を参照してください。

図7に、異なる2台のExadata Database Machineに接続された2つのSun ZFS Storage Applianceコントローラを示します。これはおそらく、互いに独立した2つの異なる本番アプリケーションが別々のExadata Database Machine上で実行されており、バックアップには共通のSun ZFS Storage Applianceを共有することになっているか、または本番システムからの定期的なリフレッシュが必要なテスト・システムがあるケースにあたります。このような構成に対する要件は、次のとおりです。

- 各Sun ZFS Storage Applianceコントローラに4枚のInfiniBand HCAカードを構成します。
- 2枚のInfiniBand HCAカードを1組として、Exadata Database Machineにあるリーフ・スイッチの空きポートにそれぞれ接続します。
- 2台のExadata Database Machineには異なるInfiniBandサブネットを構成する必要があります。

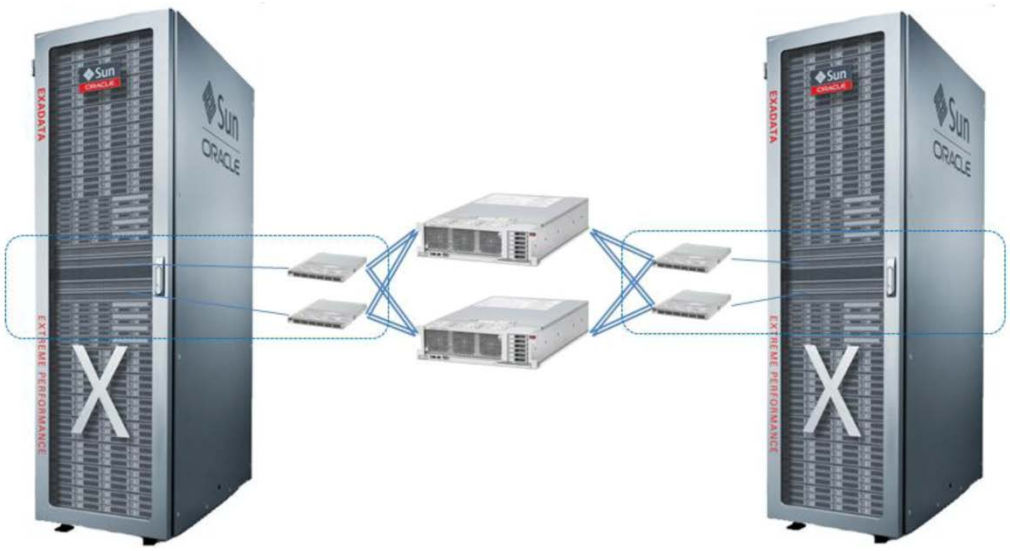


図7：2台のExadata Database Machineと1台のSun ZFS Storage Appliance

10ギガビット・イーサネット接続のためのネットワーク構成

InfiniBandネットワークを使用せずに最善の可用性とパフォーマンスを提供するため、各Sun ZFS Storage Applianceコントローラに2つのデュアル・ポート10Gb光イーサネットを構成し、Sun ZFS Storage Applianceコントローラごとに合計4個のポートを提供することを推奨します。

また、次のいずれかに影響を与える停止の発生時にも継続的な接続を提供できるように、Sun ZFS Storage Applianceコントローラを冗長化された1組の10ギガビット・ネットワーク・スイッチに接続する必要があります。

- Sun ZFS Storage Appliance内の10ギガビットNICカード
- Sun ZFS Storage Applianceコントローラ（ヘッド）の障害
- 10ギガビット・ネットワーク・スイッチの障害

最高のバックアップおよびリストア速度を実現するため、Sun ZFS Storage Applianceネットワーク構成はアクティブ/アクティブのIPマルチパス（IPMP）グループを使用して構成されています。また、パフォーマンス上の理由から、大きなMTUサイズ（ジャンボ・フレーム）に対しては10ギガビット・ネットワークを構成する必要があります。付録にあるInfiniBand接続のためのネットワーク構成を参照してください。

ネットワーク・クラスタ・リソースの構成

10ギガビット・ネットワークまたはInfiniBandのネットワーク・リソース構成が完了したら、インタフェースをSun ZFS Storage Applianceクラスタの管理下に置く必要があります。

ネットワーク・リソースをクラスタ化すると、Sun ZFS Storage Applianceのコントローラ・ヘッドに障害が発生した場合でも、途切れることなくSun ZFS Storage Appliance上のデータにアクセスできます。

詳しくは、付録の"ネットワーク・クラスタ・リソースの構成"の項を参照してください。

ストレージ構成

Sun ZFS Storage Applianceに対するデータベースのバックアップおよびリカバリを実行するには、同じサイズを持つ2個のプール内にSun ZFS Storage Applianceを構成することを推奨します。また、Exadata Database Machineに対して推奨されるバックアップ戦略では、バックアップおよびリストアのパフォーマンスのために、書込みフラッシュ（ログ）デバイスを使用する必要はありません。これは、バックアップおよびリストア処理では、Sun ZFS Storage Applianceヘッドにデータがキャッシュされないためです。ただし、非構造化データ、汎用NFSアクセス、二次的なデータベース処理などの別の処理をサポートするシステムの場合、書込みフラッシュ・デバイスの使用を検討し、これを2つのプール間で均等に分散する必要があります。各プールはローカル・ノード上に作成します。

このテストで使用されたプール構成は、2つのプールから成ります。各ディスク・トレイに含まれるドライブの半分と書込みフラッシュ・デバイスの半数がこのプールに割り当てられ、"Single parity, narrow stripes"というデータ・プロファイルと"Mirrored Log"というログ・プロファイルが構成されました。可能な場合は、"Mirrored Log"プロファイルに対してシングル・ポイント障害なしのオプションを選択します。

作成が完了したプールは、Sun ZFS Storage Applianceクラスタの管理下に置かれます。ネットワーク・リソースとプール・リソースを構成することで、Sun ZFS Storage Applianceコントローラや Sun ZFS Storage Applianceをサポートするネットワーク・コンポーネントに障害が発生した場合でも、Oracle RMANによって実行されているデータベースのバックアップまたはリストアは処理を継続できるようになります。

データベースのバックアップおよびリストアの重要な手法

Sun ZFS Storage Applianceから最大の可用性とパフォーマンスを引き出すには、次の重要な手法に従う必要があります。

- Sun ZFS Storage Applianceでのシェアの作成
- ダイレクトNFSの有効化
- oranfstabの構成
- データベースのバックアップおよびリストアに対するMAA原則の適用

Sun ZFS Storage Applianceでのシェアの作成

Oracle Exadata Backup Configuration Utility v1.0は、Oracle Technology Network（OTN）のSun NAS Storageダウンロード・ページからダウンロードできるプラグインです。このユーティリティはSun ZFS Storage Applianceのプロジェクトとシェア[3]の作成手順を自動化するとともに、クラスタ内のデータベース・ノード上に該当するエントリを作成します。また、データベース・バックアップに使用できるOracle RMANコマンド・ファイルを作成します。

Oracle Exadata Backup Configuration Utilityを実行する前に、前述のとおりに2つのストレージ・プールを作成する必要があります。

このユーティリティには、Exadata Database Machineのバックアップ先としてSun ZFS Storage Applianceを使用する際のベスト・プラクティスがすべて組み込まれています。

- 2つのプロジェクト（1つはプール用）を作成します。
- 書込みバイアスとして"スループット"を使用するようにプロジェクトを構成します。
- Exadata Database Machineからシェアへアクセスできるようにするため、NFS Exceptionsというプロジェクトを構成します。
- バックアップおよびリストア速度を最適化するため、プロジェクトごとに少なくとも8つのシェアを作成します。

詳しくは、Oracle Technology NetworkのSun NAS Storageダウンロード・ページを参照してください[4]。

ダイレクトNFSの有効化

Exadata Database Machine上で稼働するOracle Database 11g Release 2内で利用できるダイレクトNFSオプションを使用して、OracleデータベースをSun ZFS Storage Applianceにバックアップします。ダイレクトNFSオプションはデフォルトでは有効化されておらず、ダイレクトNFSを有効化するにはOracleカーネルを再リンクする必要があります。このため、Exadata Database Machine上で稼働しているデータベースを停止する必要があります。

ダイレクトNFSを有効化すると[5]、LinuxカーネルのNFSプロセスはバイパスされます。また、データベースのバックアップおよびリストアを高速化するためにOracle RMANのみによって使用される1MBのネットワーク・バッファがダイレクトNFSユーティリティによって開かれ、システムで構成されたTCPネットワーク・バッファはバイパスされます。特定のOracleホームでダイレクトNFSを有効化するには、このOracleホームを使用しているすべてのデータベースを停止する必要があります。Oracleソフトウェアの所有者（通常は"oracle"）としてログインします。

---

[3] ZFSシェアはNFSマウント・ポイントへマッピングされ、後からExadata Database Machine上のデータベース・ノードによってマウントされます。ZFSプロジェクトはZFSシェアを論理的にグループ化し、該当するすべてのシェアに影響を与えるプロパティ（ZFSシェアへの書込みを許可されたホストやネットワークの特定を含む）を集中的に構成できるようにします。

[4] http://www.oracle.com/technetwork/server-storage/sun-unified-storage/downloads/index.html

[5] ある顧客では、データベースのリストア処理に対してダイレクトNFSを有効化することで、25%のパフォーマンス向上が確認されました。またダイレクトNFSを有効化した場合に、より一貫性のあるスループットが得られました。

```
$ srvctl stop database -d <データベース名>
```

各データベース・ノードでORACLE環境変数（ORACLE_HOME）を設定してから、次のコマンドを入力します。

```
$ cd ${ORACLE_HOME}/rdbms/lib
$ make -f ins_rdbms.mk dnfs_on
```

以上でデータベースを再起動できますが、ダイレクトNFSを使用してバックアップを実行するデータベースの場合、先に/etc/oranfstabファイルを作成する必要があります。

詳しくは、『Oracle© Databaseインストレーション・ガイド, 11gリリース2（11.2） for Linux』の第5.3.9項"ダイレクトNFSクライアントの構成および使用方法"を参照してください。

/etc/oranfstabの構成

/etc/oranfstabファイルの作成手順は、前述したインストール・ガイドの項で紹介されています。Oracle DBMSでダイレクトNFSを利用するために、oranfstabファイルを作成する必要はありません。ただし、次のいずれかの構成でExadata Database Machine内のデータベース・ノードとSun ZFS Storage Applianceとの間にあるネットワーク・リソースを十分に活用するには、oranfstabファイルを作成する必要があります。

- InfiniBandを使用してExadata Database Machine X2-8からSun ZFS Storage Applianceへ接続する場合、またはExadata Database MachineでSolarisを実行している場合、もしくはSun ZFS Storage Appliance上にアクティブ/アクティブのIPMPグループが作成されている場合

/etc/oranfstabファイルのサンプルについては、付録を参照してください。

データベース・バックアップに対するMAAの原則

OracleデータベースからSun ZFS Storage Applianceへのバックアップを実行する際に推奨される戦略は、レベル0とレベル1のバックアップ・セットを組み合わせて使用する方法です。

- Oracleデータベースのファスト・リカバリ領域はSun ZFS Storage Appliance上に配置せず、Exadata Database Machineの一部であるExadata Storage Server上に配置したままにします。
- レベル0のフル・バックアップを毎週取得し、Sun ZFS Storage Applianceに書き込みます。
- レベル1の増分バックアップを毎日取得し、Sun ZFS Storage Applianceに書き込みます。

レベル0とレベル1のバックアップ取得頻度は、障害回復時のRTO要件とRPO要件によって決まります。ただし、ほとんどの場合は上記の推奨事項で十分です。

Sun ZFS Storage Applianceを使用する場合、Oracle MAA開発チームは増分更新バックアップ戦略を推奨していません。この戦略では、データファイル・コピーと増分バックアップ・セットが組み合わせて使用され、増分バックアップ・セットは後でデータファイル・コピーにマージされます。増分更新バックアップ戦略を選択した場合、バックアップ・ソリューションはマージ・プロセス中のI/Oに制約されるため、このボトルネックを軽減して要求されるIOPS値を達成す

るには、多数のディスク・スピンドルを利用できるようにする必要があります。特定の構成で達成できるSun ZFS Storage Applianceのバックアップおよびリストア速度を測定するため、Exadata MAAおよびSun ZFS Storage Applianceの開発チームによって追加のテストが実施される予定です。テスト構成は、92個の高パフォーマンス600GBドライバと多数の書込みフラッシュ・デバイスで構成される見込みです。

Oracle RMANチャネル

週次のレベル0バックアップおよび日次のレベル1バックアップの両方に対して推奨されるのは、最大16個のOracle RMANチャネルを割り当てることです。これらのチャネルは、前述のステップでOracle Exadata Backup Configuration Utilityによって作成された16個のSun ZFS Storage Applianceシェアのいずれかに対して書込みを行うように構成されます。

Oracle RMAN圧縮

Oracle RMAN圧縮を使用してZFS Storage Appliance上の領域を節約する予定であるが、データベース内に非圧縮データと圧縮データが混在しているか、これ以上圧縮できないデータがある（おそらくデータベース内にLOBデータが格納されている）場合、2つのバックアップ・ジョブを実行する必要があります。1番目のジョブは非圧縮バックアップを作成し、2番目は圧縮バックアップを作成します。一般的に高い圧縮効果を得られないデータを圧縮しようとすると、バックアップ時間が大幅に長くなり、不要なCPUリソースが浪費され、非常にわずかなスペースしか節約できない結果となります。

たとえば、データベースの大部分が非圧縮データで構成されているが、少数の表領域にLOBデータが含まれる場合、LOBデータが格納されている表領域を対象とするinclude部分と、データベースの残りの部分を対象とするexclude部分の2つにバックアップを分ける必要があります。

レベル0の週次Oracle RMANバックアップ・ジョブから一部を抜粋すると、次のようになります。

```
configure exclude for tablespace ‘LOB_DATA1’;
configure exclude for tablespace ‘LOB_DATA2’;
run {
allocate channel ch01 device type disk ….
backup as backupset incremental level 0 section size 32g
filesperset 8 tablespace LOB_DATA1, LOB_DATA2 tag
‘FULLBACKUPSET_L0’;
backup as compressed backupset incremental level 0 section size
32g filesperset 8 database tag ‘FULLBACKUPSET_L0’ plus archivelog;
…
}
```

多くの場合、圧縮を有効化するとバックアップおよびリストアの所要時間が増加するため、引き続きRTO要件を満たすことを確認するにはテストを実施する必要があります。

Oracle RMANのDuration Minimize Loadオプション

2番目の考慮事項は、一定期間にわたってバックアップ処理を制限することで重要なアプリケーションへのパフォーマンス影響を最小化するOracle RMANオプションの使用です。たとえば、20TBのデータベースに対するレベル0の週次バックアップに対して8時間のバックアップ時間枠が利用できるが、この時間中に重要な機能が引き続き実行されている場合、Oracle RMANの`duration minimize load`オプションを使用することで8時間の時間枠に対して負荷を分散することができます。

つまり、3時間以内にバックアップを完了し、ことによるとバックアップ中にアプリケーションへ影響を与える代わりに、Oracle RMANの`duration minimize load`バックアップ・オプションを使用することで、バックアップ時間枠全体を通じて影響を最小化します。

次に、Oracle RMANのバックアップ・ジョブの例を示します。

```
run {
allocate channel …
backup as backupset incremental level 0 section size 32g duration
8:00 minimize load database tag 'FULLBACKUPSET_L0' plus
archivelog;
…
}
```

Oracle RMANとBIGFILE表領域

`sectionsize`をはじめとする特定のOracle RMANオプションの使用が推奨されており、Oracle Exadata Backup Configuration Utilityによって生成されるバックアップ・スクリプトにも含まれています。`sectionsize`パラメータは、BIGFILE表領域を管理しやすいサイズに分割するために特に重要です。非常にサイズの大きい表領域は、それぞれのOracle RMANチャネルに対してサイズが均等なセクションに分割することが推奨されています。たとえば、10TBと14TBという2つの表領域があり、16個のOracle RMANチャネルがある場合、セクション・サイズは約320GBに設定する必要があります。生成されたOracle RMANバックアップ・スクリプトの例は、付録を参照してください。

Oracle Exadata Backup Configuration Utilityでは、Oracle RMANまたはSun ZFS Storage Applianceの圧縮は有効化されません。これらはバックアップするデータや購入したデータベースのライセンス・オプションに依存するためです。また、Oracle RMAN圧縮が有効になっている場合、最大限のOracle RMANチャネルのそれぞれが、実行されているデータベース・ノード上のCPUコアを消費するため、より多くのCPU処理能力がデータベース・ノードで必要になります。

すべてのバックアップ処理とバックアップ手法への変更に対して、使用できる時間内にバックアップが完了し、障害発生時にはRTO要件を満たしたリストアとリカバリが実行できることを確認してください。

データベースのリストアに対するMAAの原則

リストア用システム・リソースの最大化

予期せぬ障害のためにデータベースをリストアする必要が生じた際には、多くの場合、企業のIT部門にとってこれが最大の目的になります。できる限り早急にデータベース・リストアを実行するには、次の措置を講じる必要があります。

- 同じリソースを使用している可能性のあるすべてのバックアップ・ジョブを一時停止します。Sun ZFS Storage Applianceを使用する場合、すべてのシステム・リソースおよびネットワーク・リソースをリストア専用にする必要があります。したがって、その他のバックアップ・ジョブを一時停止し、テスト目的でクローン環境が使用されている場合はこれらの環境も停止する必要があります。
- データベース・インスタンスが通常実行されているデータベース・マシン上のすべてのノードを利用します。詳しくは、"データベース・マシン上の全ノードの活用"を参照してください。

複数のデータベースやアプリケーションがデータベース・マシンを使用している統合環境では、すべてのシステム・リソースをリストア処理用に使用することが受け入れられない場合があります。このような場合、リストアに使用するOracle RMANチャネルの数を減らすことで、使用するシステム・リソースを削減し、実行中のアプリケーションのSLAを満たすことができます。

データベース・マシン上の全ノードの活用

理想を言えば、Exadata Database Machine内のすべてのデータベース・ノードを使用して、Exadata Database Machineの処理能力を均等に利用する必要があります。個別のデータファイルや表領域のリストアが必要なデータベースでは、これは簡単に達成できますが、データベース全体を損失しており、フル・データベース・リストアを実行する必要がある場合は、より困難になります。

データベースが引き続き実行中である場合、前述のバックアップ例と同様にOracle RMANチャネルを割り当てたり、データベースが稼働しているOracleクラスタ内の全ノードで開始されたデータベース・サービスに対してOracle RMANセッションを接続したりできます。後者の場合、並列度を指定すると、Oracle RMANコマンドによって実行中のデータベース・ノード全体に対してチャネルが割り当てられます。

データベース・リストアの定期的な検証

あらゆるバックアップおよびリカバリ環境と同様に、停止時にデータベースをリストアできる場合のみこのソリューションは有効になります。したがって、データベース全体のリストアおよびリカバリとアプリケーション再起動を含むリストア手順を定期的に検証する必要があります。

本番とまったく同じテスト・システムでこの手順をテストできれば理想的ですが、テスト・システムが存在しない場合は、restore validateコマンドを使用して、データベース・バックアップの読取りとリストアを確認できます。ただし、データベースとアプリケーションのリカバリは検証されません。

restore validateコマンドはまた、Sun ZFS Storage ApplianceからExadata Database Machineのデータベース・ノードに対して、どのくらいの速さでデータを読み取れるかを評価するためにも使用できます。

たとえば、Exadata Database Machineハーフラック内の4つのノードすべてで`orcl_restore`というデータベース・サービスが実行されている場合、次のコマンドを使用すると、データベース・ノード間に8つのチャネルが割り当てられ、データベース・リストアの検証が実行されます。

```
$ rman target sys/welcome1@dm01-scan/orcl_restore

RMAN> configure device type disk parallelism 8;

RMAN> restore validate database;
```

一般的な手法としては、完全なリストアおよびリカバリのテストを少なくとも毎月実施し、リストアの検証を毎週実行します。

データベースのバックアップおよびリカバリのテスト

データベース・バックアップが存在しているだけでは、障害の発生時にそれを使用できるという保証にはなりません。したがって、ベスト・プラクティスとして、データベースのリストアおよびリカバリ手順を文書化して定期的にこれを検証することで、この手順が引き続き有効であり、アプリケーションのRTOとRPOが満たされることを確認する必要があります。

アプリケーション統合やビジネス成長のためにリストア対象のデータベースが増大していたり、アプリケーションが変更されて圧縮データや暗号化データが活用されていたりする場合、上記はあてはまりません。データベース・バックアップでOracle RMAN圧縮を活用している場合、新しく圧縮または暗号化されたデータのリストアおよびリカバリによって、アプリケーションのRTOを満たす能力に悪影響が及ぶ可能性があります。このため、リストア/リカバリ処理が正しく機能しており、RTO/RPO要件を引き続き満たすことを確認する必要があります。

- 週次のリストア検証テスト
- 定期的なリストアおよびリカバリ・テスト
- データファイル/表領域の新規追加またはバックアップ・オプションの変更（圧縮オプションや暗号化の追加または変更）を実施した場合は、必ず直後に再評価を実施

実際、データベースのサイズや構成、またはZFSのアーキテクチャおよび構成に何らかの変更があった場合、リストアおよびリカバリ処理を再評価して、データがリカバリ可能であり、RTO要件が満たされることを確認する必要があります。

HAテスト

Sun ZFS Storage Applianceは高可用性向けに構成されています。障害発生時にも可用性を提供するため、InfiniBandまたは10GigEによるネットワーク接続は冗長スイッチに接続されており、ヘッドに影響がおよぶ障害発生時には、存続するZFSコントローラへネットワーク・リソースとディスク・リソースがフェイルオーバーすることが、実施されたクラスタ構成手順によって確認されています。

次の図では、Exadata Database Machine X2-8の2つのデータベース・ノードに対するネットワーク・スループットの集計値がグラフ化されており、1つのヘッドの再起動時に発生したパフォーマンス低下が示されています。当初、スループットは低下しますが、ZFS Storage Arrayが完全に稼働した時点で元のスループットの約50%まで回復します。

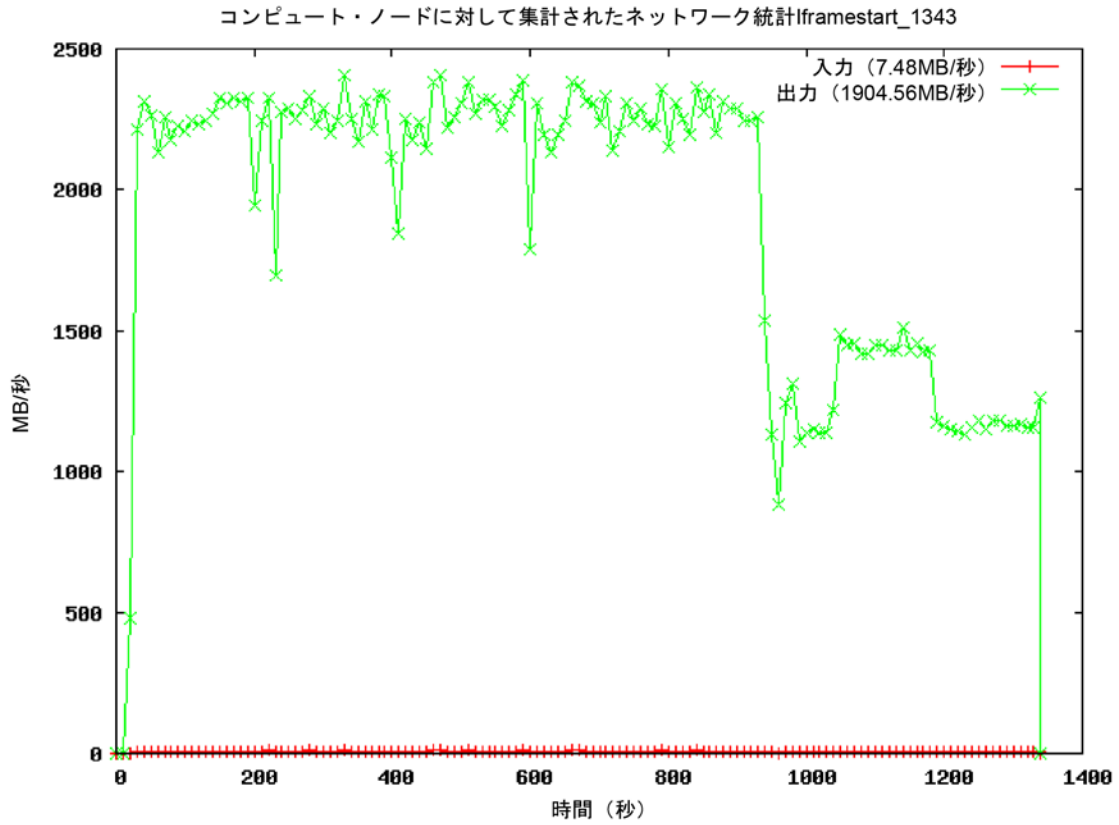


図8：ZFSコントローラの障害時のネットワーク・スループット統計

バックアップ速度は低下しますが、Oracle RMANバックアップ自体への影響はなく、バックアップ・ジョブはエラーなしで完了します。

データの保存

データベース・バックアップの保存方針は、従来のファスト・リカバリ領域使用時と同じ方法で構成されますが、古くなったバックアップの消去は、Sun ZFS Storage Applianceにバックアップが書き込まれた時点でマニュアル実施します。

データベースの保存方針

保存方針は、コピーの数または期間のいずれかで指定します。たとえば、レベル0のバックアップを常に2つ維持する予定である場合、保存方針には2という冗長レベルを構成します。

```
RMAN> configure retention policy to redundancy 2;
```

古いデータベース・バックアップの削除

目的の保存方針をサポートするために必要でなくなったデータベース・バックアップは、Oracle RMANでは古くなったバックアップとみなされます。Oracle RMANスクリプト内でdelete obsoleteコマンドを使用すると、不要になったバックアップを消去できます。

```
RMAN> delete obsolete;
```

バックアップやリストアのパフォーマンスが不十分である理由

この項では、トラブルシューティング、監視、テストのおもな要件について説明します。

backup validateを使用したExadata Database Machineの読取り速度のテスト

バックアップのパフォーマンスが十分でない場合に最初にわくのは、"データが十分な速度でセルから取得されているか"という疑問です。

この疑問は、レベル0の週次バックアップといったOracle RMANのbackupコマンドを使用することで簡単に解決できます。ただし、Sun ZFS Storage Appliance上のバックアップ場所にデータを書き込むかわりに、validate句を追加することは、Sun ZFS Storage Applianceへのデータの作成と書込みを除くすべてのOracle RMANコードを実行することを意味します。同様に、Oracle RMANの圧縮または暗号化が使用されている場合は、Oracle RMANのbackup validateコマンドに対してこれらのオプションを追加または削除することで、オプションによる影響を評価できます。

ジョブが完了したら、Oracle RMANのログ・ファイルを分析して、バックアップ開始時間とバックアップ終了時間からバックアップ・ジョブにかかった時間を特定できます。

```
Starting backup at 07-Feb-2012 06:30:47

Finished backup at 07-Feb-2012 07:15:19
```

Oracle RMANは、環境変数`NLS_DATE_FORMAT`によって定義された形式を使用してログ・ファイルに時間を書き込みます。

また、固定ビュー`$backup_async_io`に対して問合せを実行すると、各Oracle RMANチャネルに対する"1秒あたりの有効バイト数"を取得でき、OSWatcherログを分析すると、バックアップによってデータベース・ノードやExadataストレージ・セルにかかる負荷を確認できます。

バックアップを使用したExadata Database Machineの書込み速度のテスト

バックアップ読取り速度を検証したら、次にデータベースのバックアップ処理を実行して、Sun ZFS Storage Applianceへの書込み速度を検証します。

ここでも、Oracle RMANのログ・ファイルを分析することで、バックアップ開始時間と終了時間からバックアップにかかった時間を特定できます。また、ログ・ファイルから、時間がかかり過ぎているように見える特定のバックアップ・フェーズを識別できる場合もあります。

```
Starting backup at 07-Feb-2012 08:43:53

Finished backup at 07-Feb-2012 10:31:19
```

データベース・マシンの観点から見ると、今回も`v$backup_async_io`固定ビューとOSWatcherログを分析することで目的を果たせます。

OSWatcherログを使用すると、バックアップによってデータベース・ノードとExadata Database Machineストレージ・セルにかかる負荷だけでなく、データベース・ノードからの書込み時にネットワーク・スタックにかかる負荷を分析することもできます。

また、Oracle Sun ZFS Storage ApplianceでAnalytics製品を使用する方法について、Sun ZFS Storage Applianceでのリアルタイム監視およびリアルタイム・アラートに関する項を参照してください。

ネットワーク・スループットの評価

データベース・リストアの重要な手法に関する前述の項では、データベースのリストア検証テストを定期的に実行することで、Sun ZFS Storage Appliance上のバックアップの整合性を確認するという考え方を紹介しました。同じ手順を使用して、ネットワーク・スループットを評価できます。

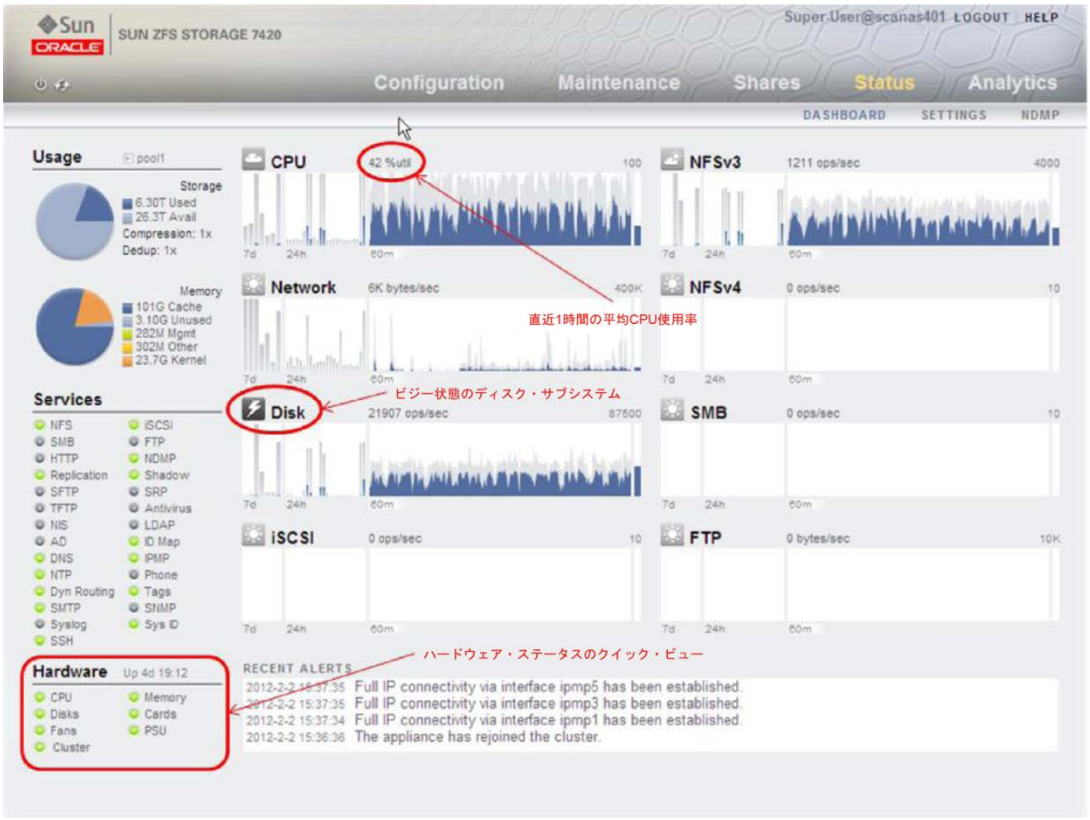
ZFSアプライアンスでのボトルネック評価

Sun ZFS Storage ApplianceにはAnalyticsと呼ばれるパフォーマンス監視機能とアラート・システムが組み込まれています。

リアルタイム監視

Sun ZFS Storage Applianceに最初にログインすると、Status→Dashboard画面がデフォルトで表示されます。この画面ではSun ZFS Storage Applianceに対するクイック・ビューが提供されており、ストレージ使用率、サービスの状態、ハードウェア・コンポーネントの状態などを確認できます。また、CPU、ネットワーク、ディスクを含む各種コンポーネントに対して、7日間、24時間、60分、およびリアルタイムのグラフが多数表示されます。

次のスクリーンショットは、クラスタに含まれる2つのヘッドのうちの1つに対するSun ZFS Storage Applianceダッシュボードを示しています。



CPUグラフでハイライトされているのは1時間のビューであり、42%という数値は、直近1時間の平均CPU使用率が42%であったことを示しています。

Diskという見出しの横には稲妻マークが表示されています。Sun ZFS Storage Applianceには各コンポーネントに対する特定のしきい値が事前に構成されており、しきい値ごとに異なる"嵐"インジケータが、晴れからくもり、雨を経てさまざまな強さの嵐を示します。各コンポーネントの横に表示されたアイコンを見るだけで、管理者はシステムが現在どの程度のビジー状態にあるかを把握できます。

ストレージ使用率の円グラフとサービス・ステータスの下に表示されているのは、ハードウェア・ステータスです。これらのコンポーネントに影響を与える障害の発生時には緑のインジケータが黄色に変わり、コンポーネントに問題があることを示します。詳しい情報は、MaintenanceのProblemsタブに表示されます。

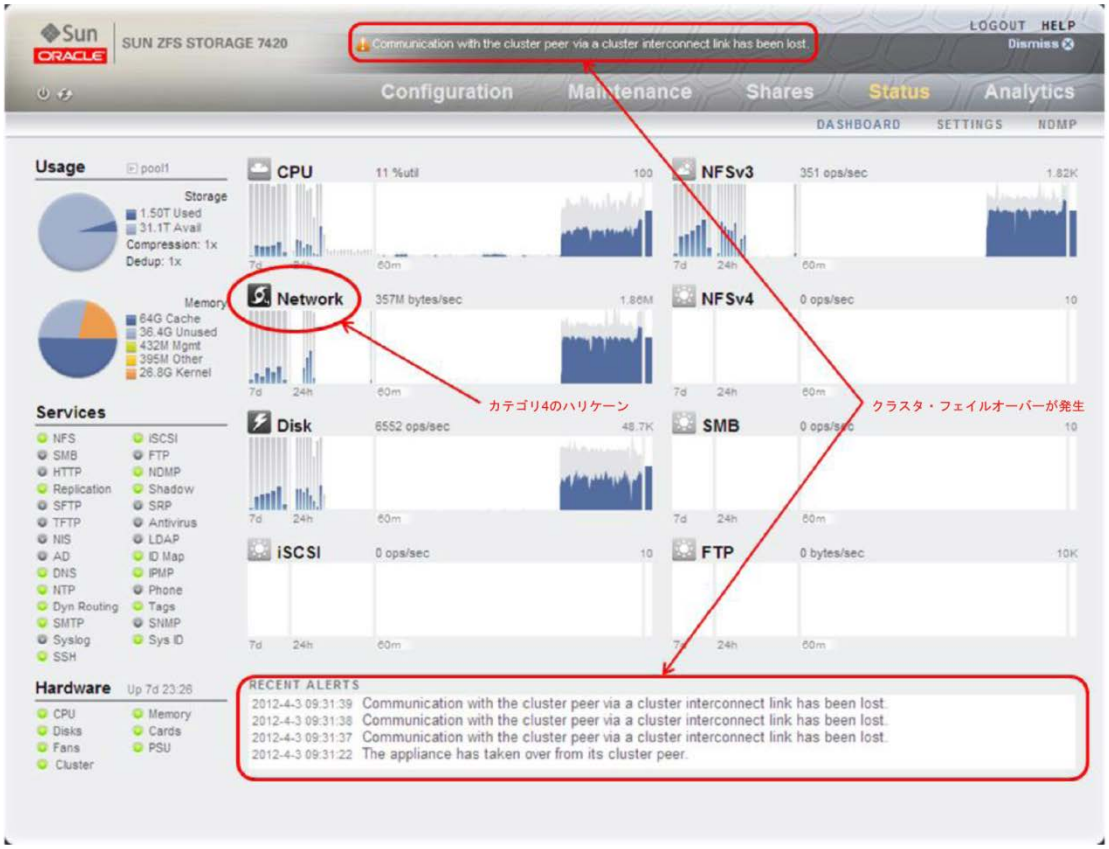
画面下部にはシステムで発生した最新のアラートが表示されます。

2番目のスクリーンショットは、クラスタに含まれる2つのヘッドのうちの1つに障害が発生した直後の、残りのヘッドに対するSun ZFS Storage Applianceダッシュボードを示しています。

画面上部のウィンドウに最新のアラートが表示されており、画面下部にも"The appliance has taken over from its cluster peer"というメッセージが表示されています。

この画面は、フェイルオーバーが実行されてから約2分後に取得されたものであり、ネットワークの"嵐"インジケータには、カテゴリ4のハリケーンが表示されています。このアイコンをクリックすると構成ダッシュボード画面が表示さ

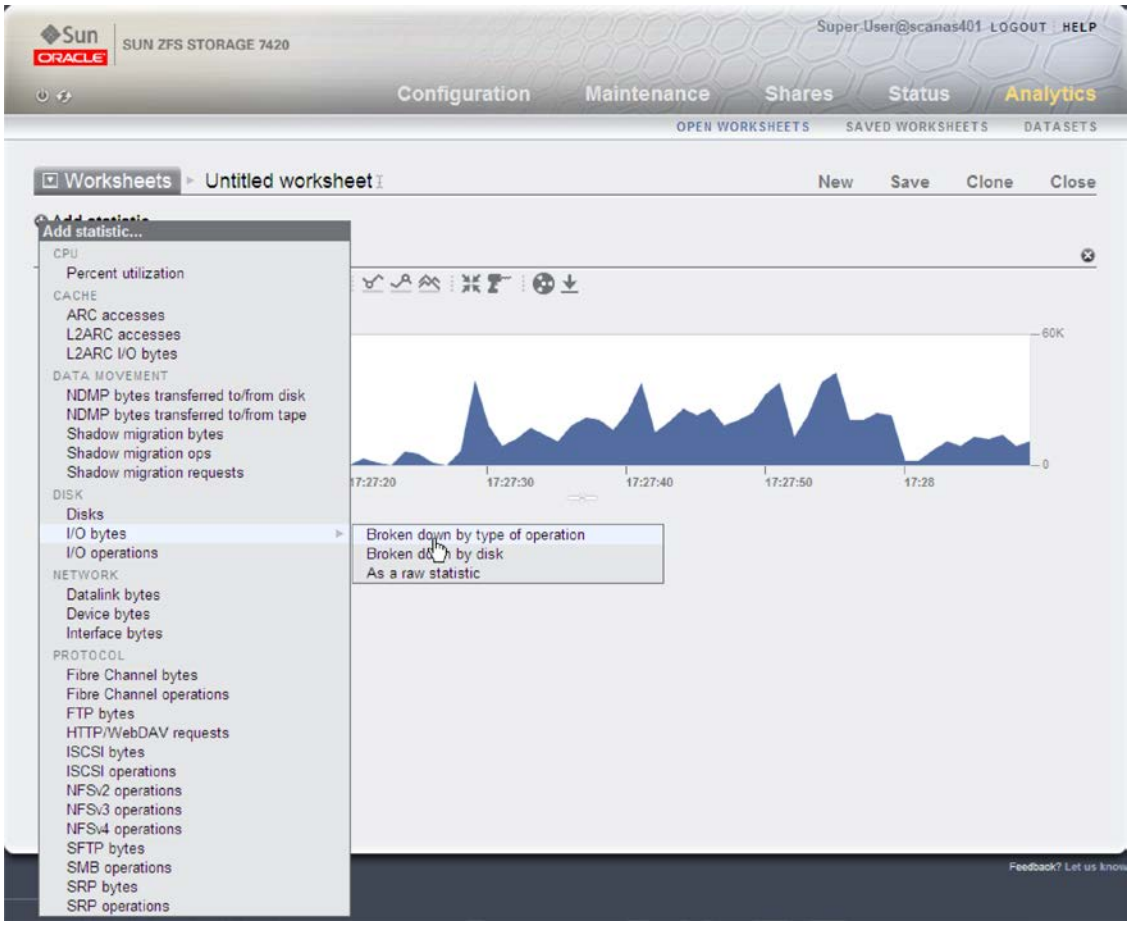
れ、カテゴリ4のハリケーンに対するしきい値が1.3GB/秒に設定されています。つまり、最近60秒での平均ネットワーク・スループットが1.3GB/秒を超えていることが分かります。



いずれかのアクティブなグラフをクリックすると、Sun ZFS Storage Applianceのリアルタイム監視機能であるAnalyticsが表示されます。Analyticsは画面上部のメニューをクリックしても表示されます。

以前に保存したAnalyticsワークシートを開いたり、Add statisticsの横にある「+」記号をクリックして各種の統計を選択したりできます。

次のスクリーンショットには画面上にあらかじめ表示された特定のメトリックとともに、各種の統計クラス、メトリック、およびサブメトリックが表示されています。このスクリーンショットでは、ディスクのI/Oバイトが処理タイプ（読取り/書込み）によって分類されています。



Sun ZFS Storage Appliance Analyticsについて、詳しくはSun ZFS Storage Appliance Analyticsページからアクセスできるオンライン・ヘルプを参照してください。

リアルタイム・アラート

Sun ZFS Storage Applianceにはアラート・システムが組み込まれており、電子メールやSNMPを使用した監視アラートの送信やSYSLOGを使用した一元化ロギング・システムへの書込みを実行できます。このアラート・システムは、ネットワーク（リンクのパフォーマンス低下または障害発生時に表示）、ハードウェア障害、クラスタ・イベント（クラスタのフェイルオーバーまたはテイクオーバー）などのさまざまなカテゴリを監視します。

また、必要に応じてNDMPやリモート・レプリケーション処理の進捗を報告します。

バックアップおよびリストア速度に影響を与えるその他の要因

バックアップの重複

当然のように思われるかもしれませんが、本書で取り上げたバックアップおよびリストア速度はExadata Database MachineではなくSun ZFS Storage Applianceによって制限されていることは注目に値します。ExadataおよびExadata以外の多数のデータベースに対する集中バックアップ・ソリューションを提供する目的でSun ZFS Storage Applianceが配置されている場合、バックアップが重複しないようにするか、バックアップが競合してもSLAが満たされるように、バックアップ・スケジュールを計画する必要があります。

たとえば、Sun ZFS Storage Applianceは約9TB/時の速度で書込みを実行できますが、2つのバックアップが同時に実行された場合、最大9TB/時の速度は2つのバックアップ間で良くても均等に分散されます。したがって、9TBのデータベースと18TBのデータベースが同意に実行されるようにスケジューリングされている場合、両方のデータベースのSLAに対して、3時間以上のバックアップ時間枠を指定する必要があります。

バックアップの圧縮

バックアップの圧縮は、通常、Sun ZFS Storage Applianceなどのバックアップ・ターゲット上の領域を節約したいという要求があった場合に検討されます。圧縮によって使用される領域が削減される可能性はありますが、圧縮を有効化するとデータの圧縮にCPUサイクルを実行する必要が生じるため、ほとんどの場合でバックアップの実行時間が長くなります。圧縮は、バックアップ・ソリューションに含まれるさまざまな個所で有効化できます。

- Oracle Database 11g Release 2 Enterprise Editionに付属している基本圧縮を使用した、Oracle RMANレベルでの圧縮
- Oracle Database 11g Release 2のAdvanced Compression Option（ACO）に付属しているLOW、MEDIUM、またはHIGH圧縮を使用した、Oracle RMANレベルの圧縮
- LZJB、GZIP、GZIP-2、GZIP-9のいずれかの圧縮アルゴリズムを使用した、Sun ZFS Storage Applianceのプロジェクト・レベルまたはシェア・レベルの圧縮

圧縮がどこで（Oracle RMANまたはZFS Storage Appliance）有効化されるかに関係なく、圧縮を実行するデータベース・ノードまたはSun ZFS Storage Applianceコントローラは CPUの制約を受ける可能性があります。これを確認するには、データベース・ノードのOSWatcherログを分析して最初の部分またはmpstatログを確認するか、またはSun ZFS Storage ApplianceのAnalyticsを使用してCPU使用率を確認します。

- データベース・ノード上のOracle RMANで圧縮が有効化される場合、Sun ZFS Storage Applianceに送信されるデータは少なくなります。
- Sun ZFS Storage Appliance上で圧縮が有効化される場合、データベース・マシンのデータベース・ノードに追加のCPU負荷はかかりませんが、Sun ZFS Storage Appliance上のCPUがよりビジーになるか、またはシステムがCPUの制約を受けます。

リストア時の解凍

バックアップに対して圧縮が有効化されていると仮定した場合、リストア時に同じ場所で解凍が実行されます。

しかし、データベース・リストア中の最大の目標は、おそらく、できる限り早急にリストアを完了してRTOとRPOを満たすことです。つまり、Oracle RMANの並列度を上げる必要が生じる場合があります。

- Oracle RMANレベルで圧縮が実行されている場合、追加のOracle RMANチャネルが新たなCPUコアを消費するため、十分な速度でデータが提供されると考えられます。
- Sun ZFS Storage Applianceレベルで圧縮が実行されている場合、Oracle RMANチャネルの数を増やすとSun ZFS Storage Applianceのコントローラ・ヘッドでボトルネックが発生し、追加のOracle RMANチャネルはストリーミング・データを受け取れない場合があります。
- データベースのバックアップ・レベルとリストア・レベルの両方でテストを実施し、圧縮が有効化されている場合もバックアップのSLAとリストアのRTO/RPOが満たされることを確認する必要があります。

I/Oリソースの競合

常に、その他のアクティビティがシステム上で実行されていない負荷の少ない時間にデータベース・バックアップをスケジューリングできるとは限りません。Exadata Database Machine上でデータ・ロードが実行されていたり、Sun ZFS Storage Applianceのスナップショットやクローンが作成されており、ここでテストが実行されていたりするかもしれません。Sun ZFS Storage Applianceは、バックアップと競合する固有のI/O要件および動作を持つアプリケーションのデータやログをホストしている場合があります。バックアップをスケジューリングする際、これらの付加処理を考慮に入れる必要があります。

Exadata Database Machineでは、リソース・マネージャを使用して各種処理のI/Oに優先順位を付けることで、特定の時間内にバックアップが完了するようにバックアップ処理を優先したり、または、朝、"オンライン"ユーザーがシステムにアクセスする時点でシステムが利用可能な状態になっているように、データ・ロードを優先したりできます。

Sun ZFS Storage Appliance上の空き領域と断片化

その他のCopy-On-Writeソリューションと同様に、Sun ZFS Storage Applianceプール内のディスクにある空き領域すべてを使い果たすことは得策ではありません。Oracle Exadata MAAおよびSun ZFS Storage Applianceの開発チームが推奨するのは、作成された各ディスク・プールに20%の空き領域を維持することです。こうすることで、コピー処理と書込み処理を最適な状態で実行できます。これを実現するには、Sun ZFS Storage Applianceのプロジェクトに対して割当て制限を設定します。

空き領域が20%という推奨値を切ると、アプライアンスは空き領域を探すために長い時間を費やさざるを得なくなるため、Sun ZFS Storage Applianceへの書込み処理に影響が及ぶ場合があります。

また、断片化は読取り処理と書込み処理の両方に影響を及ぼします。Sun ZFS Storage Applianceが最初に配置された時点では、アプライアンスは空のディスクに対してデータを順番に書き込むことができますが、データの保存方針がいった

ん満たされると（4週分の週次バックアップおよび増分日次バックアップとアーカイブ・ログをSun ZFS Storage Applianceに保存など）、古いデータが消去されて新しいデータが書き込まれるため、ディスク領域が断片化する可能性があります。断片化による影響は、テスト目的でスナップショットやクローンが作成された場合や、本書の推奨に反して増分更新バックアップ・ソリューションが実装された場合に顕著になります。

結論

Exadata Database MachineとSun ZFS Storage Applianceを組み合わせることで、非常に高速で費用効果に優れたバックアップおよびリストア・ソリューションを提供できます。MAAとZFSの構成および運用手法を一体化することで、顧客はこのソリューションを構成して目的とするHA利点を実現すると同時に、バックアップおよびリストア速度の低下につながる一般的な落とし穴を見つけ、回避できるようになります。

付録A：Sun ZFS Storage Applianceの構成

InfiniBand接続のためのネットワーク構成

最善の可用性とパフォーマンスを提供するため、各Sun ZFS Storage Applianceコントローラに2枚のデュアル・ポートQDR InfiniBand HCAカードを構成し、Sun ZFS Storage Applianceコントローラごとに合計4個のポートを提供することを推奨します。

2枚のInfiniBandカードはおそらくSun ZFS Storage ApplianceコントローラのPCIeスロット4および5にインストールされ、PCIeスロット4内のibp0とibp1およびPCIeスロット5内のibp2とibp3として識別されるでしょう。これらの2個のスロットは、縦に設置されたPCIeスロットの上部ポートと下部ポートとみなすことができます。Sun ZFS Storage Applianceのユーザー・インタフェースでConfiguration→Networkから、ibp0とibp3をアクティブ/スタンバイのIPMPグループとして、ibp2とibp1を2番目のアクティブ/スタンバイIPMPグループとしてバインドします。

InfiniBand接続のケーブル配線

Sun ZFS Storage Applianceの各コントローラ・ヘッドからExadata Database Machine内の2つのInfiniBandリーフ・スイッチに対して、4本のケーブルを配線します。InfiniBandリーフ・スイッチ上には6個のポートがあり、バックアップおよびリカバリ目的でSun ZFS Storage ApplianceからExadata Database Machineへ接続する場合などのために、顧客が自由にInfiniBandファブリックを拡張できます。このホワイト・ペーパーの目的に対して、ここでは5A、6A、6B、7Aという印の付いたポートを使用しますが、別のデバイスがすでにこれらのポートを使用している場合、InfiniBandスイッチ上で使用できる6個のポート（5B、6A、6B、7A、7B、12A）のうちのいずれを使用しても構いません。

メディア・サーバー、データ・ミュール、または中間層アプリケーション・システムなどのデバイスのせいでInfiniBandスイッチに空きがない場合、追加のInfiniBandスイッチを購入し、Database Machine Owners Guideで説明されているケーブル配線方法を使用して2つのラックを相互接続できます。

次の表に示すとおりにケーブルを接続します。

表2. InfiniBandのケーブル配線マップ

Sun ZFS Storage Applianceヘッド	Sun ZFS Storage Applianceポート	モード	InfiniBandスイッチ	InfiniBandポート
ヘッドA	ibp0	アクティブ	リーフ・スイッチ#2	5B
ヘッドA	ibp1	スタンバイ	リーフ・スイッチ#3	6A
ヘッドA	ibp2	アクティブ	リーフ・スイッチ#2	6A
ヘッドA	ibp3	スタンバイ	リーフ・スイッチ#3	5B
ヘッドB	ibp0	アクティブ	リーフ・スイッチ#2	6B
ヘッドB	ibp1	スタンバイ	リーフ・スイッチ#3	7A

ヘッドB	ibp2	アクティブ	リーフ・スイッチ#2	7A
ヘッドB	ibp3	スタンバイ	リーフ・スイッチ#3	6B

ネットワークの構成 - デバイス、データ・リンク、インタフェース

Sun ZFS Storage Applianceのネットワーク構成は、デバイス、データ・リンク、インタフェースという3つのレベルで構成されています。

- デバイスは物理ハードウェアを表します。
- データ・リンクは物理デバイスを管理します。
- インタフェースはデータ・リンク上のIPアドレスを構成します。

ネットワーク・データ・リンクの構成

InfiniBandポートが使用するネットワーク・データ・リンクは、次のとおりに構成します。

Configuration→Network画面に移動します。

Datalinks列にデバイス名をドラッグするか、またはDatalinksの横にある「+」記号をクリックして、新しいデータ・リンクを構成します。

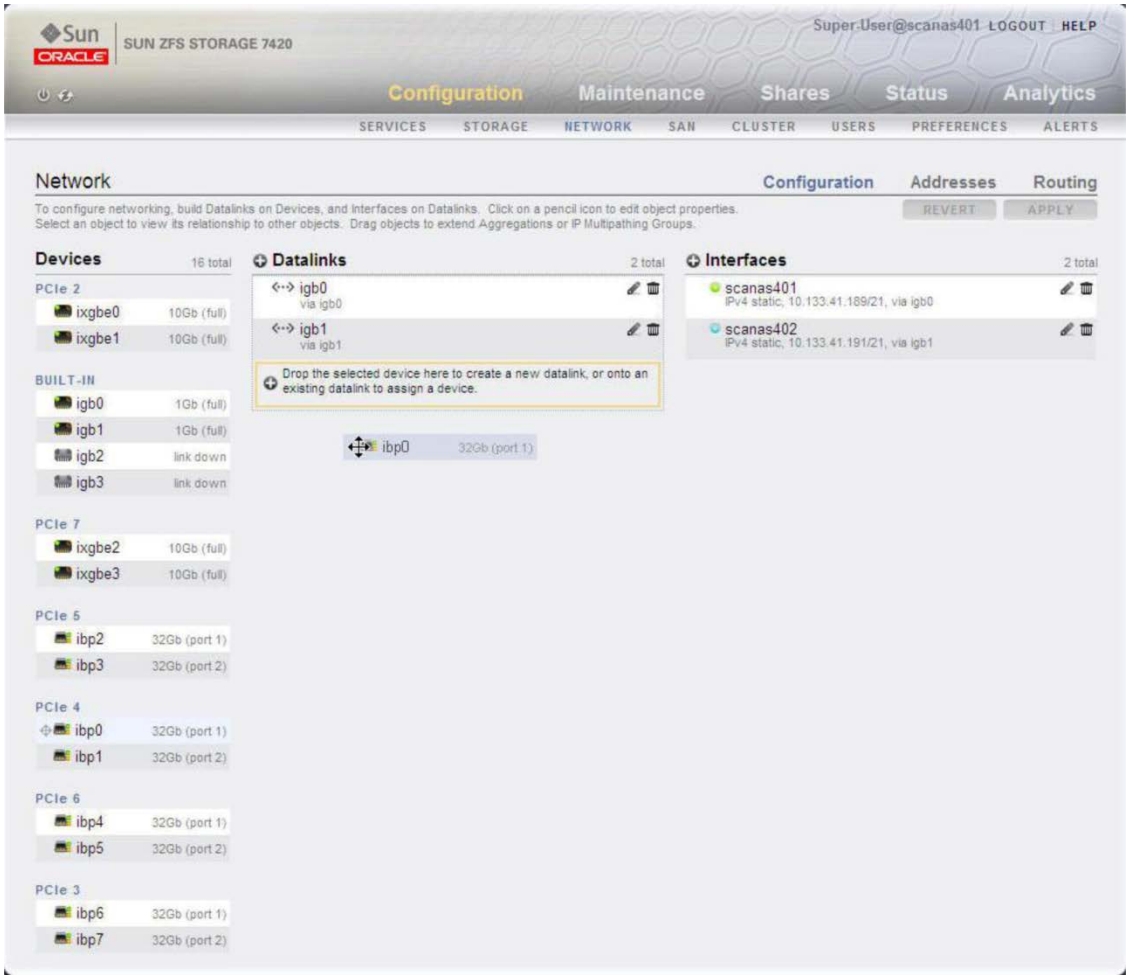


図9：ネットワーク・データ・リンクの新規追加

- 「**IB Partition**」チェック・ボックスを選択します。
- 分かりやすくするため、データ・リンクはデバイス名と同じ名前で作成します。
- 使用するPartition Keyとして"ffff"と入力します。
- パフォーマンスを最適化するため、Link Modeのドロップダウン・ボックスから「**Connect Mode**」を選択します。
- 追加するデバイスに関連付けられたラジオ・ボタンをクリックします。

画面は次のようになります。

図10：InfiniBandネットワーク・データ・リンクの構成

デバイスを管理するその他3つの新規データ・リンクに対して、同じ手順を繰り返します。

ネットワーク・インタフェースの構成

可用性のためにIPMPグループを使用するため、ネットワーク・インタフェースの構成には2つのフェーズがあります。

フェーズ1：ネットワーク・インタフェースの構成 - 静的アドレス

データ・リンクの構成と同じように、インタフェース列にデータ・リンク名をドラッグするか、または「+」記号をクリックして、新しいインタフェースを構成します。

- 分かりやすくするため、インタフェース名=データ・リンク名とします。
- 「**Use IPv4 Protocol**」チェック・ボックスを選択します。
- "Configure with"ドロップダウン・ボックスから「**Static Address List**」を選択します。
- Configure withボックスの下にあるアドレス・ボックスに"0.0.0.0/8"と入力します。

- 追加するデータ・リンクに関連付けられたラジオ・ボタンをクリックします。

画面は次のようになります。

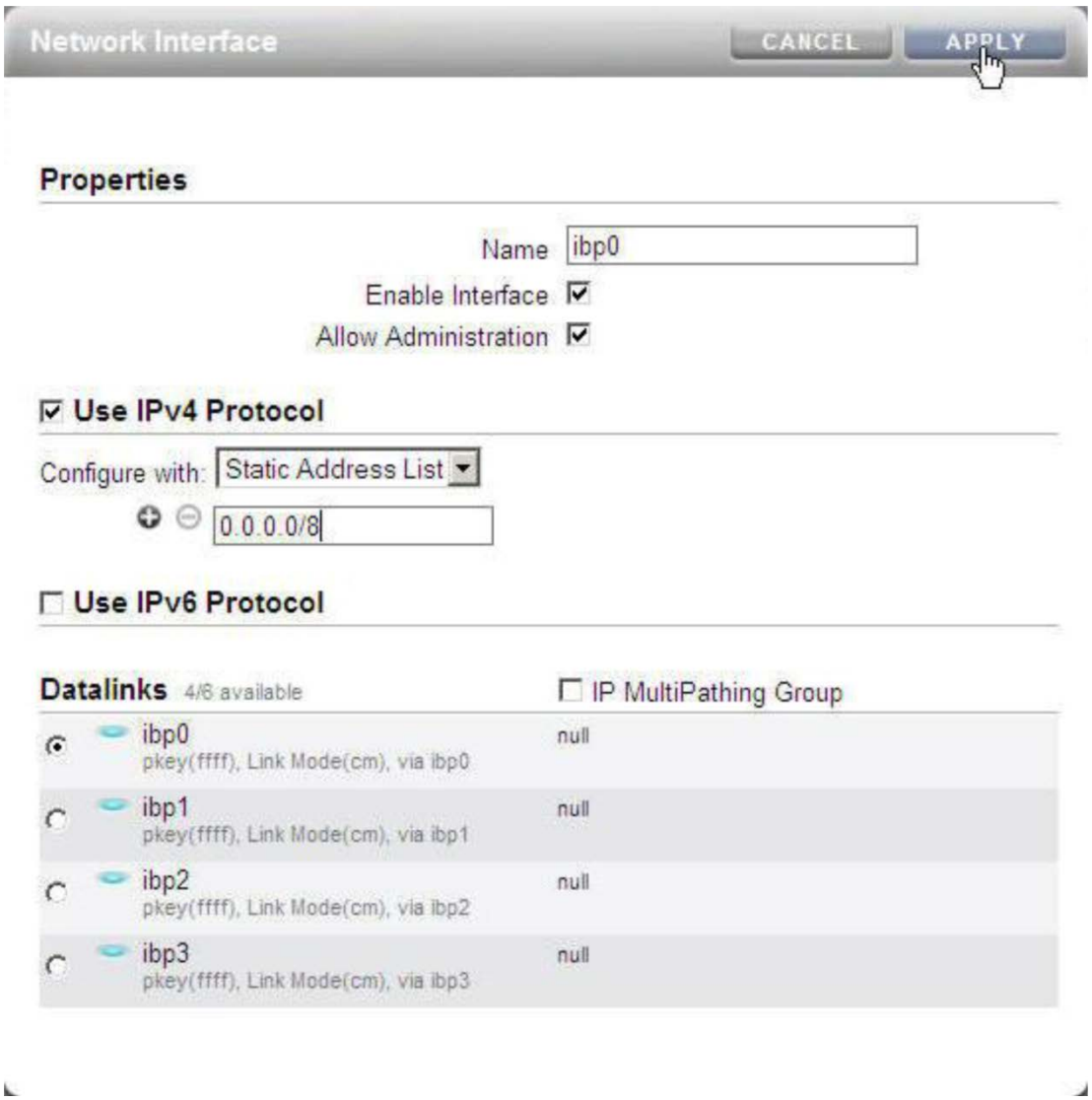


図11：InfiniBandネットワーク・インタフェースの構成

データ・リンクを管理するその他3つの新規インタフェースに対して、同じ手順を繰り返します。

フェーズ2：ネットワーク・インタフェースの構成 - IPネットワーク・マルチパス・グループ

「+」記号をクリックして、新規IPMPインタフェースを構成します。

- 分かりやすくするため、インタフェース名にはIPアドレスが解決される名前を使用します。
- 「**Use IPv4 Protocol**」チェック・ボックスを選択します。
- "Configure with"ドロップダウン・ボックスから「**Static Address List**」を選択します。

- Configure withボックスの下にあるアドレス・ボックスに、インタフェースが動作するIPアドレスとサブネット・マスク・ビットを入力します（例："192.168.41.189/21"）。
- 「**IP MultiPathing Group**」チェック・ボックスを選択します。
- IPMPグループに使用される2つのインタフェース（前述のステップで作成したもの）にチェックを付けます。このペアについては、前項を参照してください。
- チェックを付けた2つのインタフェースに対して、それぞれのドロップダウン・ボックスでActiveとStandbyが正しく設定されていることを確認します。

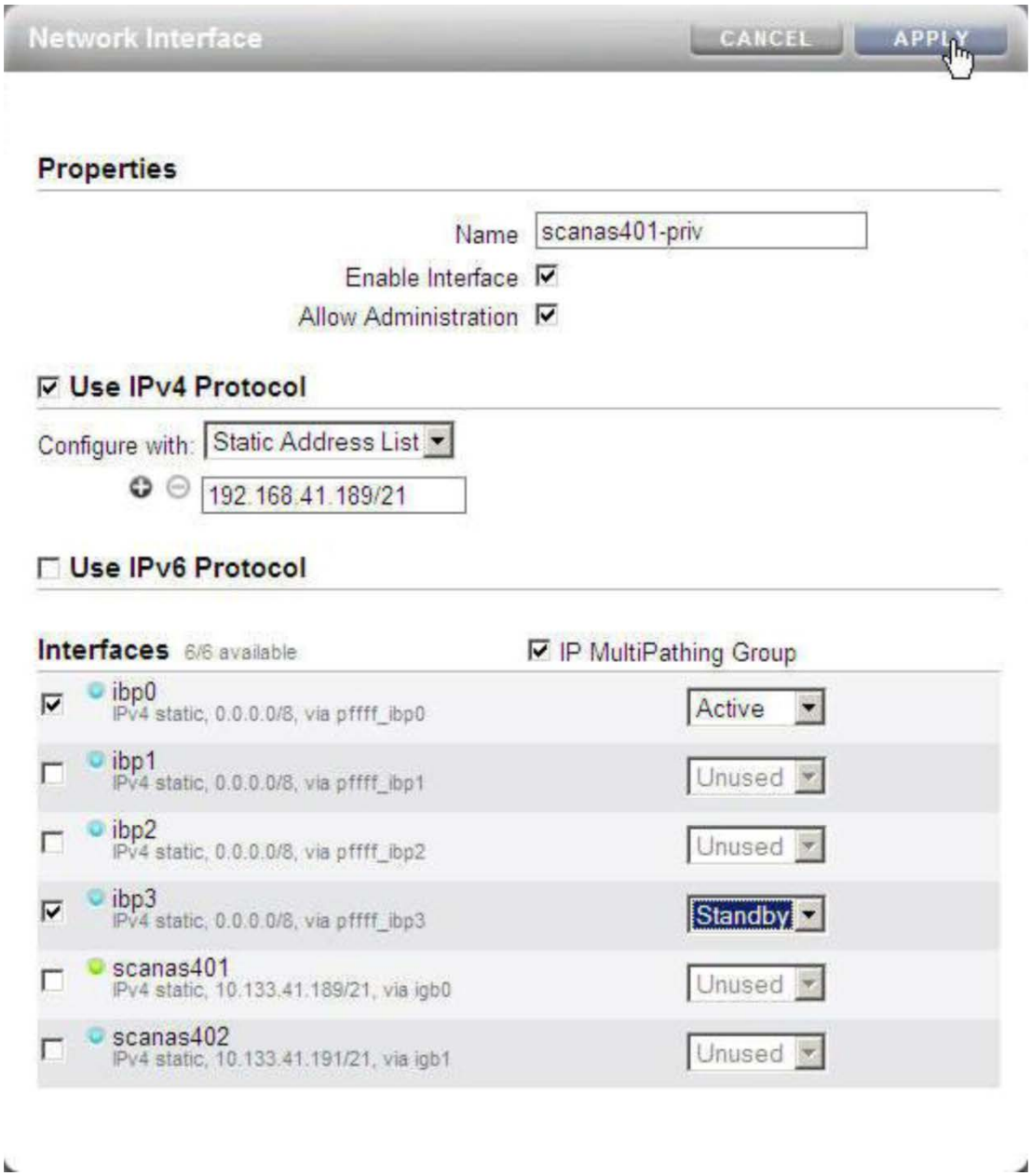


図12：InfiniBandのIPMPネットワーク・グループの構成

もう1つのInfiniBand IPMPグループに対して、同じ手順を繰り返します。

すべての変更を実施したら、メインのConfiguration→Networkページで必ずこの変更を適用します。変更を適用しない場合、ブラウザ・セッションを閉じたり画面から移動したりすると、変更が失われます。

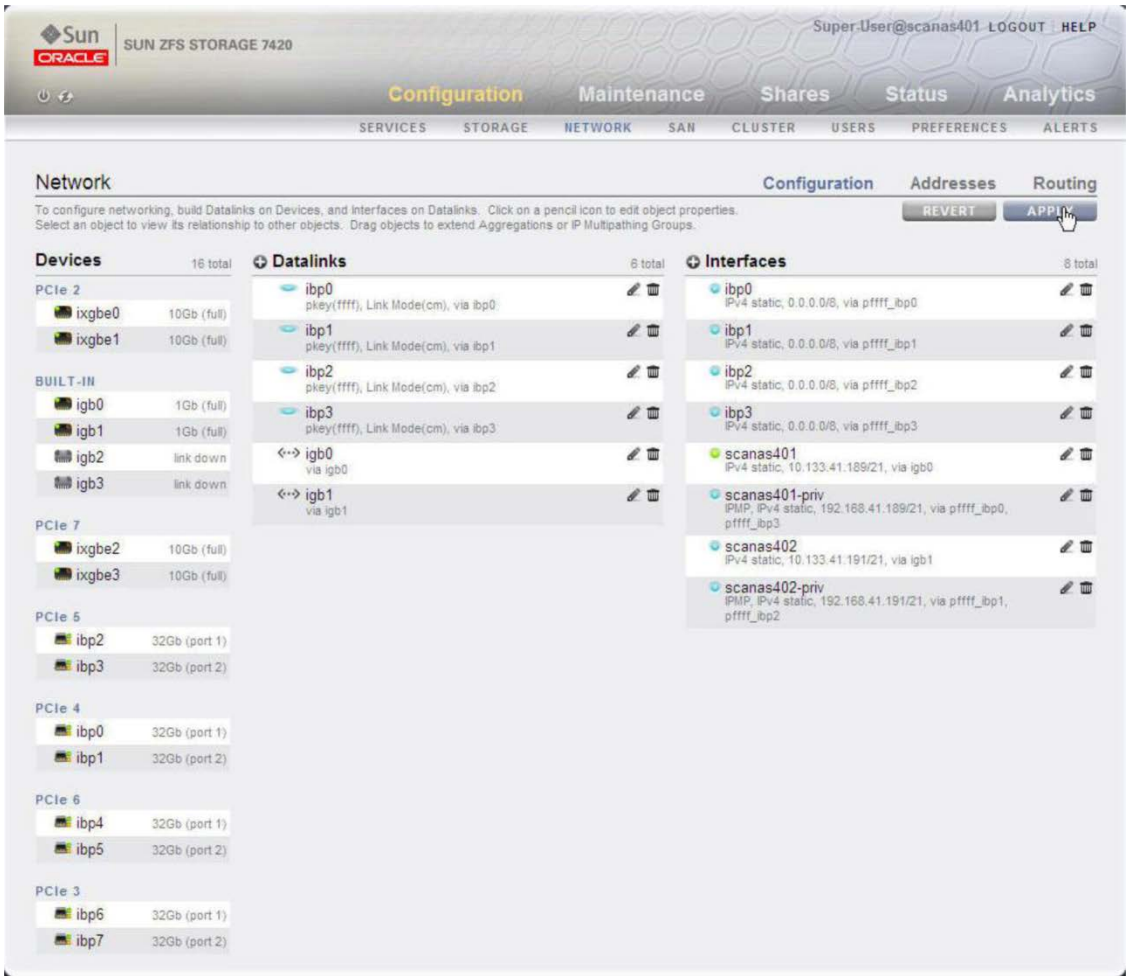


図13：InfiniBandネットワーク構成の保存

変更を適用すると、作成したネットワークIPアドレスに対してpingを実行できます。

10ギガビット・イーサネット接続のためのネットワーク構成

最善の可用性とパフォーマンスを提供するため、各Sun ZFS Storage Applianceコントローラに2つのデュアル・ポート10Gb光イーサネットを構成し、Sun ZFS Storage Applianceコントローラごとに合計4個のポートを提供することを推奨します。

これを実現するため、Sun ZFS Storage Applianceネットワーク構成はアクティブ/アクティブのIPMPグループを使用して構成します。また、パフォーマンス上の理由から、大きなMTUサイズ（ジャンボ・フレーム）に対しては10ギガビット・ネットワークを構成する必要があります。

2枚のデュアル・ポート10Gb光イーサネット・カードはおそらくSun ZFS Storage Applianceコントローラの PCIeスロット3および6にインストールされ、PCIeスロット3内のixgbe0とixgbe1およびPCIeスロット6内のixgbe2とixgbe3として識別されるでしょう。これらの2個のスロットは、縦に設置されたPCIeスロットの上部ポートと下部ポートとみなすことができます。Sun ZFS Storage Applianceのユーザー・インタフェースで、Configuration→Networkから、ixgbe0とixgbe3をアクティブ/アクティブのIPMPグループとして、ixgbe2とixgbe1を2番目のアクティブ/アクティブIPMPグループとしてバインドします。

10ギガビット・イーサネット接続のケーブル配線

10ギガビット・イーサネット接続は、顧客が供給する10ギガビット・ネットワーク・スイッチに接続されます。したがって、推奨事項はInfiniBandに対するケーブル配線よりも一般的なものになります。いずれかのスイッチに障害が発生した場合も可用性を提供できるようにするため、各IPMPグループから2つの異なる10ギガビット・ネットワーク・スイッチに対して2本のケーブルを接続すると理想的です。IPMPグループを使用するため、LACP（LAGまたは802.3ad）をネットワーク・スイッチ上に構成する必要はありません。リンク集約は、Oracle DBMSで提供されるダイレクトNFSを使用して実現されます。

大きなフレーム・パケットを送信できるようにするため、ネットワーク・スイッチにジャンボ・フレームを構成することが推奨されていますが、この場合、データベース・マシンとSun ZFS Storage Applianceの間で送信する必要のある確認応答の数が少なくなります。ジャンボ・フレームは、10ギガビット・ネットワーク・スイッチを担当するネットワーク管理者が構成します。

次の表に示すとおりにケーブルを接続します。

表3. 10ギガビット・ネットワークのケーブル配線マップ

Sun ZFS Storage Applianceヘッド	Sun ZFS Storage Applianceポート	10ギガビット・ネットワーク・スイッチ
ヘッドA	ixgbe0	スイッチ#1
ヘッドA	ixgbe1	スイッチ#2
ヘッドA	ixgbe2	スイッチ#1
ヘッドA	ixgbe3	スイッチ#2
ヘッドB	ixgbe0	スイッチ#1
ヘッドB	ixgbe1	スイッチ#2
ヘッドB	ixgbe2	スイッチ#1
ヘッドB	ixgbe3	スイッチ#2

ネットワークの構成 - Exadata Database Machine内のデータベース・ノード

本書に記載した速度を実現するには、Exadata Database Machine内の各データベース・ノードに対して、2つのネットワーク構成変更を実施する必要があります。

- 10GigEインタフェースでのTXキュー長の増加
- LinuxカーネルでのRXキュー長の増加

10GigEインタフェースでのTXキュー長の増加

TXキュー長はそれぞれの10GigEインタフェースに関連付けられた構成設定であり、Linuxのifconfigコマンドを使用して設定されます。このパラメータを、ボンディングされた10GigEインタフェースと基盤の10GigEスレーブ・インタフェースの両方に対して設定する必要があります。実行中のシステムでtxqueuelenを増やすには、rootユーザーとして次のコマンドを実行します。

```
# /sbin/ifconfig eth4 txqueuelen 10000

# /sbin/ifconfig eth5 txqueuelen 10000

# /sbin/ifconfig bondeth1 txqueuelen 10000
```

ノードまたはシステムの再起動中にこれらの構成を永続的に設定するには、データベース・ノードの/etc/rc.localファイルに同じコマンドを入力します。

LinuxカーネルでのRXキュー長の増加

RXキュー長はsysctlパラメータであり、sysctl -wコマンドまたは/etc/sysctl.conf構成ファイルを使用して設定されます。実行中のシステムでこのパラメータを設定するには、rootユーザーとして次のコマンドを実行します。

```
# /sbin/sysctl -w net.core.netdev_max_backlog=2000
```

ノードまたはシステムの再起動中にこのパラメータを永続的に設定するには、データベース・ノードの/etc/sysctl.confファイルにパラメータと値を入力します。

ネットワークの構成 - デバイス、データ・リンク、インタフェース

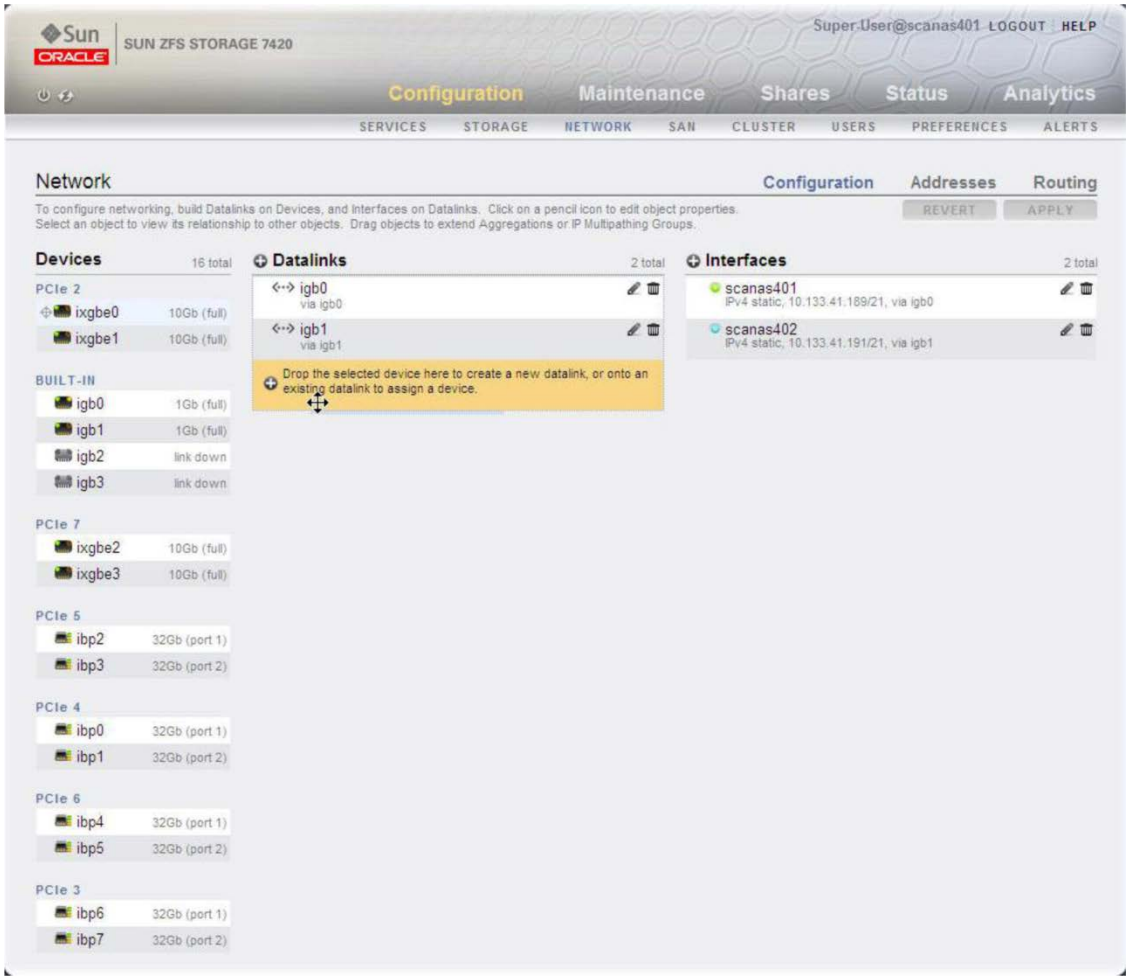
Sun ZFS Storage Applianceのネットワーク構成は、デバイス、データ・リンク、インタフェースという3つのレベルで構成されています。

- デバイスは物理ハードウェアを表します。
- データ・リンクは物理デバイスを管理します。
- インタフェースはデータ・リンク上にIPアドレスを構成します。

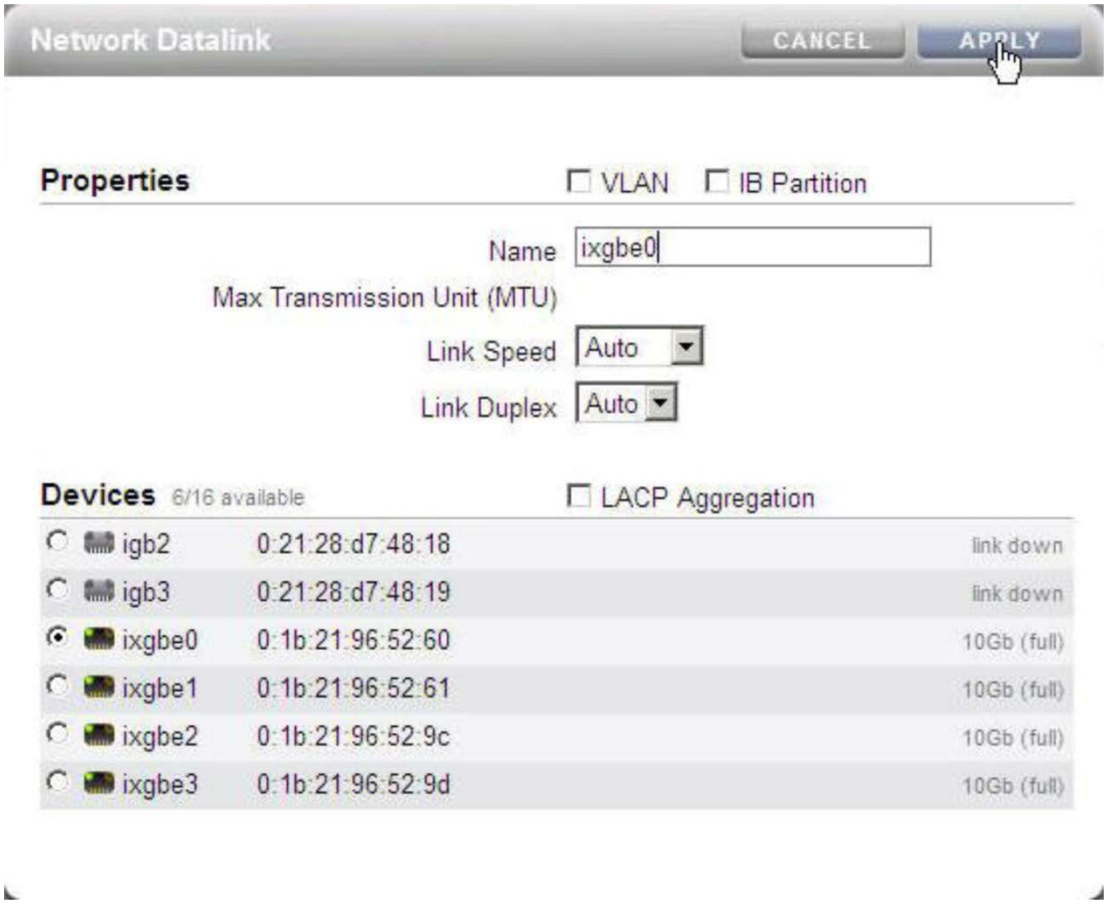
ネットワークのデータ・リンク構成

10ギガビット・ポートが使用するネットワーク・データ・リンクは、次のとおりに構成されています。

インタフェース列にデータ・リンク名をドラッグするか、または「+」記号をクリックして、新しいデータ・リンクを構成します。



- VLANとIB Partitionのチェック・ボックスは選択しないままにします。
- 分かりやすくするため、データ・リンクはデバイス名と同じ名前で作成します。
- Link SpeedとLink Duplexのドロップダウン・ボックスから望ましい値を選択します。
- 追加するデバイスに関連付けられたラジオ・ボタンをクリックします。



デバイスを管理するその他3つの新規データ・リンクに対して、同じ手順を繰り返します。

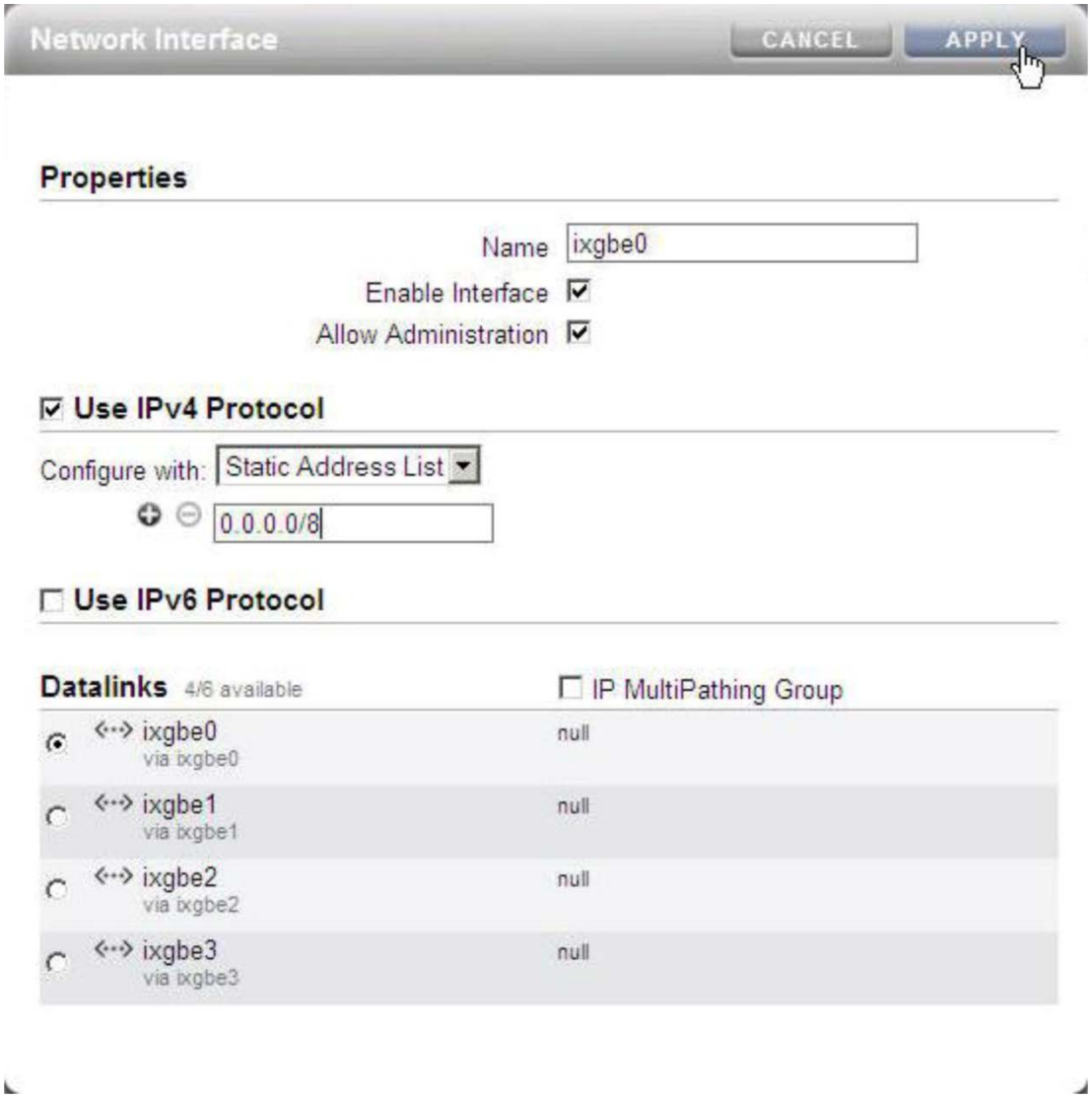
ネットワーク・インタフェースの構成

可用性のためにIPMPグループを使用するため、ネットワーク・インタフェースの構成には2つのフェーズがあります。

フェーズ1：ネットワーク・インタフェースの構成 - 静的アドレス

「+」記号をクリックして、新規インタフェースを構成します。

- 分かりやすくするため、インタフェースはデータ・リンク名と同じ名前で作成します。
- 「**Use IPv4 Protocol**」チェック・ボックスを選択します。
- "Configure with"ドロップダウン・ボックスから「**Static Address List**」を選択します。
- Configure withボックスの下にあるアドレス・ボックスに"0.0.0.0/8"と入力します。
- 追加するデータ・リンクに関連付けられたラジオ・ボタンをクリックします。



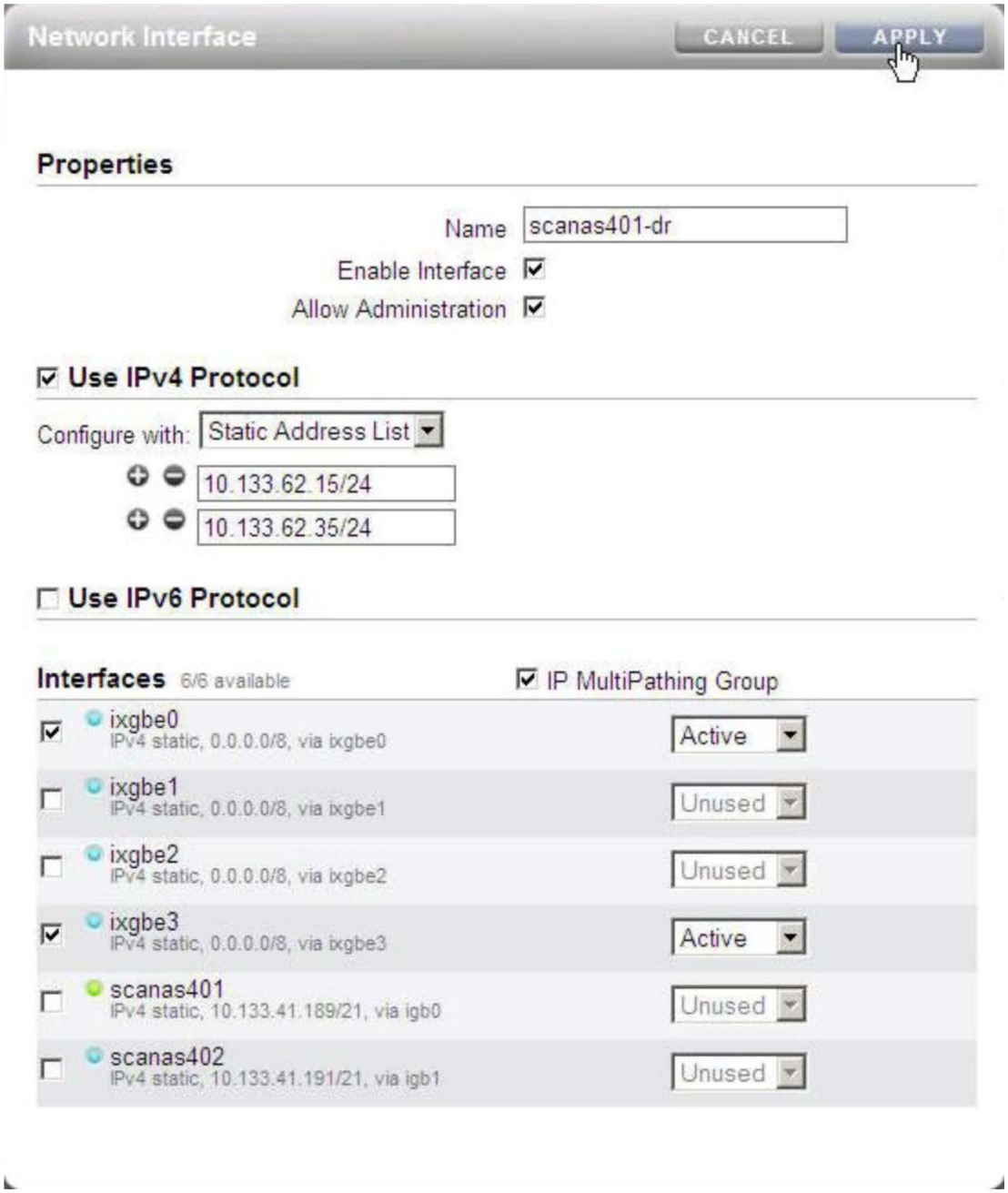
データ・リンクを管理するその他3つの新規インタフェースに対して、同じ手順を繰り返します。

フェーズ2：ネットワーク・インタフェースの構成 - IPMPグループ

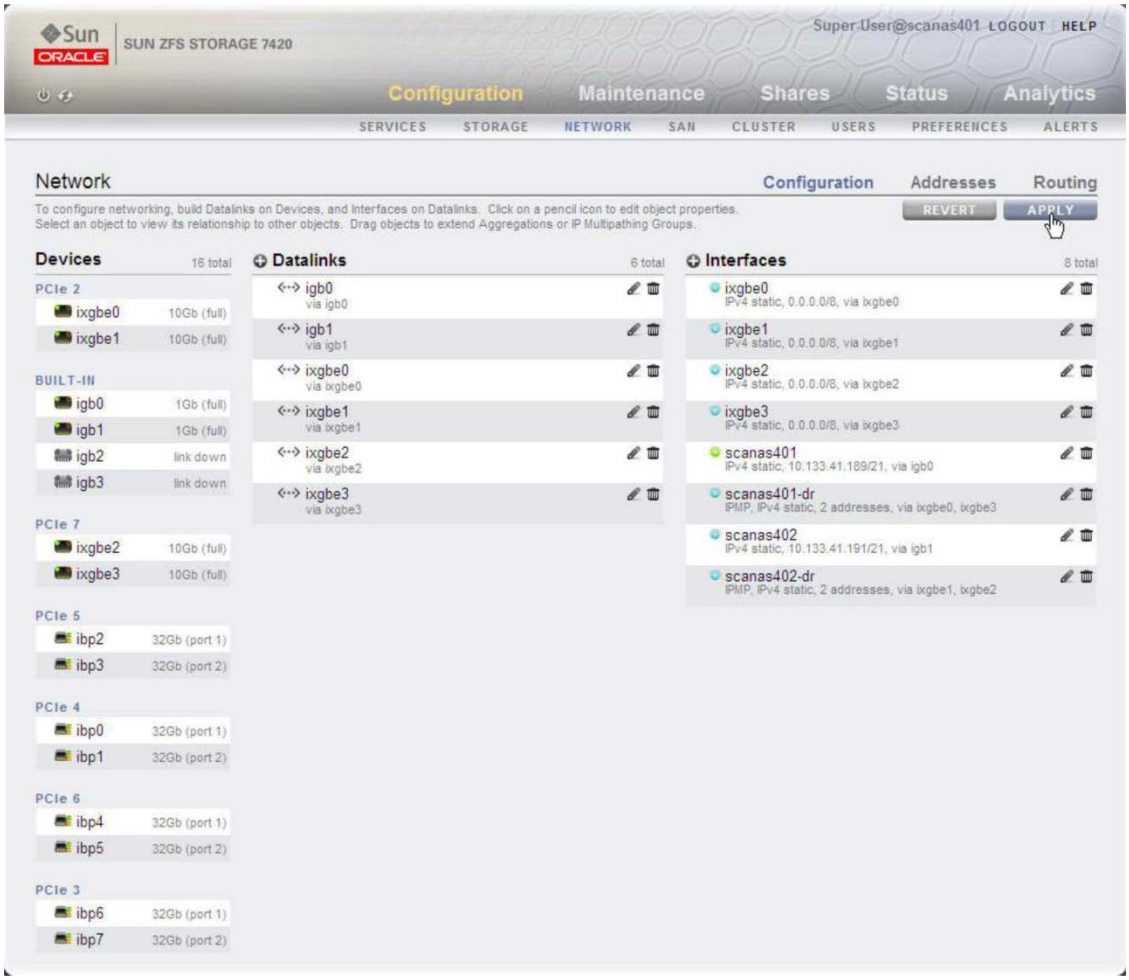
「+」記号をクリックして、新規IPMPインタフェースを構成します。

- 分かりやすくするため、インタフェース名にはIPアドレスが解決される最初の名前を使用します。
- 「**Use IPv4 Protocol**」チェック・ボックスを選択します。
- "Configure with"ドロップダウン・ボックスから「**Static Address List**」を選択します。
- Configure withボックスの下にあるアドレス・ボックスに、インタフェースが動作するIPアドレスとサブネット・マスク・ビットを入力します（例："192.168.41.129/21"）。

- 「+」記号をクリックして2番目のIPアドレスを追加し、2番目のインタフェースが動作するIPアドレスとサブネット・マスク・ビットを入力します（例："192.168.41.130/21"）。
- 「**IP MultiPathing Group**」チェック・ボックスを選択します。
- IPMPグループに使用される2つのインタフェース（前述のステップで作成したもの）にチェックを付けます。このペアについては、前項を参照してください。
- チェックを付けた2つのインタフェースに対して、両方とも、ドロップダウン・ボックスからActiveが設定されていることを確認します。



すべての変更を実施したら、メインのConfiguration→Networkページにこの変更を適用します。変更を適用しない場合、ブラウザ・セッションを閉じたり画面から移動したりすると、変更が失われます。



変更を適用すると、各インタフェース用に作成した両方のネットワークIPアドレスに対してpingを実行できます。

ネットワーク・ルーティングの構成

「**routing**」タブを選択してアダプティブ・ルーティングを有効化し、サブネットのローカルIPアドレスに対するSun ZFS Storage Applianceからの読取り要求が、要求を受け取った物理インタフェースから処理されるようにします。Exadata Database MachineとSun ZFS Storage Applianceが異なるサブネット上にある場合、Sun ZFS Storage Applianceの10Gbイーサネット・インタフェースからExadataの10Gbイーサネット・インタフェースを含むサブネットへのネットワーク・ルートを追加します。この作業は、クラスタ構成に含まれる両方のSun ZFS Storage Applianceヘッドに対して実行します。

高度なルーティング、トラフィック分離、およびサービス品質（QoS）フィルタは、システム・スループットに影響を与える場合がありますが、このようなチューニングは本書の範囲外です。詳しくは、ネットワーク管理者にお問い合わせください。

ジャンボ・フレームの構成

Sun ZFS Storage Applianceのユーザー・インタフェースでジャンボ・フレームを構成することはできませんが、Sun ZFS Storage ApplianceのCLIを使用することで構成できます。sshプロトコルを使用してSun ZFS Storage Applianceにログインし、次のとおりにMTUサイズを構成します。

- "configuration net datalinks select ixgbe0 show"コマンドを使用して、ixgbe0インタフェースに対する現在の定義を表示します。"configuration net datalinks show"を実行すると、4つの10ギガビット・データ・リンクが表示されます。
- "configuration net datalinks select ixgbe0 set mtu=9000"コマンドを使用して、Maximum Transmission Unitを9000バイトに設定します。特定のネットワークに対して使用すべきMTUサイズについては、ネットワーク・エンジニアにお問い合わせください。先ほど構成したその他3つのデータ・リンクについて、この手順を繰り返します。

```
scanas401:> configuration net datalinks select ixgbe0 show
Properties:
                        class = device
                        label = ixgbe0
                          mac = 0:1b:21:96:52:60
                        links = ixgbe0
                          mtu = 1500
                        speed = auto
                       duplex = auto

scanas401:> configuration net datalinks select ixgbe0 set mtu=9000
                          mtu = 9000
scanas401:> configuration net datalinks select ixgbe1 set mtu=9000
                          mtu = 9000
scanas401:> configuration net datalinks select ixgbe2 set mtu=9000
                          mtu = 9000
scanas401:> configuration net datalinks select ixgbe3 set mtu=9000
                          mtu = 9000
scanas401:> configuration net datalinks select ixgbe0 show
Properties:
                        class = device
                        label = ixgbe0
                          mac = 0:1b:21:96:52:60
                        links = ixgbe0
                          mtu = 9000
                        speed = auto
                       duplex = auto

scanas401:>
```

図14：10ギガビット・データ・リンクに対するMTUサイズの構成

ネットワーク・クラスタ・リソースの構成

本書の目的上、ここでは10ギガビット・ネットワーク・インタフェースに対する構成手順を説明しますが、手順はInfiniBandインタフェースの場合も同じです。

Configuration→Clusterページへ移動すると、先ほど構成した6つのネットワーク・インタフェースがあらかじめscanas401ノードに割り当てられていることが分かります。また、それぞれのシステム上で管理ネットワーク・インタフェースが実行されています。

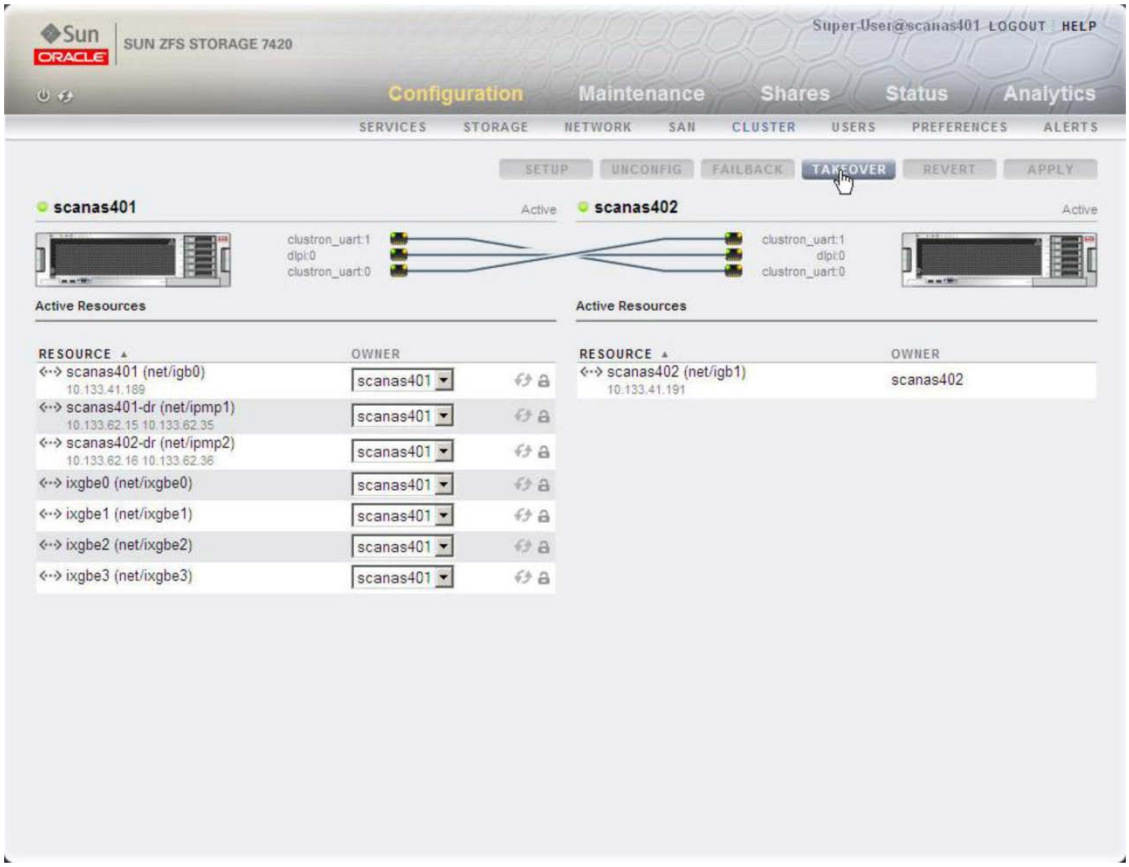


図15：クラスタ構成の初期画面

最初のステップでは、クラスタ・テイクオーバーを発行します。テイクオーバーはクラスタ内の"もう一方"のシステムを再起動し、そのシステムが所有するリソースを強制的にローカル・システムで管理します。もう一方のシステムが再起動中であるため、この時点で、最終的にこの"もう一方"のシステムで実行される3つのリソースを適用できます。「**Apply**」をクリックすると、リソースの"Apply"または"Failback"を選ぶ確認ダイアログが表示されます。"もう一方"のシステムが実行中に戻っている場合は「**Failback**」を選べますが、それ以外の場合は「**Apply**」をクリックします。

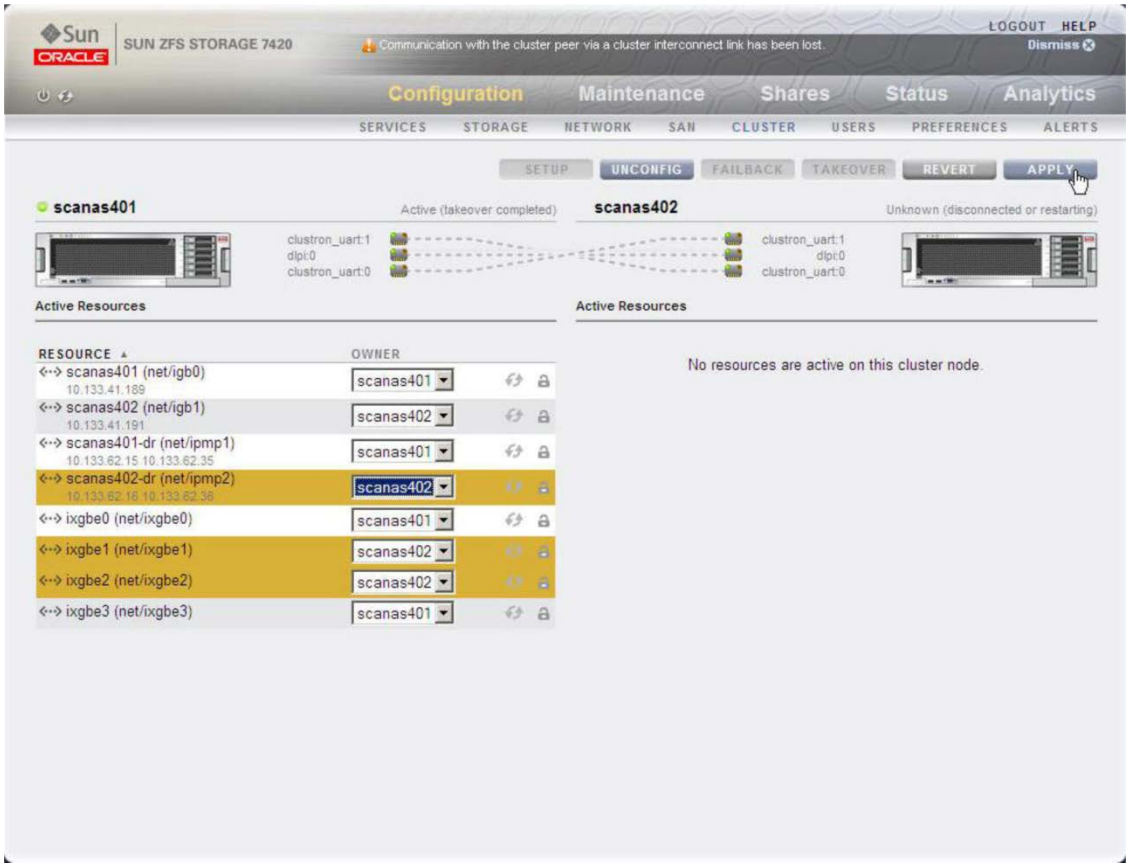


図16："もう一方"のシステムに対する10ギガビット・リソースの割当て

"もう一方"のシステムが再起動されたら、Sun ZFS Storage ApplianceのUIでは、"もう一方"のシステムのステータスとして"Ready (waiting for failback)"が表示されます。すべてのシステム・リソースはローカル・システム上で実行されていますが、「**failback**」をクリックすると"もう一方"のシステムにリソースが戻されます。

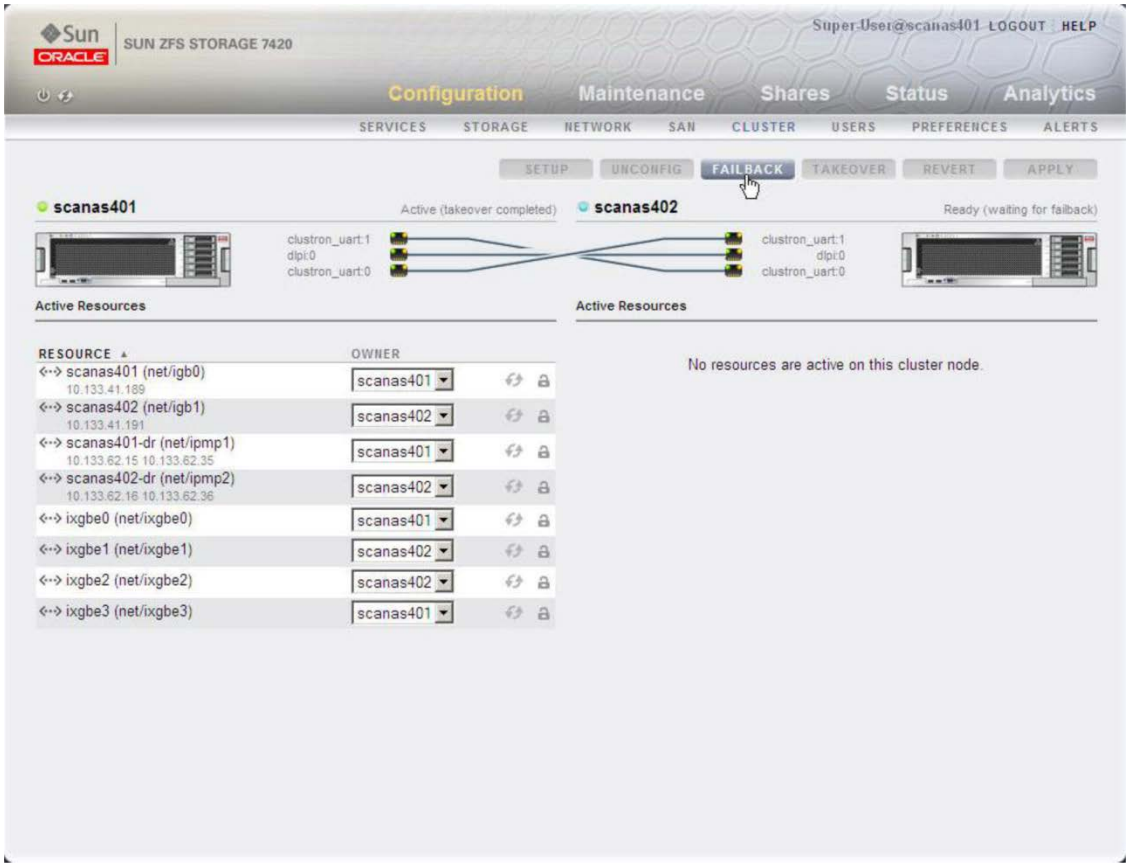


図17：所有ノードに対するクラスタ・リソースのフェイルバック

所有ノードに対するリソースのフェイルバックが成功したら、Sun ZFS Storage ApplianceのUIには正しく分散されたリソースが表示されます。

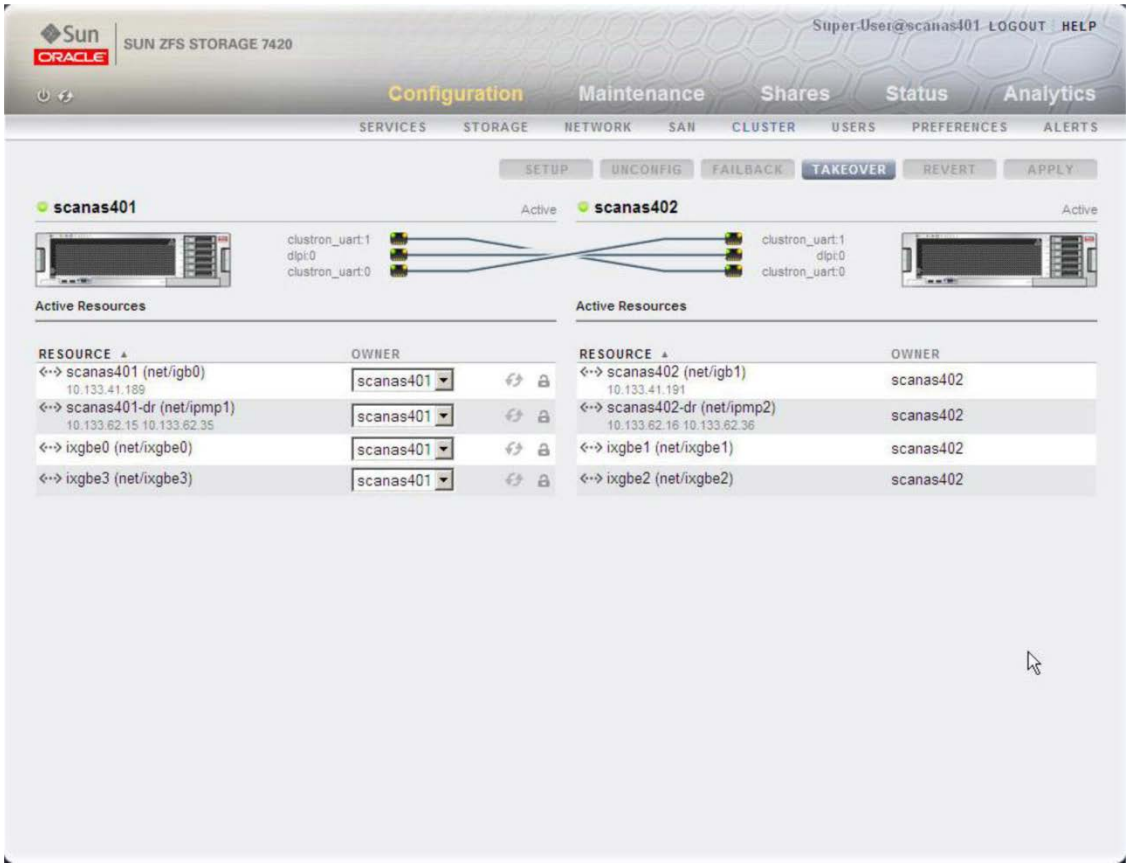
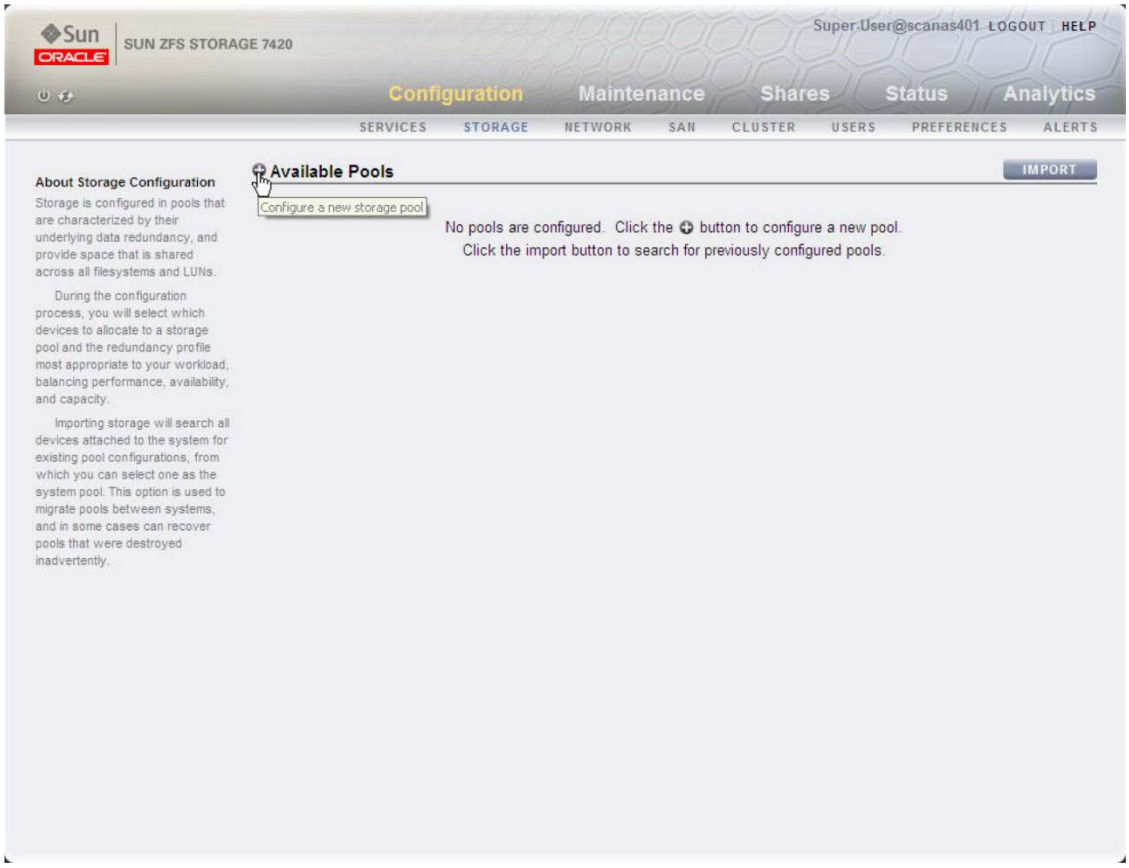


図18：クラスタ・ネットワーク・リソースの構成後

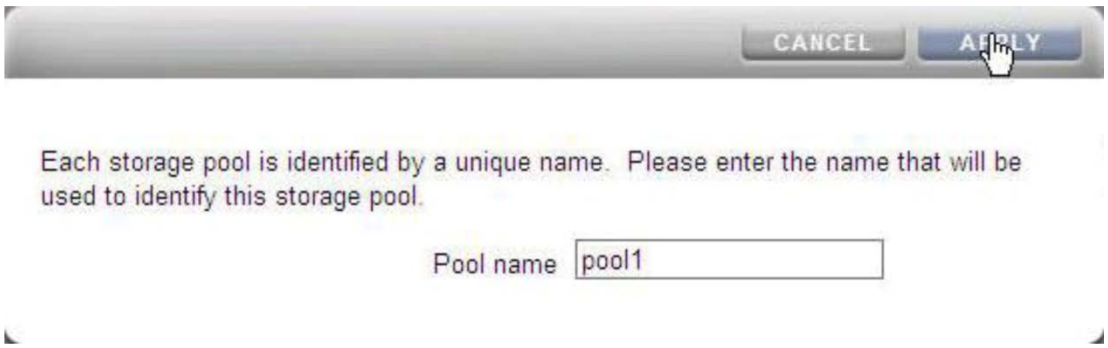
この時点でExadata Database Machineにログオンし、これらのリソースのIPアドレスへ"ping"を実行できるようになります。10ギガビット・ネットワークの場合は4つのインタフェースすべてに対して、InfiniBandネットワークの場合は両方のインタフェースに対して、pingを実行できるようになります。

ストレージ構成

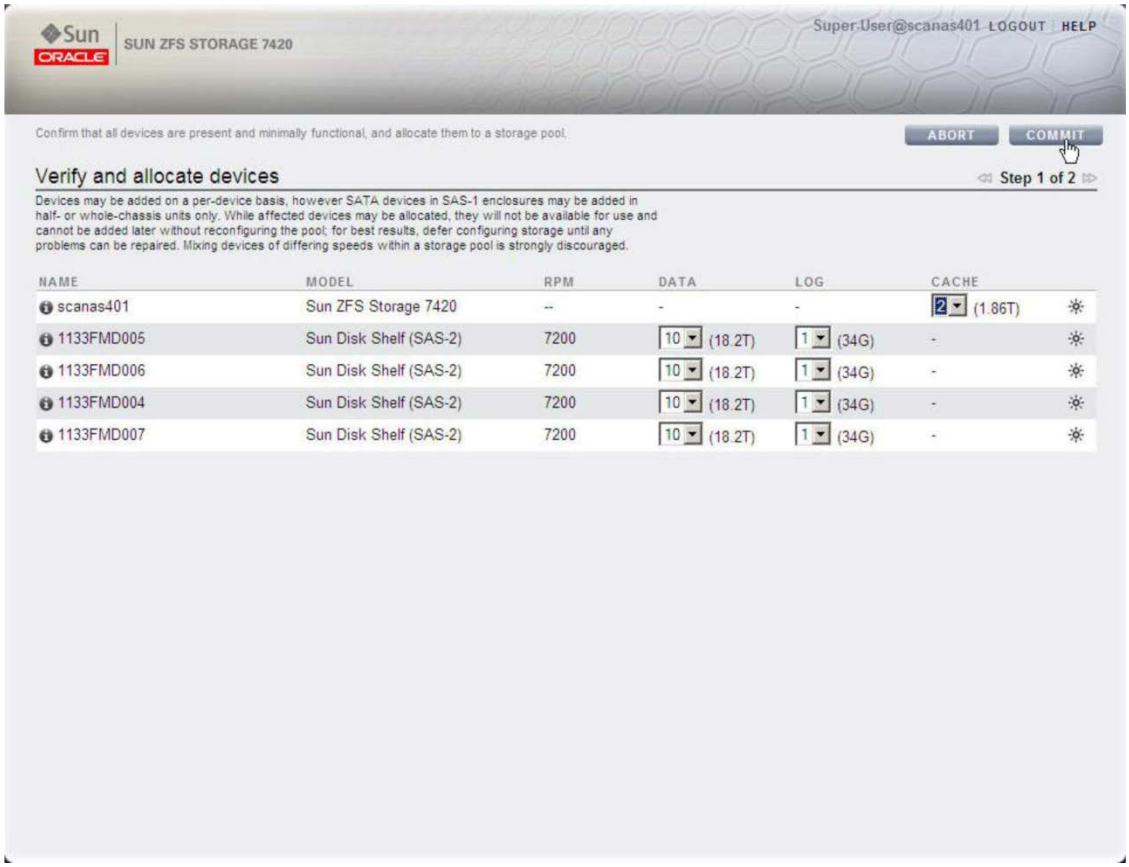
最初のシステムにログインしてConfiguration→Storageページへ移動したら、Available Poolsの横にある「+」記号をクリックします。



構成スクリプトでは、pool1およびpool2というプールを検出することになっています。「+」記号をクリックしたら、プール名としてポップアップ・ウィンドウにpool1と入力します。

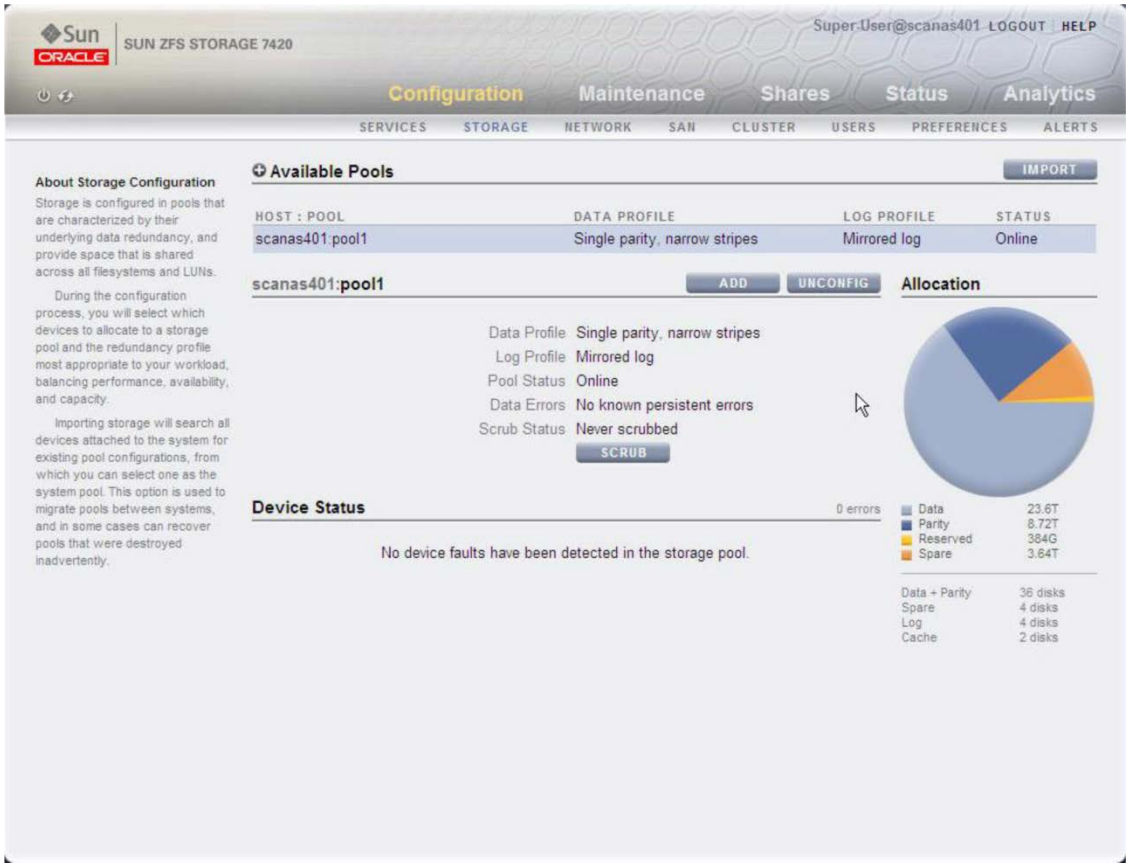


2ページのダイアログが表示されます。最初のページでは、最初のプールにマッピングされるデータ・デバイス、ログ・デバイス、キャッシュ・デバイスの数を選択します。2つの同等なプールが作成されたら、リソース数を均等に割り当て、「**commit**」をクリックします。

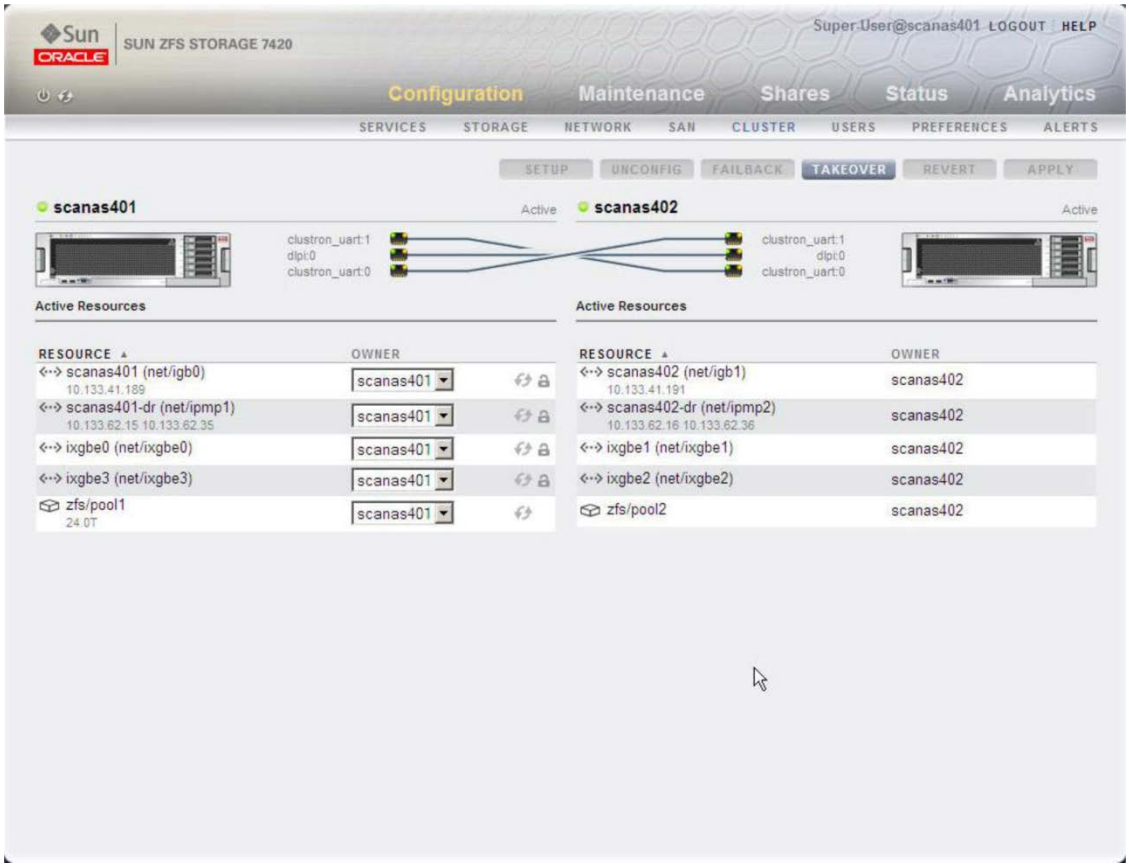


2番目のページにはData、Log、Cacheという3つのプロファイル・タブがあり、これらに入力する必要があります。このホワイト・ペーパーでは単純なOracleデータベースのバックアップおよびリカバリに焦点を合わせており、データのスナップショットとクローンを使用したテストの目的でSun ZFS Storage Applianceを使用してはいません。このため、「**Single Parity, narrow strips**」の選択を推奨します。こうすることで、適度な量の領域を維持しながら、良いパフォーマンスを提供します。

ログ・デバイスが表示されており、NSPFを提供している方の「**Mirrored Logs**」を選択できる場合はこれを選択します。それ以外の場合は、もう一方の「**Mirrored Logs**」を選択します。Cacheタブでは何も選択する必要はありません。適切なプロファイルを選択したら、「**commit**」ボタンをクリックします。ストレージ・プールが作成され、"Available Pools" 画面に使用可能な領域が表示されます。



もう一方のシステムにログインし、同じ手順を繰り返してpool2という名前の2番目のプールを作成します。作業が完了したら、Configuration→Clusterへ移動します。次のような出力が表示され、各システムがSun ZFS Storage Applianceプールと、先ほど構成したそれぞれのネットワーク・インタフェースの所有者になっていることが分かります。



サンプルの/etc/oranfstabファイル

次に、2つのoranfstabファイルを示します。1つは10ギガビット・イーサネット用であり、もう1つはInfiniBand用です。データベース名には"orcl"が使用されています。

これらのサンプル・ファイルは、クラスタ内の1つの特定ノードを対象としています。サンプル内で使用されたローカルIPアドレスは、データが送信されるデータベース・ノード上のインタフェースに関連付けられたIPアドレスを表しています。クラスタ内の各ノードで、ファイルに正しいローカルIPアドレスが指定されていることを確認してください。

10ギガビット向けの/etc/oranfstabファイル

server: scanas401

```
local: 10.133.62.27 path: 10.133.62.15
local: 10.133.62.27 path: 10.133.62.35
dontroute
export: /export/orcl/backup1 mount: /zfssa/orcl/backup1
export: /export/orcl/backup2 mount: /zfssa/orcl/backup2
export: /export/orcl/backup3 mount: /zfssa/orcl/backup3
```

```
export: /export/orcl/backup4 mount: /zfssa/orcl/backup4
export: /export/orcl/backup5 mount: /zfssa/orcl/backup5
export: /export/orcl/backup6 mount: /zfssa/orcl/backup6
export: /export/orcl/backup7 mount: /zfssa/orcl/backup7
export: /export/orcl/backup8 mount: /zfssa/orcl/backup8
server: scanas402
local: 10.133.62.27 path: 10.133.62.16
local: 10.133.62.27 path: 10.133.62.36
dontroute
export: /export/orcl/backup9 mount: /zfssa/orcl/backup9
export: /export/orcl/backup10 mount: /zfssa/orcl/backup10
export: /export/orcl/backup11 mount: /zfssa/orcl/backup11
export: /export/orcl/backup12 mount: /zfssa/orcl/backup12
export: /export/orcl/backup13 mount: /zfssa/orcl/backup13
export: /export/orcl/backup14 mount: /zfssa/orcl/backup14
export: /export/orcl/backup15 mount: /zfssa/orcl/backup15
export: /export/orcl/backup16 mount: /zfssa/orcl/backup16
```

10ギガビットのネットワーク構成では、Sun ZFS Storage Applianceがアクティブ/アクティブのIPMPグループを使用しているため、local/pathのパラメータ・ペアが2つ含まれる点に注意してください。

InfiniBand向けの/etc/oranfstabファイル（Exadata Database Machine X2-8用）

```
server:scanas401
local: 192.168.20.208 path: 192.168.20.189
local: 192.168.20.209 path: 192.168.20.189
local: 192.168.20.210 path: 192.168.20.189
local: 192.168.20.211 path: 192.168.20.189
dontroute:
export: /export/orcl/backup1 mount: /zfssa/orcl/backup1
export: /export/orcl/backup2 mount: /zfssa/orcl/backup2
export: /export/orcl/backup3 mount: /zfssa/orcl/backup3
export: /export/orcl/backup4 mount: /zfssa/orcl/backup4
```

```
export: /export/orcl/backup5 mount: /zfssa/orcl/backup5
export: /export/orcl/backup6 mount: /zfssa/orcl/backup6
export: /export/orcl/backup7 mount: /zfssa/orcl/backup7
export: /export/orcl/backup8 mount: /zfssa/orcl/backup8
server: scanas402
local: 192.168.20.208 path: 192.168.20.191
local: 192.168.20.209 path: 192.168.20.191
local: 192.168.20.210 path: 192.168.20.191
local: 192.168.20.211 path: 192.168.20.191
dontroute:
export: /export/orcl/backup9 mount: /zfssa/orcl/backup9
export: /export/orcl/backup10 mount: /zfssa/orcl/backup10
export: /export/orcl/backup11 mount: /zfssa/orcl/backup11
export: /export/orcl/backup12 mount: /zfssa/orcl/backup12
export: /export/orcl/backup13 mount: /zfssa/orcl/backup13
export: /export/orcl/backup14 mount: /zfssa/orcl/backup14
export: /export/orcl/backup15 mount: /zfssa/orcl/backup15
export: /export/orcl/backup16 mount: /zfssa/orcl/backup16
```

Exadata Database Machine X2-8ではそれぞれのデータベース・ノード上に4個のInfiniBandポートがあるため、local/pathのパラメータ・ペアが4つ含まれています。

ネットワークとスループット使用率のテスト

システムのセットアップ、Sun ZFS Storage Applianceの構成、Exadata Backup Configuration Utilityの実行、データベースでのダイレクトNFSの有効化の各手順を完了したら、テスト・バックアップを実行して、ネットワーク構成とスループット使用率を検証できます。

サンプルのレベル0バックアップ・スクリプト（付録B）のbackup句を変更し、500GBから1TBまでのデータベース・オブジェクトで構成される比較的小さいデータベース・サブセットに対してバックアップを限定します。

Oracle RMANのバックアップを実行する前に、次の監視を開始します。

- 各データベース・ノードで次のコマンドを実行します。

  ```
  sar -n DEV 1 1000
  ```

- Sun ZFS Storage Applianceの各コントローラ・ヘッドでAnalytics画面を開き、次の項目を監視します。

- ネットワーク：インタフェースごとに分類された1秒あたりのインタフェース・バイト

Oracle RMANのバックアップ・スクリプトを実行したら、oranfstabファイルのlocal/pathパラメータ・ペアで定義された、使用予定のネットワーク・インタフェースすべてが使用されていることをチェックする必要があります。

また、Oracle RMANの"restore validate"コマンドを実行して、Sun ZFS Storage Applianceからの読取り中にもすべてのネットワーク・インタフェースが引き続き使用されることを確認します。

付録B：データベース・バックアップのサンプル・スクリプト

レベル0の週次データベース・バックアップ

レベル0のバックアップ・セットに対してOracle Exadata Backup Configuration Utilityで生成されるスクリプトは、次の例のようになります。

```
run
{
sql 'alter system set "_backup_disk_bufcnt"=64 scope=memory';
sql 'alter system set "_backup_disk_bufsz"=1048576 scope=memory';

allocate channel ch01 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup1' format
'/zfssa/orcl/backup1/%U';
allocate channel ch02 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup2' format
'/zfssa/orcl/backup9/%U';
allocate channel ch03 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup3' format
'/zfssa/orcl/backup2/%U';
allocate channel ch04 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup4' format
'/zfssa/orcl/backup10/%U';
allocate channel ch05 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup5' format
'/zfssa/orcl/backup3/%U';
allocate channel ch06 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup6' format
'/zfssa/orcl/backup11/%U';
```

```
allocate channel ch07 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup7' format
'/zfssa/orcl/backup4/%U';

allocate channel ch08 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup8' format
'/zfssa/orcl/backup12/%U';

allocate channel ch09 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup1' format
'/zfssa/orcl/backup5/%U';

allocate channel ch10 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup2' format
'/zfssa/orcl/backup13/%U';

allocate channel ch11 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup3' format
'/zfssa/orcl/backup6/%U';

allocate channel ch12 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup4' format
'/zfssa/orcl/backup14/%U';

allocate channel ch13 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup5' format
'/zfssa/orcl/backup7/%U';

allocate channel ch14 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup6' format
'/zfssa/orcl/backup15/%U';

allocate channel ch15 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup7' format
'/zfssa/orcl/backup8/%U';

allocate channel ch16 device type disk connect
'system/welcome1@dm01-scan/orcl_bkup8' format
'/zfssa/orcl/backup16/%U';

configure snapshot controlfile name to
'/zfssa/orcl/backup1/snapcf_orcl.f';

backup as backupset incremental level 0 section size 32g database
tag 'FULLBACKUPSET_L0' plus archivelog tag 'FULLBACKUPSET_L0';

}
```

スクリプトは最初に、Sun ZFS Storage Applianceへのデータ書込みを最適化する2つのinit.oraパラメータを設定します。次に、16個のOracle RMANチャネルを割り当てて、各チャネルがSun ZFS Storage Appliance上の特定のシェアにバックアップ・セッ

トを書き込むようにします。また、Oracle RMANによって自動的に作成される制御ファイルのスナップショットを、Sun ZFS Storage Appliance上の1番目のシェアに割り当てます。

その後で、メインのbackupコマンドが実行されます。重要な部分は、レベル0の増分バックアップ・セットを明示的に作成している個所と、すべてのリソースを有効活用するために、32GBを超えるデータファイルを32GBごとに分割している点です。このスクリプトはまた、システム上にあるアーカイブ・ログのバックアップも行います。

レベル1の日次データベース・バックアップ

レベル1のバックアップ・セットに対してOracle Exadata Backup Configuration Utilityで生成されるスクリプトは、レベル0のバックアップと非常によく似ています。唯一の違いは、実行されるメインのbackupコマンドにあります。

```
RMAN> backup as backupset incremental level 1 database tag 'FULLBACKUPSET_L1' plus
archivelog tag 'FULLBACKUPSET_L1';
```

Sun ZFS Storage Appliance と Oracle Exadata Database Machine を使用したバックアップおよびリカバリのパフォーマンスとベスト・プラクティス
2012 年 4 月
著者：Andrew Babb
貢献者：Juan Loaiza、Lawrnece To、Jeff Wright

文書作成者：Virginia Beecher

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

海外からのお問い合わせ窓口：
電話：+1.650.506.7000
ファクシミリ：+1.650.506.7200

www.oracle.com

Oracle is committed to developing practices and products that help protect the environment

Hardware and Software, Engineered to Work Together