ORACLE
ZFS STORAGE APPLIANCE

Oracle ZFS Storage ApplianceによるExadata Database Machineの保護

構成のベスト・プラクティス

Oracle ホワイト・ペーパー | 2015 年 7 月

ORACLE

目次

概要 3
はじめに 8
データ保護戦略の選択 9
従来型のバックアップ戦略 9
増分更新バックアップ戦略 12
Oracle ZFS Storage Appliance を使用したベスト・プラクティス 13
Oracle ZFS Storage Appliance の構成 15
コントローラの選択 15
適切なディスク・シェルフの選択 16
ストレージ・プロファイルの選択 17
ストレージ・プールの構成 19
書込みフラッシュ・アクセラレータおよび読取り最適化フラッシュの使用 19
InfiniBand または 10GbE の選択 20
Oracle Direct NFS の選択 21
Oracle Intelligent Storage Protocol 21
IP マルチパスの構成 22
Oracle Exadata Database Machine の構成 24
NFSv3 または NFSv4 の選択 24
Exadata での NFS の有効化 24
NFS マウント・オプション 25
Oracle Direct NFS の構成 25
oranfstab ファイルの作成 26
Recovery Manager のバックアップ・サービスの構成 29
InfiniBand のベスト・プラクティス 29
バックアップ用のデータベースの準備 30
従来型の Recovery Manager のバックアップ向けのベスト・プラクティス 32
ネットワーク共有の構成 32
レコード・サイズ 32
同期書込みバイアス 33
読取りキャッシュ 33
データ圧縮 34
Recovery Manager の構成 34
バックアップ形式 34

チャネルの最適化 35
バッファのチューニング 36
セクションのサイズ 36
FILESPERSET 36
run ブロックの例 37
増分更新バックアップのベスト・プラクティス 38
Oracle ZFS Storage Appliance の構成 38
Recovery Manager 環境の構成 39
Oracle ZFS Storage ZS4-4 における Recovery Manager のパフォーマンス・サイジング ... 42
Oracle ZFS Storage ZS3-2 における Recovery Manager のパフォーマンス・サイジング ... 45
結論 48

概要

Oracle Exadata Database Machine 上のミッション・クリティカルなデータを保護することは、最優先事項です。Oracle ZFS Storage Appliance は、優れたパフォーマンス、強化された信頼性、卓越したネットワーク帯域幅、強力な機能、簡素化された管理、およびコスト効率の高い構成を実現しているため、このタスクに最適です。Oracle ZFS Storage Appliance の機能は、以下のとおりです。

卓越したネットワーク帯域幅 - Oracle ZFS Storage Appliance は、InfiniBand または 10GbE 周辺で構築可能な、スケーラビリティの高いアーキテクチャを備えており、Exadata Database Machine に接続する際に必要なネットワーク・パフォーマンスおよび冗長性を実現しています。

ZFS で強化されたディスクの信頼性 - copy-on-write、メタデータのチェックサム演算、バックグラウンドでのデータ・スクラビングなどの ZFS 機能が強化されたことにより、データの整合性を保証し、発見されにくいデータ破損の場合でも検出して、手遅れになる前にエラーを修正します。

強力な機能 - ストレージ分析を実施することにより、迅速かつ効率的にパフォーマンスのボトルネックを識別できます。Oracle Intelligent Storage Protocol により、一意のデータベース認識ストレージが使用可能になるため、管理が簡素化され、パフォーマンスが最適化されます。レプリケーション、ZFS のスナップ/クローン、暗号化などのテクノロジーは、発生する可能性のある、データ保護や開発/テスト・プロビジョニングに関する課題に対して、ソリューションを提供します。これらは、Oracle ZFS Storage Appliance で使用できる、エンタープライズ・クラスの機能のほんの一部に過ぎません。

シンプルな管理 - 使いやすい Web 管理インタフェースを提供して、Oracle Enterprise Manager 内で統合することにより、トレーニングの費用を削減し、管理オーバーヘッドを減らします。Oracle ZFS Storage Appliance の革新的でスケーラビリティの高いストレージ・プール、および高速なファイル・システム・プロビジョニングにより、ストレージの管理が極めて簡単になります。

Oracle Optimized Compression - Hybrid Columnar Compression (HCC) が使用できるのは、オラクルのストレージ上のみです。Oracle ZFS Storage Appliance は、Oracle Database と統合して、特定のデータ使用やワークロード・パターンに対して最適化された、幅広い圧縮オプションを提供します。

優れたパフォーマンス - Oracle ZFS Storage Appliance は、最大 42TB/時のデータベース・バックアップ速度、および最大 55TB/時のリストア速度を実現しています。また、SPC-2 および SPECsfs 業界標準ベンチマークの世界記録を樹立しています。優れたハードウェアと緊密な

統合により、競合のストレージ製品よりもはるかに高いバックアップおよびリストアのスループットを実現しています。

次のグラフは、バックアップおよびリストアの持続可能な最大パフォーマンスを示しています。Exadata と Oracle ZFS Storage Appliance 間のバックアップおよびリストアのワークロードにおける物理スループット速度を、ネットワーク・レベルで測定しました。Oracle ZFS Storage ZS4-4 および ZS3-2 を特定のストレージ構成で使用する際の詳細な速度については、このホワイト・ペーパーの後半部分のパフォーマンス・サイジングに関するセクションを参照してください。

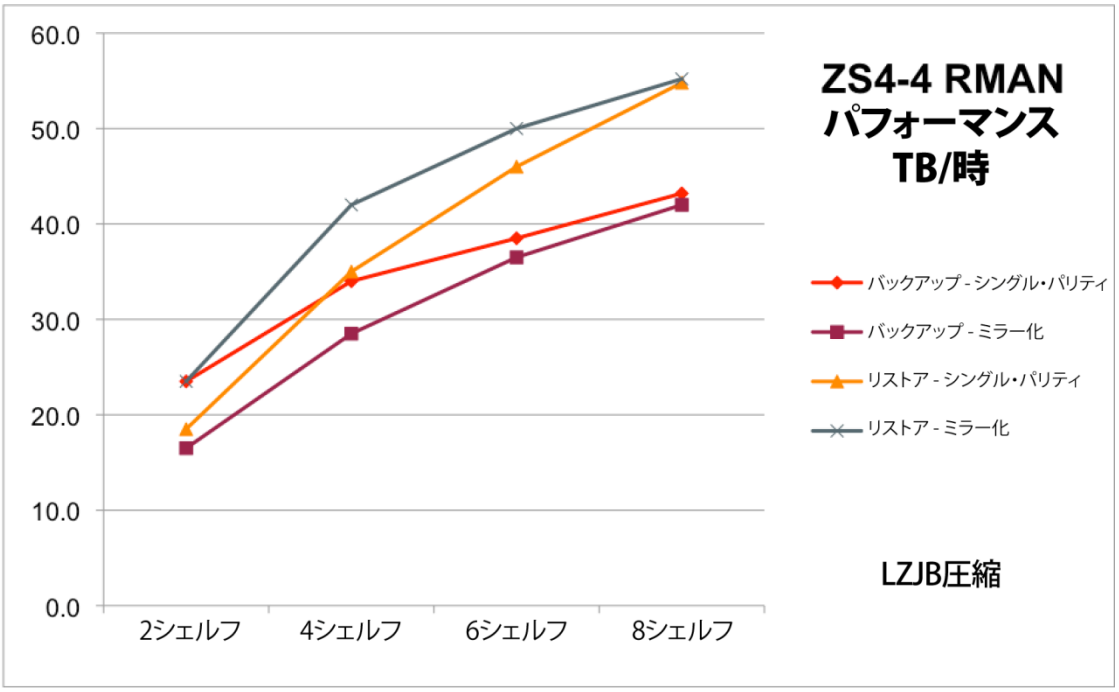

図1：Oracle ZFS Storage ZS4-4のバックアップおよびリストアで持続可能な最大スループット

これらは、Oracle Database 12c、および販売受注スキーマにお客様のサンプル・データが移入されている、大規模なオンライン・トランザクション処理（OLTP）データベースを使用した場合の実際の結果です。OLTPワークロードを実行中のお客様のベスト・プラクティスの推奨事項に対応するために、データベース・レベルでAdvanced Row Compressionを使用しています。これらのスループット速度は、データベースまたはI/Oジェネレータ・テスト・ツールを使用すると、誤った結果を導く可能性があるため、この手法では取得しませんでした。また、低レベルのシステム・ベンチマークからは予測されません。このホワイト・ペーパーで収集されたバックアップおよびリストアのパフォーマンス・データは、レベル0のバックアップ/リストア時以外はアイドル状態のデータベースを使用して測定されました。

リストアで 55TB/時のスループット速度を達成するには、80:20 DATA/RECO 分割および DATA ディスク・グループの標準 ASM 冗長性を備えた、Exadata Database Machine X5-2 フルラックが必要になります。または、同一の Oracle ZFS Storage ZS4-4 を利用して、同時に複数の小規模な Exadata 構成を使用した場合も、同じスループット速度を達成できます。小規模な構成に関するフル・パフォーマンス・サイジングの詳細については、このホワイト・ペーパーの後半部分で説明します。データベース・レベルの圧縮または増分バックアップ戦略に関しては、前のグラフで記録された物理速度よりもかなり速い有効バックアップ速度が、常に観察されています。

Oracle ZFS Storage ZS4-4 へのレベル 0 の全体データベース・バックアップでは、40TB/時を越える持続可能なスループット速度が達成されました。毎日 5-10%の更新を伴う、増分バックアップについては、40～80TB/時のバックアップ速度が得られました。ミラー化ストレージ・プロファイルを使用した場合、Oracle ZFS Storage ZS4-4 からのリストア速度で、50TB/時を超える値が測定されました。これを全体的に把握するために、ミラー化の際に一時ファイルを考慮した、250TB を超えるディスク領域を消費する 110TB データベースについて見てみると、レベル 0 のバックアップを実行した場合、3 時間未満でバックアップを実行できます。また、増分バックアップを実行した際には、その約半分の時間で完了することになります[1]。障害発生時には、同じデータベースを、2.5 時間未満でリストアできます。競合のソリューションを使用した場合、このサイズのデータベースのリストアに、数日を要することがあります。リカバリ時間目標（RTO）要件について議論する際には、数時間と数日というのは大きな差になります。Oracle ZFS Storage Appliance のリストア・スループット機能は優れているため、重要なデータベースをできる限り早くリカバリして使用可能な状態にします。

Oracle Exadata Database Machine を保護するソリューションを選択する際、重要になる考慮事項は高いパフォーマンスです。Oracle ZFS Storage Appliance が、これらのバックアップおよびリストア速度を達成可能なのは、以下のテクノロジーが採用されているためです。

InfiniBand のサポート - Oracle ZFS Storage Appliance を、非常に高い冗長性を備えたスケーラブルな InfiniBand アーキテクチャで構成できます。これにより、Oracle Exadata Database Machine とシームレスな統合が可能になり、CPU オーバーヘッドが比較的低い、高帯域幅で低レイテンシの I/O パスが実現されます。

Recovery Manager の統合 - Oracle Recovery Manager（Oracle RMAN）は高度にパラレル化されたアプリケーションであり、Oracle Database に常駐して、バックアップおよびリカバリ操作を最適化します。Oracle ZFS Storage Appliance は、最大 2,000 個の同時スレッドを利用して、Recovery Manager と統合できるように設計されています。これにより、複数のコントローラにわたる多数のチャネルに、I/O が分散されます。さらに、大部分のバックアップおよびリストアで一般的な、大規模ブロックの順次ストリーミング I/O ワークロードに関するパフォーマンスが著しく向上します。

Oracle Direct NFS - Oracle の最適化された NFS クライアントは、アグレッシブ実装です。これは、オペレーティング・システムをバイパスして、バッファを直接ユーザー領域に書き込むことにより、CPU とメモリのオーバーヘッドを低減し、データベース・プロセスごとに個別の TCP 接続を割り当てています。

[1]毎日の更新割合を 5%と仮定しています。

1MB のレコード・サイズ - 現在、Oracle ZFS Storage Appliance では、1MB という大きいレコード・サイズが使用できます。これにより、ディスクで必要な IOPS の数が低減され、Recovery Manager のバッファからストレージへの I/O サイズを保持したうえで、大規模ブロックの順次操作のパフォーマンスを向上させます。

ハイブリッド・ストレージ・プール - Oracle では、メモリ、フラッシュ、およびディスク間で動的ストレージ層を利用する、革新的なハイブリッド・ストレージ・プール（HSP）を導入しています。ダイナミック・ランダム・アクセス・メモリ（DRAM）と、特にマルチレベル・ストレージ向けに設計されたエンタープライズ・クラスのソフトウェアを有効に使用することは、Oracle ZFS Storage Appliance の優れたパフォーマンスを促進するための重要な要素です。

Oracle ZFS Storage Appliance は、世界最高レベルのスループットを達成しており、別個に監査済みの SPC-2 業界標準ベンチマークの価格/パフォーマンス・メトリックにおいて、第 1 位にランキングされています。これを、強力な機能、簡略化された管理、Oracle 間の統合と組み合わせることにより、Oracle Exadata Database Machine 上のミッション・クリティカルなデータを保護するための魅力的なソリューションを実現しています。

はじめに

データベースやシステム、ストレージの管理者は誰でも、データベースのバックアップやリカバリ時に、より多くのデータをより頻繁に、より短い時間でバックアップしながら、予算内に収めるにはどうすれば良いかという難問を抱えています。さらに、実世界での機能停止にまつわる実際的な課題のため、危険にさらされた状況でも確実に円滑な操作を実行できるように、データ保護システムはシンプルで信頼性の高いものにしなければなりません。Oracle ZFS Storage Appliance は、このような難問に取り組む管理者を支援するために、NFS プロトコルの簡便性と ZFS で強化されたディスクの信頼性を併せ持つ、コスト効率のよい高帯域幅ストレージ・システムを提供します。管理者は、Oracle ZFS Storage Appliance テクノロジーを介して、データ保護に関連する資本費用や運用費用を削減しながら、最終顧客との厳しい品質保証契約を守ることができます。

Oracle ZFS Storage Appliance は、Oracle Exadata Database Machine 内のデータの保護に最適な、他に類を見ない、展開しやすい統合ストレージ・システムです。ネイティブな InfiniBand（IB）や 10 ギガビット（Gb）イーサネットに接続できる Oracle ZFS Storage Appliance は、Oracle Exadata には理想的です。このような高帯域幅インターコネクトにより、従来型のネットワーク接続ストレージ（NAS）のシステムと比較して、バックアップやリカバリにかかる時間が短縮されるだけでなく、バックアップ・アプリケーションのライセンスやサポートにかかる費用も削減されます。Oracle ZFS Storage Appliance ソリューションでは、従来の階層型バックアップと増分更新バックアップの両方の戦略がサポートされているため、リカバリ時間をさらに削減してシステム管理を簡略化できるような、拡張ストレージでの効率性が実現されます。

Exadata データベース上のミッション・クリティカルなデータを保護するために Oracle ZFS Storage Appliance を展開する際に、災害時のタイムリーなリカバリを保証するには、バックアップ期間とリカバリ時間目標（RTO）が一致していなければなりません。このホワイト・ペーパーでは、Oracle データベースの最適なバックアップとリカバリのために Oracle ZFS Storage Appliance をセットアップするためのベスト・プラクティスを説明するとともに、Oracle Exadata のチューニングに関する具体的なガイドラインを示します。

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。

- データ保護戦略の選択
- Oracle ZFS Storage Applianceの構成
- Oracle Exadata Database Machineの構成
- 従来型および増分更新バックアップのベスト・プラクティス
- Oracle ZFS Storage ZS4-4およびZS3-2でのRecovery Managerのバックアップ/リストア・パフォーマンス・サイジング

データ保護戦略の選択

Oracle データベースに最適なバックアップ戦略は、現在のリカバリ時間目標（RTO）、リカバリ・ポイント目標（RPO）、およびバージョン保存目標（VRO）に適合するデータ保護ソリューションを構築する際に、重要になるステップです。さらに、もっとも重要な点は、バックアップは即座に効率的に実行する必要があることです。これにより、エンドユーザーのアプリケーションには一切影響を与えず、本番環境のハードウェアで消費されるリソースも最小限になります。Oracle ZFS Storage Appliance で Recovery Manager のワークロードを最適化する際には、バックアップ戦略のタイプにも関連が出てきます。以下の章では、特定のベスト・プラクティスについて説明します。

従来型のバックアップ戦略

従来型のバックアップ戦略は、変更されていない全体バックアップ、または変更されていないレベル 0、レベル 1 の累積増分とレベル 1 の差分増分バックアップを組み合わせて使用し、物理的または論理的な障害の発生時に、データベースの一部をリストアおよびリカバリする戦略のことです。

従来型のバックアップ戦略のうちもっとも単純な実装は、データベース全体の定期的な全体バックアップです。このバックアップは、データベースが開かれていてアクティブな場合に実行できます。全体バックアップは、毎週または毎日のユーザー定義のスケジュールに従って実施されます。データベースのトランザクション・アーカイブ・ログは、Oracle ZFS Storage Appliance 上に直接格納されている 1 つのコピーを使用して、多重化する必要があります。このアーカイブ・ログは、リストアされた全体バックアップをリカバリし、障害発生前の最後のオンライン REDO ログ切り替え時間までのすべてのトランザクションに適用されます。REDO ログのサイズが適切に設定されている場合、このソリューションの RPO が 20 分を越えることはありません。一般的に、VRO では、最低 2 つの全体バックアップを常時アクティブにすることが要求されます。この際、Recovery Manager は自動的に有効期限が切れ、最終的に以前のバックアップは消去されます。厳格な RTO が指定されている小規模なデータベースでは、毎日の全体バックアップが理想的です。

一般的な実装は、増分レベル 0 とレベル 1 のバックアップを組み合わせた階層型の手法です。大抵の場合、レベル 0 の増分バックアップを毎週、レベル 1 の差分または累積増分バックアップを毎日実行します。

Recovery Manager ブロック・チェンジ・トラッキングは、増分バックアップのパフォーマンスを向上させるために使用されます。レベル 0 の増分バックアップは、データベース全体をスキャンしますが、レベル 1 増分バックアップは、ブロック・チェンジ・トラッキング・ファイルを使用して、最後のバックアップ以降に変更されたブロックのみをスキャンします。これにより、データベースで必要な読取り量が大幅に低減されます。

次に、毎週のレベル 0 バックアップと毎日のレベル 1 差分バックアップを組み合わせて利用する、従来型の Recovery Manager のバックアップ戦略の例を示します。レベル 0 のバックアップは、内容的には全体データベース・バックアップに相当します。差分とは、各レベル 1 において、最後のレベル 0 またはレベル 1 のバックアップ以降に変更されたデータのみをバックアップするということです。

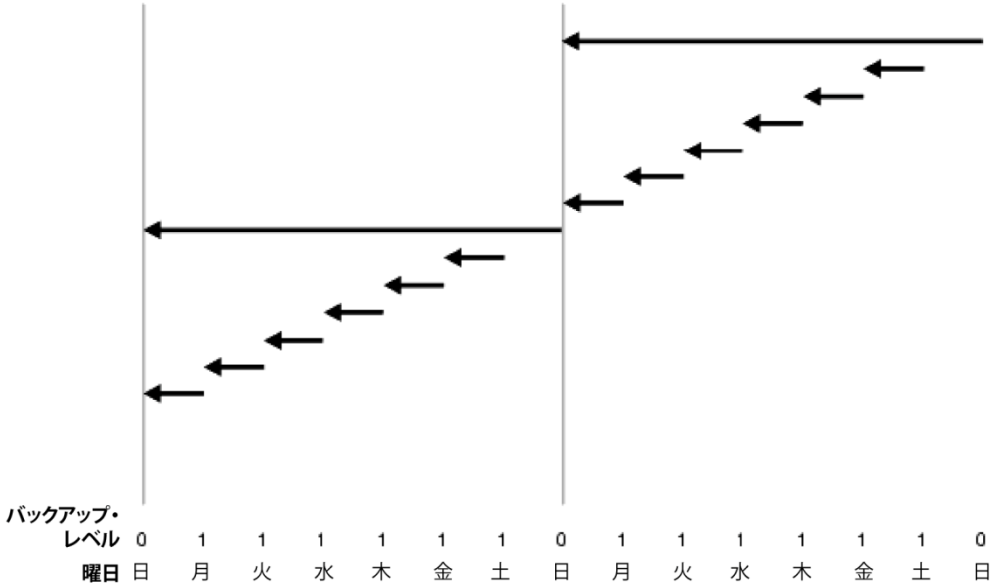

図2：毎日の差分増分バックアップを使用する従来型のバックアップ戦略

この例の VRO では、バックアップ・データが常時 2 週間保持されるように指定されています。レベル 0 増分バックアップは、毎週日曜日に実行され、レベル 1 の差分増分バックアップは、その他のすべての日に実行されます。レベル 1 の差分増分バックアップでは、最後の 24 時間以内に変更されたデータのみがバックアップされます。Recovery Manager のブロック・チェンジ・トラッキングが使用されているため、月曜日から土曜日に実行されるバックアップ操作では、データベースのごく一部分のみがスキャンされ、ネットワークを介して少量のデータのみが送信されて、ディスクに書き込まれます。この例では、アーカイブ・ログは、Oracle ZFS Storage Appliance 上に直接格納されている 1 つのコピーを使用して、多重化されます。

データベースまたはデータベースの一部は、2 週間の範囲内の任意のポイントにリストアしてリカバリできます。Recovery Manager は、指定したリカバリ・ポイントより前の最新のレベル 0 バックアップをリストアし、レベル 0 とリカバリ・ポイント間の後続のレベル 1 バックアップをすべてリストアしてから、必要なアーカイブ・ログ・トランザクションを適用します。バックアップからのデータのリストアは、アーカイブ・ログでトランザクションを適用するより高速です。この例では、月曜日に発生した障害からのリカバリは、比較的高速で簡単に実行されますが、土曜日に発生した障害からのリカバリは、5 つの差分バックアップをリストアする必要があるため、かなり時間を要するプロセスになります。

次に、毎日のレベル 1 差分増分とレベル 1 累積増分を混在させ、毎週のレベル 0 増分バックアップと組み合わせて利用する例を示します。累積増分には、最後のレベル 0 バックアップ以降に変更されたデータのみが含まれています。また、累積増分は、差分よりも大きい領域を消費しますが、リストア・プロセスが簡略化されます。

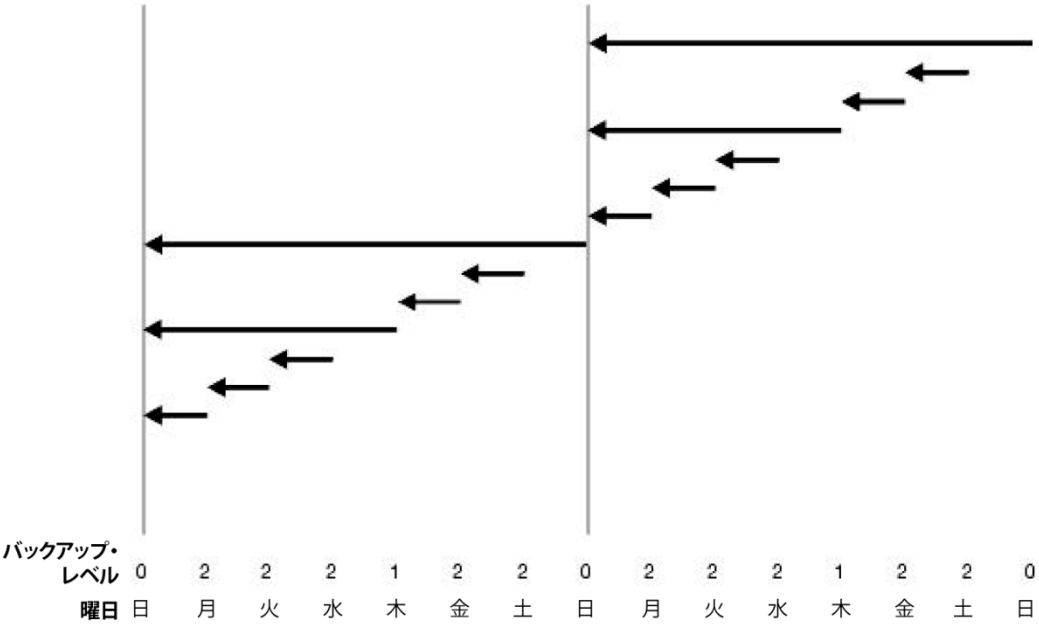

図3：毎日の差分および累積増分バックアップを使用する従来型のバックアップ戦略

この例での相違点は、週の半ばのバックアップがレベル 1 の累積増分バックアップになっていることです。これにより、木曜日のバックアップが大規模になりますが、リストア・プロセスが簡略化されます。たとえば、金曜日に発生した障害からリカバリを行う場合、2 つのバックアップのリストアのみですみます。

従来の階層型の増分バックアップ戦略には、以下のいくつかのメリットがあります。

- スループット速度の向上ー最大バックアップ速度 42TB/時および最大リストア速度 55TB/時が使用可能です。この場合、並列ストリーミング I/O により、大規模なレコード・サイズを処理できるため、Oracle ZFS Storage Appliance と従来型の Recovery Manager のワークロードを使用します。
- ブロック・チェンジ・トラッキングを使用することにより、帯域幅消費量を削減し、毎日のバックアップを高速化。
- 増分変更データ手法を使用することにより、毎日の全体バックアップと比較して、バックアップの消費容量を大幅に低減。
- テープ・アーカイブとの相乗効果により、バックアップ・セット用のネイティブの最適化を実現。
- 未使用のデータファイル・ブロックをバイパスした、全体マルチセクションをサポート。
- 2 番目のレベル 0 により、単一障害点(SPOF)を発生させずに、前のアクティブなレベル 0 を暗黙的に検証。

増分更新バックアップ戦略

増分更新バックアップ戦略では、最初にレベル 0 イメージ・コピー・バックアップが作成されます。これは、Oracle ZFS Storage Appliance に格納されている、データファイルの一貫性のあるイメージ・コピーと同等です。以降のバックアップは、すべて差分増分であり、最後のバックアップ以降に変更されたデータのみを取得します。通常は、最後の 24 時間以内に変更されたデータのみを含む Recovery Manager バックアップ・セットにより、毎日実行されます。次に、イメージ・コピー・バックアップに対して前の増分バックアップが適用され、時間内にロール・フォワードされます。これにより、アクティブなデータベースより 1 日以上遅れて証跡される、レベル 0 の一貫性のあるイメージ・コピーを指定することにより、リストアが簡略化されます。次に示す図は、このプロセスを視覚的に表現したものです。

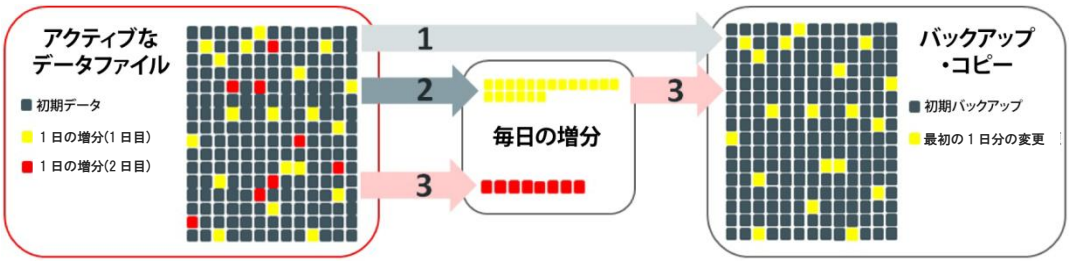

図4：Recovery Managerの増分更新バックアップ戦略

この例では、Recovery Manager の同じジョブが毎晩実行されます。最初の夜は、このジョブにより、データベース内のすべてのデータファイルに対して、レベル 0 のイメージ・コピー・バックアップが作成されます。これは、灰色のブロックで表示されます。2 日目の夜は、Recovery Manager により、最後の 24 時間以内に変更されたデータのみのバックアップ・セットが作成され、別の共有に格納されます。これは、黄色のブロックで表示されます。3 日目の夜、Recovery Manager により、最後の 24 時間以内に変更されたデータのみのバックアップ・セットが作成され（赤色のブロック）、次にバックアップ・コピーに対して、前夜に変更されたデータが適用されます（黄色のブロック）。以降は夜ごとにバックアップが実行されますが、最後の 24 時間以内に変更されたデータのみが同じ機能でバックアップされ、バックアップ・コピーに対して、前に変更されたデータが適用されます。

従来型の Recovery Manager の戦略と同様に、開かれていてアクティブなデータベースにおいて、増分更新バックアップが実行できます。アーカイブ・ログは、Oracle ZFS Storage Appliance 上に直接格納されている 1 つのコピーを使用して、多重化する必要があります。この場合、ソリューションの RPO が 20 分を越えることはありません。

増分更新バックアップ戦略には、以下のようなメリットがあります。

- 実行されるレベル 0 バックアップの回数を制限
- 大部分の状況でリストア・プロセスを簡略化
- 開発およびテスト用として、プロビジョニングとイメージ・コピーによる相乗効果を実現
- 単一のレベル 0 バックアップのみが保守されている場合、ディスク領域の消費量を低減

Oracle ZFS Storage Applianceを使用したベスト・プラクティス

Oracle ZFS Storage Appliance により、Exadata 上のデータベースを保護する際は、大部分の状況において、従来型の Recovery Manager のバックアップ戦略を使用する必要があります。

従来型の Recovery Manager のバックアップ戦略では、未使用ブロックのスキップ、NULL ブロックの圧縮、マルチセクションのサポート、および複数の入力ファイルの単一のバックアップ・セットへの組込みなどのテクノロジーによるメリットがあります。また、従来型の Recovery Manager のバックアップ戦略では、ZFS 共有で 1M のレコード・サイズが使用できるため、大規模なストリーミング I/O が排他的に利用されます。Recovery Manager では、非共有バッファが使用されます。これにより、ビジー状態で別のバッファがクリアされるのを待機する時間が低減され、チャネルごとのスループットが向上します。レベル 0 のバックアップが高速で完了し、ディスク上で消費する領域が低減されます。リストア操作のスループット速度も向上します。

増分更新バックアップ戦略では、レベル 0 のバックアップが完了するまでに時間を要し、ディスク上で消費する領域も増大します。レベル 0 のバックアップおよびマージ操作では、レベル 0 のバックアップに対して増分の変更が適用されますが、この操作では、チャネルごとのスループットが低減された、Recovery Manager の共有バッファが利用されます。通常、ZFS 共有で使用されるレコード・サイズは小さくする必要があります。また、マルチセクションのサポートには制限があります。その結果、リストアのスループット速度が低くなります。

従来型のバックアップ戦略の別のメリットは、特定のお客様のニーズおよびアーカイブからテープへのシームレスな標準の統合に対して最適化されたバックアップ戦略の設計が、より柔軟に行えることです。従来型の Recovery Manager のバックアップにおける特定のユースケースに対して、Oracle ZFS Storage Appliance を構成する場合、書込み最適化フラッシュは必要ないため、低価格で実現できます。

増分更新バックアップ戦略にもメリットがありますが、ソース・データベースが大規模で、毎日の更新割合が非常に小さい状況で使用する必要があります。運用時のレベル 0 バックアップでは、増分更新の必要性はないかまたは低く、変更されたデータのみがネットワークを介して送信されるため、これらのシナリオでこのバックアップ戦略を実装することは大きなメリットです。

一般的なガイドラインとして、増分更新バックアップ戦略は、以下の両方が該当する場合に検討する必要があります。

- ソース・データベースが十分に大規模で、毎週のレベル 0 バックアップが、ネットワーク、ディスク、またはサーバーの各リソースに対して、何らかの影響を与える可能性がある場合
- データベースの毎日の更新割合が 3%以下の場合

増分更新バックアップ戦略では、開発およびテスト・プロビジョニング用の ZFS スナップショット・クローン作成との相乗効果を実現しており、リストア・プロセスを簡略化できます。ただし、保守しているのが 1 つのレベル 0 バックアップのみの場合、単一障害点(SPOF)となる能性があるため、大部分の実装では、バックアップからリストアできる短い期間のみが設定されます。ここでは、n-1 証跡更新が標準です。

増分更新および従来型戦略の両方において、毎日の増分バックアップでインラインの最適化重複排除が用意されています。変更されたデータのみが、ネットワーク経由で送信されます。両方の戦略ともに、要求の厳しいリカバリ・ポイント目標を満たしています。障害ポイントと最新のアーカイブ・ログ・バックアップ間で、20 分を越えることはありません。

アーカイブ・ログ・トランザクションの適用は、バックアップのリストアよりもリソース集約型です。リカバリ時間目標が厳格なデータベースの場合、従来型増分戦略では、次回のレベル 0 バックアップが行われる直前に障害が発生すると、その目標を満たすのが困難になることがあります。この場合、複数のリストアが必須で、その後アーカイブ・ログから REDO トランザクションが適用されます。このような状況では、差分の代わりに累積増分バックアップを使用して、リストア・プロセスを簡略化できます。

Oracle ZFS Storage Appliance の優れたリストア・スループットを組み合わせた、バックアップ戦略をカスタマイズする機能が備えられているため、RTO が満たされたうえで、それを上回る結果が得られます。

Oracle ZFS Storage Applianceの構成

次のセクションでは、Oracle ZFS Storage Appliance を最適化して、Oracle Exadata 環境でデータベースを保護するためのベスト・プラクティスについて説明します。

コントローラの選択

Oracle ZFS Storage Appliance は、ZS4-4 および ZS3-2 の 2 つのモデルが使用できます。

表1：Oracle ZFS Storage Applianceの詳細

機能	ZS3-2	ZS4-4
CPU コア数	32	120
DRAM	512GB または 1TB	2TB
書込み最適化フラッシュ	4TB	10.5TB
読取り最適化フラッシュ	12.8TB	12.8TB
物理容量（最大）	1.5PB (4TB HDD) 3.1PB (8TB HDD)	3.5PB (4TB HDD) 6.9PB (8TB HDD)
HA/クラスタ・オプション	あり	あり
フォーカス	ミッドレンジ	スケーラビリティ

Oracle ZFS Storage ZS4-4 は、フラッグシップ製品であり、最大レベルのスケーラビリティ、CPU、および DRAM が搭載されています。DRAM は最大 3TB、書込み最適化フラッシュは最大 10TB、読取り最適化フラッシュは最大 12TB が使用可能な、非常にスケーラブルなプラットフォームであり、物理ストレージ容量は最大 6.9 ペタバイト（PB）をサポートできます。

Oracle ZFS Storage ZS3-2 は、コスト効率の高いモデルであり、極めて高いレベルのスループットおよび冗長性を達成できます。また、DRAM は最大 1TB、書込み最適化フラッシュは最大 4TB、読取り最適化フラッシュは最大 12TB、および物理ストレージ容量は最大 3.1PB をサポートできます。

これらのモデルはいずれも、Oracle Exadata Database Machine を保護するための最適な選択肢です。現在の環境に最適な製品を決定する際は、以下のいくつかの要因について検討してください。

- 通常、大規模な順次ストリーミング・ワークロードでは、書込みおよび読取り最適化フラッシュが存在してもメリットはありません。さらに、これらの条件下で優れたパフォーマンスを達成するには DRAM が重要ですが、DRAM の容量を過度に増やす必要はありません。また、これ以上パフォーマンスは向上しません。従来型の Recovery Manager のバックアップおよびリストア・ワークロード（大規模なストリーミング I/O）に対して、Oracle ZFS Storage Appliance が排他的に 100%使用されている場合、書込みおよび読取り最適化キャッシュが存在しないシステムを利用することを推奨します。
- 非バックアップ・データベース I/O、増分更新バックアップ・ワークロード、開発/テスト・プロビジョニング用のクローンの作成、およびその他多数の I/O が混在するシナリオで、良好なパフォーマンスとユーザビリティを達成するには、大抵の場合、書込みおよび読取り最適化キャッシュが推奨されています。大規模な容量の書込み最適化フラッシュを備えた Oracle ZFS Storage Appliance では、効率的に使用できるタイプのワークロードでの柔軟性が向上します。これは、現在または将来のアクティビティ計画において重要になる場合があります。

- システムの主要な目的が、直接トランザクション・データベース・ワークロードを実行することである場合、大容量の DRAM と CPU リソースを搭載することが重要なメリットになります。
- Oracle ZFS Storage ZS3-2 は、極めて低価格で優れたスループットと冗長性を提供しています。これは、ディスク・シェルフが 2～4 つの小規模な構成に最適です。ただし、ストリーミング I/O に注目した大規模な構成でも適切に実行されます。

適切なディスク・シェルフの選択

Oracle ZFS Storage Appliance では、ディスク・シェルフに 2 つのオプションが用意されており、両方とも同程度の価格帯です。Oracle Storage Drive Enclosure DE2-24C には、4TB または 8TB の大容量ディスクが備えられており、Oracle Storage Drive Enclosure DE2-24P には、900GB の高パフォーマンス・ディスクが備えられています。各シェルフには 24 台のディスクが含まれており、両方とも同じ書込み最適化フラッシュ・オプションで構成できます（シェルフごとの最大 4 台のディスクは、SSD 書込みフラッシュ・アクセラレータで置き換え可能です）。Oracle ZFS Storage ZS4-4 および ZS3-2 は、ディスク・シェルフおよび書込み最適化フラッシュ要件に基づいて、カスタマイズできます。

表2：Oracle Storage Drive Enclosureモデル（ディスク・シェルフ）の詳細

ディスク・シェルフ	サイズ/ディスク	RPM	IOPS/ディスク	MBPS/ディスク	ラック・ユニット
DE2-24P[2]	900GB	10K	160	170	2
DE2-24C	4TB	7.2K	120	180	4

Oracle ZFS Storage Appliance により、Oracle Exadata を保護する場合は、大容量の Oracle Storage Drive Enclosure DE2-24C ディスク・シェルフを推奨します。容量が大きくなり、スループットも多少高くなることは、大部分のバックアップ・ユースケースにおいては、重要なメリットになります。Oracle Storage Drive Enclosure DE2-24P は、IOPS が高く低レイテンシであることが重要なメリットになる場合、またはラック・スペースに制限があり小規模な部分ラック構成でパフォーマンスを最大限にする必要がある場合に、検討が必要になることがあります。

[2] Oracle Storage Drive Enclosure DE2-24P は、データ保護のユースケースでは推奨されていませんが、ここには記載されていない 300GB ディスク・オプションでも使用できます。

ストレージ・プロファイルの選択

Oracle Exadata Database Machine を保護するためのストレージ・プロファイルを選択する際には、ミラー化、シングル・パリティ、およびダブル・パリティのすべてが考慮に値する項目です。

表3：ストレージ・プロファイルの比較

ストレージ・プロファイル	使用可能容量[3]	メリット	デメリット
ミラー化	42.2%	高いリストア・パフォーマンス 最大限の保護 最大限の柔軟性	高コスト
シングル・パリティ	69.3%	高いバックアップ・パフォーマンス 中程度の柔軟性	冗長性が制限
ダブル・パリティ	76.7%	高いストリーミング・パフォーマンス もっとも効率的	IOPS が制限

ミラー化

ミラー化は、特にリストアに関して、強力な冗長性と堅牢なパフォーマンスを実現しているため、大抵の場合に推奨されるストレージ・プロファイルです。シングル・パリティ実装時の 2 倍の仮想デバイス（vdevs）が生成されるため、ミラー化ストレージ・プールでは、多数の IOPS を処理できます。これは、従来型の Recovery Manager のワークロードなどの大規模な順次 I/O でも適切に実行される柔軟性を備えています。また、直接的なデータベース OLTP トランザクションなどの小規模なランダム I/O を生成するワークロードでも、優れたパフォーマンスを達成しています。

ミラー化プロファイルを選択した場合のデメリットは、他の 2 つのオプションの場合よりも大きいディスク領域を消費して、書込み時に生成される内部帯域幅が大きくなることです。これは、SAS または PCI の帯域幅が制限要因である状況では顕著な影響を与えます。

ミラー化が推奨されるのは、最適なパフォーマンス、特にリストアのパフォーマンスを達成して、重要なビジネス・アプリケーションを実行中のデータベースに対して、可能な限り最短の RTO を提供することが重要になる場合です。

また、増分更新バックアップ戦略、開発/テスト・プロビジョニング用のクローンの作成、または直接データベース・ワークロードに集中する状況においても、推奨されます。データベースで OLTP ワークロードを実行する場合（大部分は、データの既存の行の変更と読取りに焦点を当てた、小規模のトランザクションが特徴）、または、1 日を通して分散される書込みトランザクションの要素が存在する場合、小規模のランダム I/O が生成され、これらのユースケースではストレージ・プール上に IOPS 負荷が配置されます。ミラー化プロファイルは、大量の IOPS を処理する際には最適です。

[3] 使用可能容量は、パリティ、スペア、ファイル・システムのオーバーヘッドなどが原因で、物理容量が欠落することに関係があります。同様に、各ディスクで少量の領域が欠落するのは、OS のオーバーヘッド、ドライブ・メーカーのオーバーヘッド、スクラッチ領域の予約などが原因です。ただし、ストレージ・プールのサイズに応じて多少異なります。この例では、4 つのディスク・シェルフ構成を仮定しています。

シングル・パリティ

シングル・パリティは、データ保護シナリオで推奨されることが多い、中程度の手法オプションです。これは、従来型の Recovery Manager のワークロードに対して最適なバックアップ・パフォーマンスを実現しており、特に使用可能容量の問題により、ミラー化が適切ではない場合に魅力的なオプションです。

シングル・パリティは、幅の狭い 3+1 ストライプを実装して、強力な ZFS 機能を利用することにより、大規模のストリーミング I/O で優れたパフォーマンスを実現します。ただし、いくつかのランダムまたは小規模の I/O ワークロードを処理するのに十分な柔軟性も備えられています。

シングル・パリティが推奨されるのは、バックアップ・パフォーマンスに注目している場合、使用可能容量の問題により、ミラー化が適切ではない場合、および増分更新バックアップ、開発/テスト・プロビジョニング、直接的なデータベース・ワークロードなどのユースケースにおいて多少は使用されるが、それがおもな目的ではない場合です。シングル・パリティでは幅の狭いストライプが使用されるため、中程度の量の vdevs（ミラー化の際の半分）が生成されますが、大規模な非ストリーミング I/O を生成するワークロードを処理する柔軟性も備えられています。ただし、OLTP データベースでの増分更新バックアップ戦略などの IOPS 集約型ワークロードでは、データファイル・コピーに対してミラー化プロファイルを利用する必要があります。

ダブル・パリティ

ダブル・パリティでは、最適な使用可能容量が得られ、従来型の Recovery Manager のワークロードでは一般的な大規模のストリーミング I/O に関しては、シングル・パリティの場合とほぼ同じように実行されます。これを実行するために、幅の広いストライプが使用されます。この幅は、ストレージ・プール作成時に構成内にあるディスク数に応じて異なりますが、最大で 14 ディスクです。その結果として、ダブル・パリティのストレージ・プール内にある vdevs の数は、ミラー化またはシングル・パリティの場合と比較してはるかに少なくなります。ただし、IOPS 集約型のワークロードを処理する能力は、かなり低下します。

ダブル・パリティが推奨されるのは、Oracle ZFS Storage Appliance が、従来型の Recovery Manager のバックアップおよびリストアなどの大規模の順次ワークロードに対して、100%の能力で対応できる状況のみです。将来、開発/テスト・プロビジョニング用のクローン作成や増分更新バックアップ戦略の利用などのユースケースを追加で導入する可能性がある場合、この手法は推奨されません。ミラー化またはシングル・パリティ・プロファイルは、短いレイテンシで大量のディスク IOPS が発生する可能性があるユースケースの処理に対しては、かなり柔軟性があります。

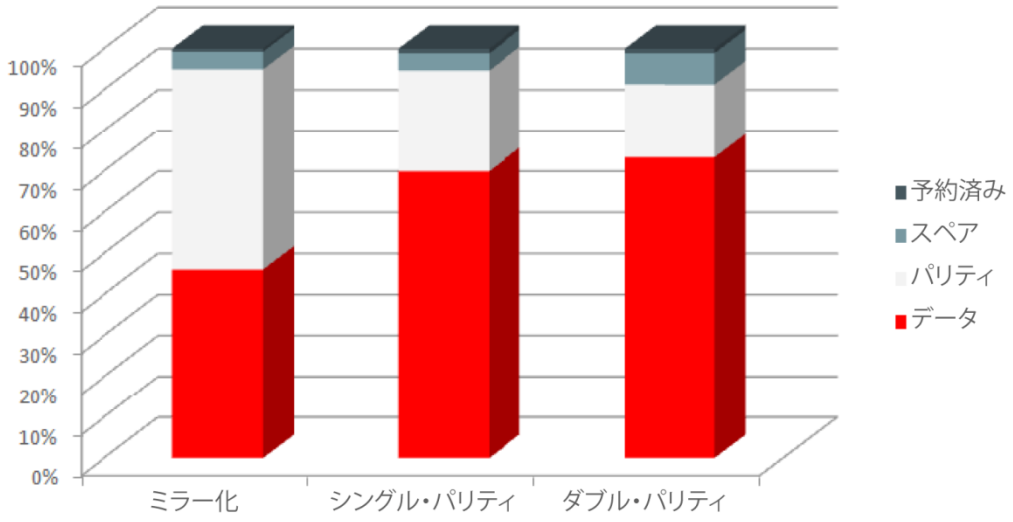

図5：物理ディスク容量の分散

ストレージ・プールの構成

大抵の場合、コントローラごとに単一のストレージ・プールを構成することを推奨します。各ストレージ・プールは、ディスク・シェルフごとに使用できるハード・ディスク・ドライブ（HDD）の半分で構成する必要があります。これにより、最大のパフォーマンスと冗長性が実現できます。

ストレージ・プールを構成する際には、単一障害点のない（NSPF）オプションを選択することを推奨します。これにより、ディスク・シェルフ全体の欠落によって、データの可用性が低下することがなくなります。

書込みフラッシュ・アクセラレータおよび読取り最適化フラッシュの使用

Oracle ZFS Storage Appliance では、コスト効率の高い、一意の高パフォーマンス・ストレージ・アーキテクチャが提供され、これはフラッシュ優先のハイブリッド・ストレージ・プール・モデル上に構築されます。パフォーマンス・オンデマンド手法では、オプションの書込みフラッシュ・アクセラレータおよび読込み最適化フラッシュ・デバイスを、ストレージ・プールに構成できます。

フラッシュベースのキャッシュを利用して、Oracle のハイブリッド・ストレージ・プール・アーキテクチャの能力を引き出すことは、トランザクションまたは混合 I/O ワークロードにより、最適なパフォーマンスを達成するために重要です。ただし、従来型の Recovery Manager のバックアップおよびリストア・ワークロードでは、フラッシュ・デバイスが存在することによるメリットはありません。これらのワークロードでは、1MB のストリーミング I/O が生成され、データセットが非常に大きくなります（大抵の場合、1TB を超える）。通常、バックアップ時のシステム・スループットは、メモリ内のデータが HDD に同期できる速度によって決定されます。さらに、Recovery Manager のバックアップは、レベル 2 の読取りキャッシュの移入に対しては優先度が低くなります。その結果、データセットが大きくなり、リストアの頻度が低下します。

ソリッド・ステート・ディスク（SSD）は、HDD と比較してコストがかなり高くなるため、従来型の Recovery Manager のワークロードで排他的に使用される、Oracle ZFS Storage Appliance 構成に対して、フラッシュ・デバイスは推奨されません。これらの環境では、ディスク・シェルフ（HDD）を追加して、パフォーマンスと容量を向上させる方がメリットがあります。ただし、開発/テスト・プロビジョニング用のデータベース・クローン作成がお客様のユースケースであるか、または増分更新バックアップ戦略が実装されている場合、フラッシュ・デバイスが推奨されるか、または必須になることもあります。従来型の Recovery Manager のワークロードおよび増分更新バックアップ戦略のベスト・プラクティスのセクションでは、フラッシュを使用するタイミングに関する詳細なガイドラインについて説明します。

InfiniBandまたは10GbEの選択

Oracle ZFS Storage Appliance は、Exadata Database Machine 上のデータを保護する目的で、InfiniBand または 10GbE のいずれかで構成できます。

Exadata は、ネイティブの InfiniBand インフラストラクチャを利用しますが、これにより、10GbE の場合と比較して帯域幅が向上し、レイテンシが短くなり、システム・リソースの利用率が低減されます。Oracle ZFS Storage Appliance は、Exadata ラックで事前構成されている 2 つの InfiniBand リーフ・スイッチ（NM2-36P）に接続することにより、シームレスに統合できます。

大抵の場合、Exadata に接続する際には、InfiniBand を利用することを推奨します。このホワイト・ペーパーに記載されている、Recovery Manager のバックアップ/リストアのスループット速度およびパフォーマンス・サイジングのデータは、InfiniBand を使用してラボで収集しました。ただし、まれにですが、10GbE の方が適している場合があります。これらの状況に関する例として、InfiniBand の展開を禁止する距離の制限事項や、単一の Oracle ZFS Storage ZS4-4 への 5 つ以上の分離 Exadata のバックアップなどがあります。

10GbE 構成での Recovery Manager のバックアップおよびリストアのスループット速度は、ディスクが制限されている小規模な構成の場合と大きく異なることはありません。システム・リソースによって制限を受ける大規模な構成では、10GbE に切り替えた場合と比較しても、スループットは 10%未満低下するだけです。開始時にネットワーク帯域幅が制限され、アクティブな 10GbE リンクが追加されていない構成では、スループットはかなり低下します。

Oracle ZFS Storage Appliance を Exadata InfiniBand ファブリックに接続する際は、各 Exadata ラックの 2 つのリーフ・スイッチで使用可能なポートがすべて利用できます。Exadata Database Machine X5-2 フルラック構成では、ポート 5B、6A、6B、7A、7B、8B、および 12A が使用できます。部分ラック構成では、追加のポートが使用できます。

可用性を最適化して、InfiniBand スイッチの欠落を許容するには、通常の構成では、各 HCA のポート 1 を下位のリーフ・スイッチに接続し、ポート 2 を上位のリーフ・スイッチに接続する必要があります。詳しくは、AIE ホワイト・ペーパー『Configuring an Oracle ZFS Storage Appliance with Multiple Oracle Exadata Database Machines』に記載されています。

Oracle Direct NFSの選択

Oracle Direct NFS（dNFS）は、Exadata と Oracle ZFS Storage Appliance 間のすべてのデータベースの Recovery Manager のワークロードに対して強く推奨します。ここでは、最適なパフォーマンスを達成する必要があります。

dNFS は、カスタム NFS クライアントで、データベース・カーネル内に常駐し、以下に示すおもなメリットを提供します。

- オペレーティング・システム（OS）をバイパスして、ユーザー領域で 1 度だけデータをキャッシュし、カーネル領域には 2 番目のコピーを作成しないことにより、システム CPU の利用率を大幅に低減
- データベース・プロセスごとに、個別の伝送制御プロトコル（TCP）接続を開くことにより、パラレル I/O のパフォーマンスを向上
- ラウンドロビン法により、バッファを複数の IP アドレスに交互に置き換えることにより、複数のネットワーク・インタフェース間にスループットを分散
- 障害の発生した I/O を代替アドレスに自動的にリダイレクトすることにより、高可用性（HA）を実現

これらの機能により、帯域幅が向上し、CPU オーバーヘッドが低減されます。

NFS サーバー・スレッドの最大数を、デフォルトの 500 から増やすことを推奨しますが、Oracle ZFS Storage Appliance では、dNFS を有効にするための追加の手順は必要ありません。これを実行するには、Oracle ZFS Storage Appliance BUI にアクセスして、「Configuration」・「Services」・「NFS」を選択し、スレッド数を 1000 に設定します。

Oracle Intelligent Storage Protocol

Oracle Intelligent Storage Protocol は、dNFS とともに Oracle Database の 12c バージョンで導入されました。これを使用して、Oracle ZFS Storage Appliance でレコード・サイズと同期書込みバイアスを動的にチューニングすることにより、データベース認識ストレージが使用可能になります。これにより、構成プロセスが簡略化され、構成エラーが原因であるパフォーマンスへの影響が低減されます。Oracle データベース・カーネルから、Oracle ZFS Storage Appliance にヒントが渡されます。これらのヒントは、ストレージ設定を動的に最適化するワークロード・プロファイルを構築するために解析されます。

Oracle Intelligent Storage Protocol は、オプションのプロトコルです。現在の実装では、このホワイト・ペーパーにおいてベスト・プラクティスに準拠し、適切に構成されている環境は、Oracle Intelligent Storage Protocol なしでも同様に、問題なく実行されます。ここでは、NFSv4 と SNMP が必要です。Oracle Intelligent Storage Protocol を有効にする手順については、My Oracle Support（MOS）Document 1943618.1 を参照してください。

IPマルチパスの構成

IP マルチパス・グループ（IPMP）は、完全な HA 冗長性を提供するために推奨されています。dNFS は、あるレベルの HA は提供できますが、現在では、ファイルを開くまたは作成する際は、カーネル NFS マウントに依存しています。IP マルチパスは、すべての状況において、完全な HA を提供するために必須になります。

この例では、各 ZFS コントローラに、2 つの InfiniBand HCA が取り付けられていると仮定しています。IP マルチパスを構成するには、以下の手順を実行します。

1. ibp0、ibp1、ibp2、および ibp3 の InfiniBand データ・リンクを作成します。パーティション・キーを ffff に設定し、リンク・モードを接続モードに設定します。
2. ibp0、ibp1、ibp2、および ibp3 のネットワーク・インタフェースを作成します。各ネットワーク・インタフェースでは、アドレス 0.0.0.0/8 を使用します。
3. ibp0 と ibp3 を使用して、両方のポートをアクティブに設定し、最初の IPMP ネットワーク・インタフェース（ib-ipmp-controller1）を作成します。dNFS でパフォーマンスを最適化するために、このリンクで 2 つの別の IP アドレスを作成します。IP アドレスの数は、アクティブな IB インタフェースの数に一致している必要があります。
4. ibp1 と ibp2 を使用して、両方のポートをアクティブに設定し、2 番目の IPMP ネットワーク・インタフェース（ib-ipmp-controller2）を作成します。dNFS でパフォーマンスを最適化するために、このリンクで 2 つの別の IP アドレスを作成します。IP アドレスの数は、アクティブな IB インタフェースの数に一致している必要があります。この IPMP グループは、2 番目のコントローラが所有します。これは、2 番目のコントローラで BUI を使用して直接作成するか、またはテイクオーバー・モードで構成している場合は、最初のコントローラで作成し、フェイルバックする前に、BUI の「Configuration」・「Cluster」画面で所有権を変更できます。

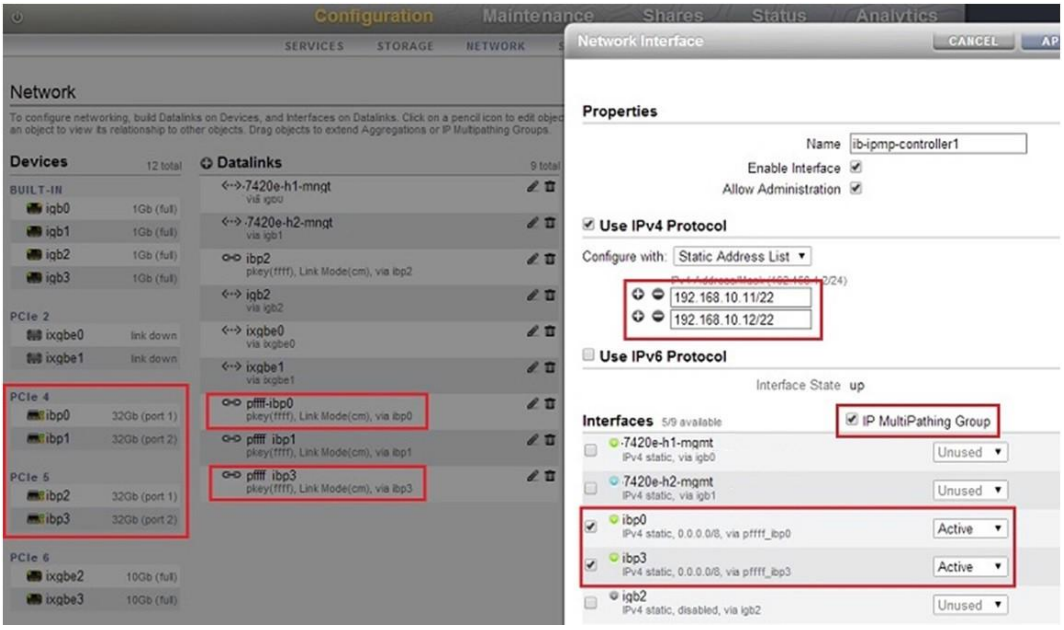

図 6：コントローラごとに 2 つの HCA を備えた IPMP 構成

Oracle ZFS Storage Appliance 上のアクティブ/アクティブ IPMP では、すべてのアクティブなリンクでトラフィックを処理するために、アクティブなリンクと同数の IP アドレスが必要になります。これを Recovery Manager のバックアップおよびリストア・アプリケーションで適用するには、oranfstab ファイルを使用して、複数のネットワーク・インタフェース間に分散されている dNFS 負荷を構成する必要があります。

各コントローラに 4 つの InfiniBand または 10GbE カードが取り付けられている場合、コントローラごとに 2 つの IPMP ネットワーク・インタフェースを構成することを推奨します。各 IPMP グループには、冗長性を最適化するために、複数のカード、PCI ブリッジ、およびネットワーク・スイッチ間に分散されている 2 つのアクティブなインタフェースが含まれています。2 つずつを 2 つのグループに分けると、4 つを 1 つのグループとした場合とは対照的に、オーバーヘッドが低減されると同時に、完全な HA 冗長性も維持されます。

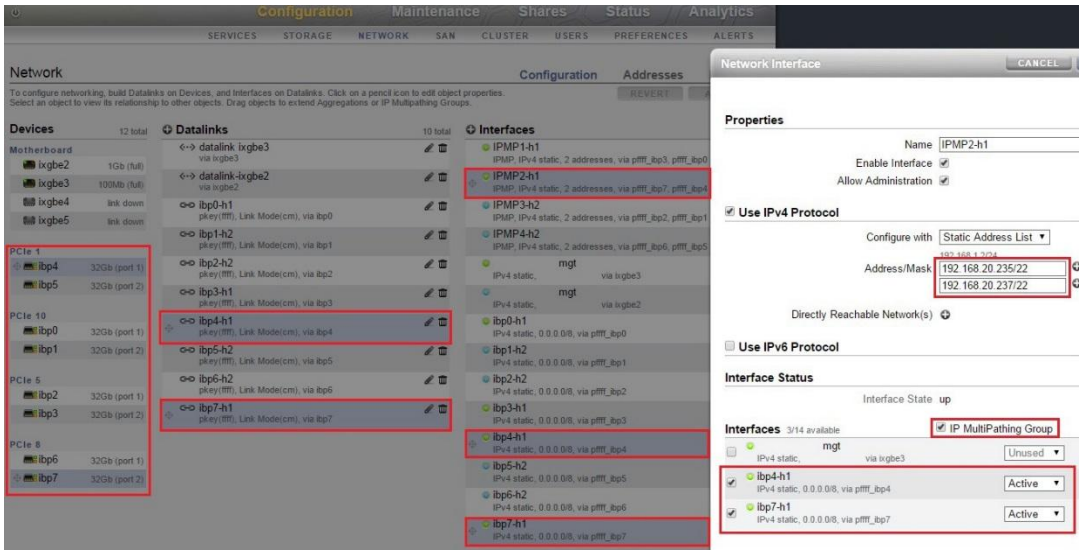

図 7：コントローラごとに 4 つの HCA を備えた IPMP 構成

IPMP フェイルオーバーのデフォルト設定では、10 秒でリンク障害を検出して、自動的にフェイルバックしますが、この設定は Exadata とともに使用するには最適ではありません。IPMP を Exadata および Oracle ZFS Storage Appliance とともに使用する場合、リンク障害検出をデフォルトの 10000ms（10 秒）から 5000ms（5 秒）に変更してください。これにより、Oracle ZFS Storage Appliance IPMP のフェイルオーバー構成が Exadata の構成に一致します。BUI または CLI からは、以下のようにして実行できます。

1. 「Configuration」→「Services」→「IPMP」を選択します。
2. Failover Detection Latency を 10000ms から 5000ms に変更して、適用します。

IPMP を使用する際には、マルチホーミング・ポリシーで適応型ルーティングを有効にして、Oracle ZFS Storage Appliance からのアウトバウンド・トラフィックが、複数のネットワーク・リンクおよび IP アドレスにわたって負荷分散されるようにします。BUI にアクセスし、「Configuration」→「Network」→「Routing」を選択してから、「multihoming=adaptive」ラジオ・ボタンを選択します。

Oracle Exadata Database Machineの構成

次のセクションでは、Exadata Database Machine を Oracle ZFS Storage Appliance とともに構成する際のガイダンスおよびオプションについて説明します。

NFSv3またはNFSv4の選択

NFSv3 と NFSv4 はともに、dNFS を Exadata および Oracle ZFS Storage Appliance とともに使用する際の優れたプロトコルです。NFSv4 では、強固なセキュリティ・モデル、コア・プロトコル内で管理されているファイル・ロック、およびクライアント側のキャッシュ精度を向上させるための委任など、いくつかの拡張機能が実装されます。

NFSv4 は、かなり大きなオーバーヘッドを受けます。ただし、dNFS ワークロードが大きいパケット・サイズを利用しているため、この環境ではパフォーマンスへの影響は無視できます。NFSv4 は、Oracle Database の 12c バージョンとともに使用して、Oracle Intelligent Storage Protocol を有効にするためには必須です。

ExadataでのNFSの有効化

NFSv4 だけを使用する場合、Exadata データベース・サーバーで共有を構成してマウントする前に、追加の手順は必要ありません。

NFSv3 または NFSv2 接続性をサポートする必要がある場合は、リモート・プロシージャ・コール（RPC）を追加し、プロトコルをマウントしてロックする必要があります。Oracle Linux 5（`cat /etc/redhat-release`）に基づいた Exadata バージョンで、これらのサービスを有効にするには、portmap、nfslock、および nfs サービスを開始し、リブート間で永続的に設定する必要があります。

以下の例では、Exadata の `dcli` コマンドを使用して、すべてのデータベース・サーバー上で portmap、nfslock、および nfs サービスを有効にします。

```
# dcli -l root -g /home/oracle/dbs_group chkconfig portmap on
# dcli -l root -g /home/oracle/dbs_group service portmap start
# dcli -l root -g /home/oracle/dbs_group chkconfig nfslock on
# dcli -l root -g /home/oracle/dbs_group service nfslock start
# dcli -l root -g /home/oracle/dbs_group chkconfig nfs on
# dcli -l root -g /home/oracle/dbs_group service nfs start
```

現在の Exadata バージョンが Oracle Linux 6（12.1.2.1.0 以上）に基づいている場合、portmap が rpcbind で置き換えられます。さらに、rpcbind は、`/etc/hosts.allow` と `/etc/hosts.deny` に対して、読取りアクセスできる必要があります。

以下の例では、Exadata の `dcli` コマンドを使用して、これらのファイルへの読取りアクセスを有効にし、すべてのデータベース・サーバー上で rpcbind、nfslock、および nfs サービスを有効にします。

```
# dcli -l root -g /home/oracle/dbs_group chmod 644 /etc/hosts.allow
# dcli -l root -g /home/oracle/dbs_group chmod 644 /etc/hosts.deny
# dcli -l root -g /home/oracle/dbs_group chkconfig rpcbind on
# dcli -l root -g /home/oracle/dbs_group service rpcbind start
```

```
# dcli -l root -g /home/oracle/dbs_group chkconfig nfslock on
# dcli -l root -g /home/oracle/dbs_group service nfslock start
# dcli -l root -g /home/oracle/dbs_group chkconfig nfs on
# dcli -l root -g /home/oracle/dbs_group service nfs start
```

NFSマウント・オプション

共有が従来型の Recovery Manager のバックアップ専用に設定されている場合は、以下のマウント・オプションを利用します。

```
rw,bg,hard,nointr,rsize=1048576,wsize=1048576,tcp,vers=3,timeo=600
```

共有が増分更新バックアップで利用されているか、または Recovery Manager のスイッチを利用して、緊急のデータベース・リカバリ用にコピーする可能性がある場合は、以下のマウント・オプションを利用します。

```
rw,bg,hard,nointr,rsize=32768,wsize=32768,tcp,actimeo=0,vers=3,timeo=600
```

Oracle ファイルのタイプと使用する NFS マウント・オプションの関係について詳しくは、My Oracle Support (MOS) Document 359515.1『Mount Options for Oracle files when used With NFS on NAS devices』を参照してください。

Oracle Direct NFS では、NFS マウント・オプションは利用されません。ただし、dNFS が使用できず、システムを NFS に戻す必要がある場合に、データベースの要件に準拠して、パフォーマンスと機能性を向上させるために、適切なマウント・オプションを設定することを推奨します。

また、Recovery Manager のバックアップおよびリストア・ジョブを実行中のデータベース・ノードには、バックアップ共有をマウントする必要があります。ただし、すべてのノードでバックアップまたはリストア操作を実行できるように、すべてのデータベース・サーバーに共有をマウントすることを推奨します。これは特に、データベース・インスタンスおよび Recovery Manager のバックアップ・サービスが、通常はアクティブではない他のデータベース・サーバーへ自動的に移行されることがある障害発生のシナリオで有効です。

Oracle Direct NFSの構成

Oracle Database 12c では、デフォルトで dNFS が有効にされています。Oracle Database 11g では、単一のデータベース・ノードで Oracle Direct NFS を有効にするには、次のコマンドを使用します。

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

すべてのデータベース・ノードで同時に dNFS を有効にするには、Exadata の dcli を使用します。

```
$ dcli -l oracle -g /home/oracle/dbs_group make -f $ORACLE_HOME/rdbms/lib/ins_rdbms.mk dnfs_on
```

Oracle Direct NFS を有効にした後、データベース・インスタンスを再起動する必要があります。

データベースの起動後、Oracle ODM メッセージのデータベース・アラート・ログをチェックして、dNFS が有効になっていることを確認します。

```
Oracle instance running with ODM:Oracle Direct NFS ODM Library Version 3.0
```

dNFS アクティビティは、SQL 問合せを使用して確認することもできます。

```
SQL> select * from v$dnfs_servers;
```

dNFS を無効にするには、次のコマンドを使用します。

```
$ make -f $ORACLE_HOME/rdbms/lib/ins_rdbms.mk dnfs_off
```

推奨されるパッチの詳細なリストについては、My Oracle Support（MOS）Document 1495104.1『Recommended Patches for Direct NFS Client』を参照してください。

---

oranfstabファイルの作成

公開されているバックアップおよびリストア速度を達成するには、oranfstab ファイルが必須です。ファイルは、$ORACLE_HOME/dbs/oranfstab に作成され、Oracle ホームを共有しているすべてのデータベース・インスタンスに適用されます。

oranfstab ファイルは、Oracle ZFS Storage Appliance 上の複数のアドレス（「path」で表示）または Exadata データベース・サーバー上の複数のアドレス（「local」で表示）を介して、dNFS 接続の負荷を分散するように構成します。複数のインタフェースにわたるロードバランシングにより、潜在的な 2 つのシステム・ボトルネック、つまりネットワーク・インタフェース帯域幅と TCP/IP バッファリングが低減されるか、または解消されます。

oranfstab で定義したパスの IP アドレスの数を、Oracle ZFS コントローラ上のアクティブなインタフェースの数と一致させることを推奨します。さらに、oranfstab で定義したローカル（local）IP アドレスの数を、データベース・サーバー上のアクティブなインタフェースの数と一致させることを推奨します。

Exadata Database Machine X3-2 または X2-2 のような以前の Exadata モデルでは、各コンピュート・ノードでデータパス（bondib0）に対してアクティブなインタフェースは 1 つのみです。

ただし、Exadata Database Machine X4-2 および X5-2 のような新しいモデルには、ib0 と ib1 の両方で定義した個別の IP アドレスにより、アクティブ/アクティブ・ロールで動作する機能があります。X3-8 および X4-8 のような一部のモデルでは、アクティブなデータパス・インタフェースは 4 つあります。これらのシナリオでは、oranfstab で複数のローカルの IP アドレスを定義して dNFS を有効にし、使用可能なパスをすべて利用する必要があります。

次に示すのは、2 つのストレージ・プールをホストする（1 つはヘッドごと、もう 1 つはプールごとに 1 つの共有）クラスタ化 Oracle ZFS Storage Appliance とともに使用する場合の oranfstab ファイルの例です。共有は、Oracle ZFS Storage Appliance で NFS エクスポート・パスにより定義し、ローカルのマウント・ポイントの所有者は oracle ユーザーです。また、共有は、関連のあるパ

ス・アドレスの下にリストされています。Exadata は X3-2 で、各 Oracle ZFS Storage コントローラは、2 つのアクティブなインタフェースにより構成されます。ローカル・アドレスは、Exadata コンピュート・ノードごとに固有であるため、別のコンピュート・ノードの oranfstab ファイルでは、別のローカル・アドレスが定義されます。これは、dNFS がデフォルトの NFSv3 を利用している例です。

```
server: zfs-storage-a
local: 192.168.10.1 path: 192.168.10.50
local: 192.168.10.1 path: 192.168.10.51
export: /export/backup1 mount: /zfs/backup1
server: zfs-storage-b
local: 192.168.10.1 path: 192.168.10.52
local: 192.168.10.1 path: 192.168.10.53
export: /export/backup2 mount: /zfs/backup2
```

次に示すのは同様の例で、Exadata Database Machine X5-2 を、アクティブ/アクティブの InfiniBand インタフェースで使用しています。ここでは、使用可能なすべてのパス間でワークロードが分散されるように、別のローカル・アドレスを定義しています。2 番目の「copy」という名前の共有はプールごとに作成され、dNFS クライアントに提示されます。この例では、サーバーごとに nfs_version パラメータを定義することにより、dNFS が NFSv4 を利用します。

```
server: zfs-storage-a
local: 192.168.10.1 path: 192.168.10.50
local: 192.168.10.2 path: 192.168.10.51
nfs_version: nfsv4
export: /export/backup1 mount: /zfs/backup1
export: /export/copy1 mount: /zfs/copy1
server: zfs-storage-b
local: 192.168.10.1 path: 192.168.10.52
local: 192.168.10.2 path: 192.168.10.53
nfs_version: nfsv4
export:/ export/backup2 mount: /zfs/backup2
export:/ export/copy2 mount: /zfs/copy2
```

次に示すのは、各コントローラで構成された 1 つのアクティブなネットワーク・インタフェースのみを使用して、Exadata Database Machine X4-8 をクラスタ化 Oracle ZFS Storage Appliance に接続する場合の oranfstab ファイルの例です。この oranfstab 構成は、各コントローラにおける単一のアクティブなパス・アドレスへの Exadata コンピュート・ノードで、4 つのアクティブなインタフェースすべてに負荷を分散します。

```
server: zfs-storage-a
local: 192.168.10.1 path: 192.168.10.50
local: 192.168.10.2 path: 192.168.10.50
local: 192.168.10.3 path: 192.168.10.50
local: 192.168.10.4 path: 192.168.10.50
export: /export/backup1 mount: /zfs/backup1
server: zfs-storage-b
local: 192.168.10.5 path: 192.168.10.51
local: 192.168.10.6 path: 192.168.10.51
local: 192.168.10.7 path: 192.168.10.51
local: 192.168.10.8 path: 192.168.10.51
export: /export/backup2 mount: /zfs/backup2
```

最後に示すのは、各コントローラで構成された 4 つのアクティブなネットワーク・インタフェースを使用して、Exadata X4-8 をクラスタ化 Oracle ZFS Storage Appliance に接続する場合の oranfstab ファイルの例です。

```
server: zfs-storage-a
local: 192.168.10.1 path: 192.168.10.50
local: 192.168.10.2 path: 192.168.10.51
local: 192.168.10.3 path: 192.168.10.52
local: 192.168.10.4 path: 192.168.10.53
export: /export/backup1 mount: /zfs/backup1
export: /export/stage1 mount: /zfs/stage1
server: zfs-storage-b
local: 192.168.10.5 path:192.168.10.54
local: 192.168.10.6 path:192.168.10.55
local: 192.168.10.7 path:192.168.10.56
local: 192.168.10.8 path:192.168.10.57
export: /export/backup2 mount: /zfs/backup2
export: /export/stage2 mount: /zfs/stage2
```

`oranfstab` ファイルでローカル・アドレスが指定されていない場合、dNFS は検出した最初のルーティング可能なアドレス/インタフェースを検出して利用します。

このセクションの例は、Exadata のネイティブの InfiniBand インフラストラクチャに対して相乗効果を実現するように接続されているため、InfiniBand 向けに調整されていますが、`oranfstab` の構文には依存性がなく、10GbE にも同様に適用されます。`oranfstab` ファイルで使用可能なオプションについて詳しくは、特定のバージョンの Oracle Database ソフトウェア用の管理ガイドを参照してください。

Recovery Managerのバックアップ・サービスの構成

Recovery Manager のバックアップ・サービスは、すべての Exadata コンピュート・ノード間で Recovery Manager のワークロードのバランスを取るように、作成および使用する必要があります。複数の Oracle RAC ノード間でバックアップを分散させることにより、パフォーマンスが向上し、並列タスクが増加して、単一のコンポーネントでの利用率の負荷が低減します。Recovery Manager バックアップ・サービスは、優先インスタンスが使用できないときには、Oracle RAC 内の他のデータベース・サーバーに自動的に移行されます。

以下の構成手順では、4 つの Oracle RAC ノードを備えた Exadata Database Machine X5-2 ハーフラックを仮定しています。「Hulk」は、データベース名です。

構文は、以下のとおりです。`srvctl add service -d <db_name> -r <preferred instance> -a <alternate instance(s)> -s <name for newly created service>`

```
[oracle@ex01db01 ~]$ srvctl add service -d hulk -r hulk1 -a hulk2,hulk3,hulk4 -s hulk_bkup1
[oracle@ex01db01 ~]$ srvctl add service -d hulk -r hulk2 -a hulk1,hulk3,hulk4 -s hulk_bkup2
[oracle@ex01db01 ~]$ srvctl add service -d hulk -r hulk3 -a hulk1,hulk2,hulk4 -s hulk_bkup3
[oracle@ex01db01 ~]$ srvctl add service -d hulk -r hulk4 -a hulk1,hulk2,hulk3 -s hulk_bkup4

[oracle@ex01db01 ~]$ srvctl start service -d hulk -s hulk_bkup1
[oracle@ex01db01 ~]$ srvctl start service -d hulk -s hulk_bkup2
[oracle@ex01db01 ~]$ srvctl start service -d hulk -s hulk_bkup3
[oracle@ex01db01 ~]$ srvctl start service -d hulk -s hulk_bkup4

[oracle@ex01db01 ~]$ srvctl status service -d hulk
Service hulk_bkup1 is running on instance(s) hulk1
Service hulk_bkup2 is running on instance(s) hulk2
Service hulk_bkup3 is running on instance(s) hulk3
Service hulk_bkup4 is running on instance(s) hulk4
```

データベースの再起動時に、Recovery Manager のバックアップ・サービスをリバランスする必要があります。これは、以下のようにして実行します。

```
[oracle@ex01db01 ~]$ srvctl stop service -d hulk; srvctl start service -d hulk
```

InfiniBandのベスト・プラクティス

最適な IPoIB パフォーマンスと安定性を実現するために、Exadata コンピュート・ノードでいくつかの変更が必要になる場合があります。

1. スキャッタ・ギャザー機能を実装します。

- InfiniBand バッファリングに対して、不連続なメモリ割当てを有効にします。これにより、IPoIB スタック内でのレイテンシが短くなり、Linux のメモリ断片化が原因のページ割当て障害が発生することがなくなります。
- Exadata Storage Server Software 12.1.2.1.0 以降および Oracle ZFS Storage Appliance OS 8.2.7（2013.1.2.7）以降の両方を使用する際には、デフォルトで有効化されており、使用可能です。
- Exadata ソフトウェアのバージョン 12.1.2.1.0 は、Database の 12c および 11g の両方のバージョンをサポートしています。ただし、関連のないセキュリティ拡張機能が原因で、推奨される最小バージョンの 12.1.2.1.1 で置き換えられました。
- Exadata データベース・サーバーで、スキャッタ・ギャザー機能が使用できるかどうかを確認します。

  ```
  $ cat /sys/module/ib_ipoib/parameters/cm_ibcrc_as_csum
  ```

  1 = 端末装置のサポートが有効、0 = 無効

2. Exadata データベース・サーバーの MTU を、接続モード設定の 65520 に戻します。
   - これを 7000 の設定と比較した場合、処理する TCP パケット数が 8 倍も少なくなっており、r レイテンシとキューの混雑が低減されます。それと同時に、IPoIB ワークロードのパフォーマンスと安定性も向上しています。

3. Exadata データベース・サーバーが、サイズ 2048 の IPoIB 受信キューを使用していることを確認します。
   - 以前の Exadata バージョンで使用されていたデフォルト設定は 512 であり、この場合、IPoIB 受信キューがいっぱいになり、再試行遅延が必要な Receiver Not Ready（RNR）メッセージの形式で、IPoIB スタックへの長いレイテンシを発生させてしまうことがあります。
   - キュー・サイズを 2048 に増やすと、RNR による再試行遅延が解消され、IPoIB が順序付けしたキューの混雑を回避する際に重要な影響を与えます。
   - IPoIB 受信キュー・サイズを確認するには、次を実行します。

     ```
     $ cat /sys/module/ib_ipoib/parameters/recv_queue_size
     ```

バックアップ用のデータベースの準備

バックアップ用のデータベースの準備する際には、以下の構成を推奨します。

ARCHIVELOGモード

データベースが ARCHIVELOG モードで動作するように構成されている場合、オンライン REDO ログのアーカイブは有効です。ARCHIVELOG モードを使用するメリットは、以下のとおりです。

- メディア障害発生時の保護
- 最新のバックアップ後に発生した、データベース・トランザクションのリカバリ機能

- データベースが開かれていてアクティブな場合のバックアップの実行
- データベースをリストアするために、非一貫性バックアップを使用可能

データベースを ARCHIVELOG モードで実行し、Exadata ストレージ上に 1 つのコピー、および Oracle ZFS Storage Appliance 上に 1 つのコピーを配置して、アーカイブ・ログを多重化することを推奨します。

ブロック・チェンジ・トラッキング

ブロック・チェンジ・トラッキングは、データファイル内で変更されたブロックを記録する Recovery Manager の機能です。レベル 0 のバックアップは、データファイル全体をスキャンしますが、その後の増分バックアップは、ブロック・チェンジ・トラッキング・ファイルを基にして、最後のバックアップ以降に変更されたとマークされているブロックのみをスキャンします。

増分バックアップのパフォーマンスを向上させるために、ブロック・チェンジ・トラッキングを有効にすることを推奨します。選択したバックアップ戦略に含まれているのが、全体バックアップまたはレベル 0 のバックアップのみである場合、ブロック・チェンジ・トラッキングを有効にしないでください。

従来型のRecovery Managerのバックアップ向けのベスト・プラクティス

このセクションでは、従来型の Recovery Manager のバックアップ戦略を使用して、Oracle ZFS Storage Appliance により Oracle Exadata Database Machine を保護する際に、最適なパフォーマンスと機能性を達成するために必要な推奨される構成手順について説明します。このセクションで示されているベスト・プラクティスでは、特に従来型の Recovery Manager のバックアップ戦略を取り扱いますが、これまでのセクションに含まれている推奨事項は、一般的なアプリケーションにも該当します。

ネットワーク共有の構成

従来型の Recovery Manager の戦略を使用してデータベースをバックアップする際には、ストレージ・プールごとに単一の共有を設定することを推奨します。通常、ストレージ・プールは、最大限のパフォーマンスを実現するために、コントローラごとに構成され、両方のコントローラでハードウェア・リソースを全面的に利用します。Oracle データベースは、コントローラごとに 1 つを所有する、2 つの共有を使用してバックアップできます。

または、複数のデータベースを同時にバックアップするか、あるいは複数の Oracle Exadata Database Machine をバックアップする際には、別のコントローラを効率的に利用している、他の Recovery Manager のバックアップ・ワークロードにより、1 つのコントローラが所有する単一の共有のみを使用するように、個別の Recovery Manager のバックアップ操作を構成できます。完全な HA 冗長性は、他のコントローラに対してストレージ・リソースをフェイルオーバーする機能とともに、この構成で提供されています。

共有のアクセス権限では、oracle ユーザーのユーザーID と dba グループのグループ ID が一致するように調整する必要があります。標準的な Exadata 構成では、それぞれ 1001 と 1002 です。大部分の構成では、ディレクトリ権限は rwxr-x---に設定されています。

レコード・サイズ

ZFS レコード・サイズは、共有レベルで構成される設定で、バックエンド・ディスク I/O のサイズに影響を与えます。最適な設定は、このケースの Recovery Manager でアプリケーションが使用する、ネットワーク I/O のサイズに応じて異なります。Oracle Direct NFS による従来型の Recovery Manager のワークロードでは、ネットワーク層において、大規模な 1M の書込みおよび読取りが生成されます。このケースでは、1M レコード・サイズ設定を使用する必要があります。大規模なレコード・サイズを使用する機能には、スループット・パフォーマンスの向上などの大きなメリットがあり、これは帯域幅集中型のワークロードにとって重要です。その他、コントローラ CPU リソースの利用率が低下するというメリットもあります。

最近では、HDD 容量が今までにないほど急速に増加していますが、これらのディスクが提供できる IOPS は横ばい状態のままです。Recovery Manager のワークロードは、大抵の場合、少ない読取り頻度のみで、TB スケールのデータセットを生成します。このように、IOPS の処理では、キャッシュは最適なソリューションではありません。スループットを最大限にして、ディスクへの IOPS を制限することは、バックアップ・ソリューションで最適なパフォーマンスを達成するために重要な要因です。従来型の Recovery Manager のバックアップ戦略でこれを実行するには、ZFS 共有の大きいレコード・サイズで非常にメリットのある、大規模なマルチチャネル・ネットワーク I/O を用意します。

同期書込みバイアス

同期書込みバイアスは、同期書込みのサービス提供において、動作を制御する共有設定です。レイテンシまたはスループットに対して最適化できます。

最初のうち、すべての書込みは、非同期、同期、レイテンシの最適化済み、またはスループット最適化済みであるかどうかに関係なく、ZFSadaptive replacement cache（ARC）に書き込まれます。さらに、すべての書込みは、ARC からストレージ・プールにコピーされます。非同期書込みは、ARC への書込みの完了後、クライアントに対して確認応答を返します。同期書込みをスループットに対して最適化すると、書込みがストレージ・プールにコピーされてはじめて、確認応答が返されます。また、レイテンシに対して最適化すると、確認応答が即座にクライアントに返されるように、別のコピーが永続ストレージに書き込まれます。ストレージ・プールで書込み最適化フラッシュが構成されている場合、レイテンシ重視の同期書込みにおいて、永続ストレージとして使用されます。

同期書込みバイアス共有設定は、スループットに対して構成する必要があります。これは、以下の 2 つの理由によります。

1. 書込み最適化フラッシュは限定リソースで、HDD と比較してはるかに高コストです。これは、レイテンシ重視の書込みにとっては大きな支援となりますが、従来型の Recovery Manager のバックアップは、帯域幅重視の大規模 1M ストリーミング書込みです。ストレージ・プール構成にフラッシュが存在しない場合、フラッシュを追加せずに、低価格が達成できます。HDD は、フラッシュが占有し、プロセスでスループットと容量を追加するようなスロットをふさぐのに使用できます。ストレージ・プール内で、書込み最適化フラッシュが構成されている場合、これは他のネットワーク共有や LUN がアクセス可能な共有リソースとなります、また、メリットを提供できるように、レイテンシ重視のワークロードに対して予約しておく必要があります。
2. 同期書込みバイアスをレイテンシに設定すると、別のデータ転送が生成され、帯域幅重視のワークロードの処理でパフォーマンスが低下します。レイテンシに対して最適化すると、ARC から別のコピーが読み取られ、フラッシュに書き込まれます。ログ・デバイスをミラー化すると、別の 2 つのコピーがフラッシュに書き込まれます。書込み最適化フラッシュ・デバイスは、多数の IOPS をサービス提供するように設計されていますが、高スループットのワークロードにより、帯域幅が簡単に飽和することがあります。ストレージ・プール内で構成されている多数のフラッシュ・デバイスがアイドル状態で、十分な量のフラッシュ帯域幅が使用できる場合でも、SAS 帯域幅は制限要因になります。ミラー化ストレージ・プール、ミラー化ログ・デバイスが存在して、同期書込みバイアスがレイテンシに設定された構成では、最初に同期書込みが ARC に書き込まれてから、4 つの追加コピーが書き込まれます。SAS 帯域幅は、ネットワーク帯域幅の 4 倍以上です。スループット最適化書込みでは、帯域幅重視のワークロードで良好なパフォーマンスが実現されます。

読取りキャッシュ

読取り最適化フラッシュは、従来型の Recovery Manager のワークロードのキャッシュには使用しないでください。Recovery Manager のバックアップ・セットをキャッシュに格納することには、ほとんどメリットがないためです。さらに、レベル 2 ARC は、ストリーミング・ワークロード向けではありません。キャッシュ・デバイスの使用共有設定では、キャッシュ・デバイスを使用しないように構成する必要があります。

データ圧縮

データ圧縮設定により、共有レベルで、Oracle ZFS Storage Appliance が圧縮アルゴリズムを適用するかどうかが決定されます。圧縮は、データベース・レベルでもっとも効率的ですが、ベスト・プラクティスは、読取り集中型のトランザクション・ワークロードでは、Hybrid Columnar Compression（HCC）を使用し、書込み集中型のトランザクション・ワークロードでは、Advanced Row Compression を使用することです。

従来型の Recovery Manager のワークロードについては、共有レベルで LZJB 圧縮を有効にする必要があります。また、データベース圧縮と組み合わせて、バックエンド・ディスクへの帯域幅を低減することにより、CPU 利用率への影響が最低限になるという別のメリットがあります。物理ネットワーク・スループットは、LZJB を使用すると実際に向上します。これは、従来型の Recovery Manager のワークロードでは通常、SAS 帯域幅と HDD 利用率が制限要因であるためです。OLTP データベースで Advanced Row Compression を有効にすると、LZJB では、1.4～2.1 倍の範囲で領域が節減されることがあります。gzip ベースの圧縮アルゴリズムは、高パフォーマンスのワークロードでは、CPU オーバーヘッドで高コストになるため、このコンテキストでは検討する必要はありません。

図 8 は、このセクションで説明した設定を使用して作成されたバックアップ共有を示します。

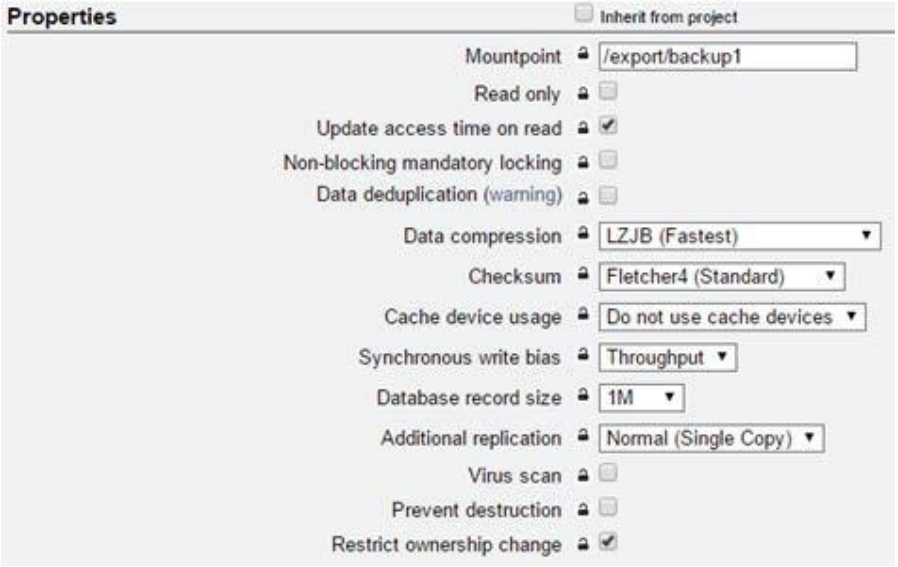

図8：従来型のRecovery Managerのバックアップの共有の例

Recovery Managerの構成

従来型の Recovery Manager のバックアップ構成では、以下の要因について検討する必要があります。

バックアップ形式

レベル 0 のバックアップの形式として、バックアップ・セットまたはイメージ・コピーのいずれも使用できますが、バックアップ・セットの使用を推奨します。データ保護戦略の選択時、セクションに強調表示されている多数のメリットを達成するには、バックアップ・セット形式が必須です。

チャネルの最適化

使用する Recovery Manager チャネルの数を決定することは、バックアップ・ソリューションのチューニングにおいて重要な側面です。Recovery Manager が新しいチャネルを開くと、入力および出力バッファの新しいセットが割り当てられます。各チャネルには、データファイルまたはデータファイルのセクションを取得して、バックアップまたはリストア・ジョブを並列に処理し、他のチャネルに作業させる機能が備えられています。チャネルは、Oracle RAC 内の複数のノードに割り当てることができ、Oracle ZFS Storage Appliance の別のコントローラが潜在的に所有している共有により、別のバックアップ先を指定できます。

チャネルを追加すると、スケーラビリティが向上するため、パフォーマンスが大幅に向上し、リソースの利用率が効率化されます。さらに、データベース・ノード間でロードバランシングが実行され、HA アーキテクチャの堅牢性が増大して、ストレージ・コントローラ間にワークロードが分散されます。

ハードウェアの限度に近づくと、追加の Recovery Manager チャネルが割り当てられ、収穫逓減点が指定されます。Recovery Manager のバッファに追加のメモリや CPU リソースが割り当てられても、パフォーマンスはほとんど向上せず、バックアップ・ピースが作成されて複雑さが増大するだけであるため、チャネルを過度に割り当てることは推奨されません。

特定の構成で推奨されるチャネルの数は、最適に構成されているソリューションで全体のパフォーマンスを制限するハードウェア要因に応じて決定されます。パフォーマンスを制限するコンポーネントは多数存在し、Exadata、ネットワーク、HDD、CPU または SAS リソースなどがその中に含まれています。本番環境に大きな変更を実装した際には、必ず徹底的なテストを実施することを推奨します。ただし、以下のマトリックスは、ハードウェア構成に応じて、従来型の Recovery Manager のバックアップ戦略で構成する Recovery Manager チャネルの数に関するガイダンスを示しています。この表では、Exadata Database Machine X5-2 および Oracle ZFS Storage ZS4-4 の両方のコントローラ間で、ストレージのバランスが取れていると仮定しています。また、ネットワークと SAS 帯域幅は制限要因ではなく、このホワイト・ペーパーでのベスト・プラクティスが実装されており、バックアップ時に他の重要な同時ワークロードは存在していないと仮定しています。

表4：従来型のRecovery Managerのバックアップ戦略 – 構成ごとに推奨されるRecovery Managerチャネルの数

	1/8 ラック	1/4 ラック	ハーフラック	フルラック
1 ディスク・シェルフ	8	8	8	8
2 ディスク・シェルフ	8	12	12	12
3～4 ディスク・シェルフ	8	16	16	16
5～6 ディスク・シェルフ	8	16	24	24
7～8 ディスク・シェルフ	8	16	32	32
9 以上のディスク・シェルフ	8	16	32	40

Recovery Manager チャネルを構成または割り当てると、Oracle RAC ノードとストレージ共有間で交互に指定されます。

バッファのチューニング

各チャネルには、特定の数の入力バッファと出力バッファが割り当てられています。バックアップ時には、これらのメモリ・バッファを使用して、データファイルのセクションの読取り、出力バッファへのコピー、およびバックアップ共有上のバックアップ・セット・ピースへの書込みが実行されます。

ASM ディスク・グループから実行されている Recovery Manager のジョブには、動的にチューニングされるバッファが用意されています。ただし、最適なリストア・スループットを達成するには、以下の設定を使用する必要があります。ただし、バックアップ・スループットは変更されません。

```
_backup_disk_bufcnt=20

_backup_disk_bufsz=2097152
```

セクションのサイズ

高度に並列化された Recovery Manager のワークロードを有効にすることは、バックアップ・ソリューションで最適なパフォーマンスとリソース利用率を達成するために重要です。課題は、非常に大規模なデータファイルが発生した場合です。単一の Recovery Manager チャネルで処理すると、スループットが大幅に低下し、環境内の他のハードウェア・リソースが、異常値データファイルの処理が完了するのを待機して、アイドル状態のままになります。

Recovery Manager は、この問題に対するソリューションを提供していますが、それは大規模なファイルを小さいピースに分割して、複数のチャネルで並列に処理するという機能です。これは、マルチセクション・サポートと呼ばれ、SECTION SIZE パラメータによって決定されます。SECTION SIZE を 100G に設定することを推奨します。

FILESPERSET

FILESPERSET パラメータでは、各バックアップ・セット内に含まれる、データファイルまたはデータファイルのセクションの数を決定します。単一のバックアップ・セットを作成するために、複数の入力ファイルを読み取っている場合、特に読取りまたはコピー・フェーズが制限要因であるときには、パフォーマンスが向上する可能性があります。デフォルトは 64 です。ただし、この値は、小規模なセクションしか使用されていない場合でも、バックアップ・セット全体を読み取る必要があるような、単一ファイルまたは部分データベース・リストアでは悪影響を及ぼします。さらに、FILESPERSET を過度に大きく設定すると、Recovery Manager のロードバランシングおよびパフォーマンス・スケーリング・プロパティに影響を与える場合があります。目標は、バックアップ全体で、すべての Recovery Manager チャネルが効率的に利用されることです。データファイルまたはデータ・セクションの数に制限がある場合、すべてのチャネルで全体バックアップ・セットが作成できない可能性があります。

一般的な手法として、FILESPERSET を 3 に設定することを推奨します。テストにより示されたことは、このパラメータは部分データベース・リストアの処理、すべてのチャネル間でのロードバランシング、および各バックアップ・セットを供給する複数の入力ファイルを使用したバックアップ・パフォーマンスの最適化において、適切なバランスを備えているということです。

runブロックの例

次に示すのは、毎週のレベル 0 バックアップに対する run ブロックの例で、増分バックアップ戦略の一部として含めることができます。この例では、Exadata Database Machine X5-2 ハーフラックによる、4 つのディスク・シェルフで構成された Oracle ZFS Storage Appliance の両方のコントローラへのバックアップを仮定しています。Recovery Manager バックアップ・サービスを使用し、4 つの RAC ノードすべてに対して、チャネルを均一に分散させます。チャネルは、コントローラごとに 1 つを所有する、2 つのストレージ共有間で交互に指定されます。

```
run
{
sql ' alter system set "_backup_disk_bufcnt"=20 scope=memory' ;
sql ' alter system set "_backup_disk_bufsz"=2097152 scope=memory' ;
allocate channel ch1 device type disk connect 'sys/passwd@exa-scan/bkup_db1' format  '/zfs/bkup1/%U';
allocate channel ch2 device type disk connect 'sys/passwd@exa-scan/bkup_db2' format '/zfs/bkup2/%U';
allocate channel ch3 device type disk connect 'sys/passwd@exa-scan/bkup_db3' format '/zfs/bkup1/%U';
allocate channel ch4 device type disk connect 'sys/passwd@exa-scan/bkup_db4' format '/zfs/bkup2/%U';
allocate channel ch5 device type disk connect 'sys/passwd@exa-scan/bkup_db1' format '/zfs/bkup1/%U';
allocate channel ch6 device type disk connect 'sys/passwd@exa-scan/bkup_db2' format '/zfs/bkup2/%U';
allocate channel ch7 device type disk connect 'sys/passwd@exa-scan/bkup_db3' format '/zfs/bkup1/%U';
allocate channel ch8 device type disk connect 'sys/passwd@exa-scan/bkup_db4' format '/zfs/bkup2/%U';
allocate channel ch9 device type disk connect 'sys/passwd@exa-scan/bkup_db1' format '/zfs/bkup1/%U';
allocate channel ch10 device type disk connect 'sys/passwd@exa-scan/bkup_db2' format '/zfs/bkup2/%U';
allocate channel ch11 device type disk connect 'sys/passwd@exa-scan/bkup_db3' format '/zfs/bkup1/%U';
allocate channel ch12 device type disk connect 'sys/passwd@exa-scan/bkup_db4' format '/zfs/bkup2/%U';
allocate channel ch13 device type disk connect 'sys/passwd@exa-scan/bkup_db1' format '/zfs/bkup1/%U';
allocate channel ch14 device type disk connect 'sys/passwd@exa-scan/bkup_db2' format '/zfs/bkup2/%U';
allocate channel ch15 device type disk connect 'sys/passwd@exa-scan/bkup_db3' format '/zfs/bkup1/%U';
allocate channel ch16 device type disk connect 'sys/passwd@exa-scan/bkup_db4' format '/zfs/bkup2/%U';
configure snapshot controlfile name to ' /zfs/bkup1/snapcf_dbname.f' ;
BACKUP AS BACKUPSET
  SECTION SIZE 100G
  INCREMENTAL LEVEL 0
  DATABASE
  FILESPERSET 3
  TAG ' BKUP_SUNDAY_L0' ;
}
```

増分更新バックアップのベスト・プラクティス

このセクションでは、Recovery Manager の増分更新バックアップ戦略を利用して、Oracle ZFS Storage Appliance により Oracle Exadata Database Machine を保護する際に、最適なパフォーマンス、信頼性、およびユーザビリティを達成するために必要な推奨される構成手順について説明します。このセクションで示されているベスト・プラクティスでは、特に増分更新バックアップ戦略を取り扱います。これは、以前のセクションで十分に説明した、テクノロジーの詳細および一般的なベスト・プラクティスに対する補足の意図があります。

Oracle ZFS Storage Applianceの構成

増分更新バックアップ戦略では、ミラー化ストレージ・プロファイルと SSD 書込みフラッシュ・アクセラレータを使用することを推奨します。前の増分レベル 1 のバックアップ・セットがイメージ・コピーにマージされるプロセスでは、OLTP ワークロードを実行中のデータベースに対して、大量の小規模ブロック・ランダム I/O が作成されます。IOPS 集約型ワークロードの処理には、ミラー化ストレージ・プールが最適です。バックアップ・コピーへの小規模の I/O 更新は、レイテンシ重視であり、最適なパフォーマンスを実現するためには、ストレージ・プールに書込み最適化フラッシュが存在する必要があります。

共有の作成

増分更新バックアップ戦略を使用してデータベースをバックアップする際には、2 つの共有を利用することを推奨します。初期のイメージ・コピー・バックアップは、1 つの共有に配置し、もう 1 つには毎日の増分を配置する必要があります。

共有のアクセス権限では、oracle ユーザーのユーザーID と dba グループのグループ ID が一致するように調整する必要があります。標準的な Exadata 構成では、それぞれ 1001 と 1002 です。大部分の構成では、ディレクトリ権限は rwxr-x---に設定されています。

レコード・サイズ

毎日の増分共有を含む読取りと書込みでは、大規模な 1M の I/O が生成されます。この共有の ZFS レコード・サイズを 1M に構成することを推奨します。

バックアップ・コピー・ファイルへの更新により、小規模の書込みが生成されます。これは、バックアップ・データファイル内において、アクティブなデータファイルで変更されたブロックのみが更新されるためです。コピー共有での最適な ZFS レコード・サイズは、ソース・データベースに対するトランザクション・ワークロードの特性に応じて決定されます。

ソース・データベースが、オンライン・トランザクション処理（OLTP）ワークロードを実行している場合、ZFS レコード・サイズは、32K に設定する必要があります。OLTP ワークロードの特徴は、大抵の場合、データの既存の行の変更または読取りに注目した、小規模のトランザクションです。これにより、比較的小規模な書込み I/O が生成され、これはデータファイル全体に分散できます。ZFS レコード・サイズを 32K に設定することにより、読取り-修正-書込みのオーバーヘッドが最小化されると同時に、初期のバックアップ、その後のリストア、またはリストア検証操作などのストリーミング・ワークロードにおいて、適切なパフォーマンスが実現されます。

ソース・データベースが、オンライン・アプリケーション処理（OLAP）ワークロードを実行している場合、ZFS レコード・サイズは、128K に設定する必要があります。OLAP ワークロードの特徴は、バックアップ・コピーの更新時に、大規模な書込み I/O を生成する、大量の問合せおよびバッチ追加です。

同期書込みバイアス

同期書込みバイアスは、同期書込みのサービス提供において、動作を制御する共有設定です。レイテンシまたはスループットに対して最適化できます。

同期書込みバイアスは、毎日の増分共有のスループット、およびコピー共有のレイテンシに対して構成する必要があります。バックアップ・コピーに変更を適用することは、レイテンシ重視のワークロードで、I/O を書込み最適化フラッシュにコピーする必要があります。これにより、クライアントに対して即座に確認応答が返されます。また、共有の IOPS 機能、およびソリューション全体のパフォーマンスが大幅に向上します。

読取りキャッシュ

読取り最適化フラッシュは必須ではありませんが、増分更新バックアップ戦略の更新フェーズ時のパフォーマンスを最適化することを推奨します。コピー共有に関するキャッシュ・デバイスの使用共有設定は、すべてのデータおよびメタデータで構成する必要があります。

データ圧縮

毎日の増分共有では、LZJB 圧縮のみの使用を推奨します。コピー共有では、ZFS 共有レベルの圧縮を有効にしないでください。

Recovery Manager環境の構成

従来型の Recovery Manager のバックアップ構成では、以下の要因について検討する必要があります。

バックアップ形式

増分更新バックアップ戦略では、レベル 0 のバックアップをイメージ・コピー形式にする必要があり、また、毎日の増分は常にバックアップ・セット形式にします。他の設定は使用できません。

チャネルの最適化

イメージ・コピー操作に割り当てられたチャネルは、バックアップ・セットの操作に割り当てられたチャネルと比較して、チャネルごとのスループット容量が低くなります。これは、入力および出力関数で、共有バッファが使用されるためです。Recovery Manager は、共有バッファがビジー・ステータスをクリアしない待機状態では、さらに時間を消費します。これは、環境内のバックアップ・ソリューションのパフォーマンス全体には影響を与えませんが、複数のハードウェア・プラットフォーム間に分散された多数のチャネルを使用するように適切にスケーリングできます。ただし、バックアップ・セットの操作として同量のスループットを達成するには、チャネルを追加することが必要になる場合があります。

バッファのチューニング

各チャネルには、特定の数の入力バッファと出力バッファが割り当てられています。バックアップ時には、これらのメモリ・バッファを使用して、データファイルのセクションの読取り、出力バッファへのコピー、およびバックアップ共有上のバックアップ・セット・ピースへの書込みが実行されます。

増分更新バックアップで最適なスループットを達成するには、以下の設定を使用する必要があります。

```
_backup_disk_bufcnt=64
_backup_disk_bufsz=1048576
_backup_file_bufcnt=64
_backup_file_bufsz=1048576
```

セクションのサイズ

イメージ・コピーのマルチセクション・サポートは、Oracle Database 12c で挿入されました。以前のバージョンのデータベース・ソフトウェアを実行している場合、イメージ・コピー形式を使用して bigfile 表領域をバックアップすることは推奨されません。これは、Recovery Manager のパフォーマンス・スケーリングおよびロードバランシングが保証できないためです。これは、標準の表領域では問題にはなりません。複数のデータファイルにより、すべての Recovery Manager チャネル間での最適なパフォーマンス・スケーリングおよびロードバランシングが保証されるためです。

RUNブロックの例

次に示すのは、データベースのバックアップが実行されるたびに、繰り返し実行されるように設計された RUN ブロックの例です。これは 2 つの共有と増分更新バックアップ戦略を使用して、各ファイルのレベル 0 のイメージ・コピーを bkup1 に配置し、その後のレベル 1 の増分バックアップを bkup2 に配置します。毎晩、同じ RUN ブロックが実行されますが、前のレベル 0 およびレベル 1 のバックアップが存在するかどうかに応じて、別のタスクが実行されます。

最初の夜の実行時、データベース内のすべてのデータファイルに対して、レベル 0 イメージ・コピー・バックアップが作成されます。2 日目の夜、最後の 24 時間以内に変更されたデータのみのバックアップ・セットが作成され、別の共有に格納されます。3 日目の夜、最後の 24 時間以内に変更されたデータのみのバックアップ・セットのみが作成され、次にバックアップ・コピーに対して、前夜に変更されたデータが適用されます。以降は夜ごとにバックアップが実行されますが、最後の 24 時間以内に変更されたデータのみが同じ機能でバックアップされ、バックアップ・コピーに対して、前に変更されたデータが適用されます。Recovery Manager の tag コマンドを使用して、これを追跡します。

この例では、4 つのデータベース・サーバーを備えた、ハーフラック Exadata を仮定しています。接続文字列内で Recovery Manager のバックアップ・サービスを使用して、Oracle RAC 内で使用可能なすべてのノード間にチャネルを分散します。

```
RUN
{
sql ' alter system set "_backup_disk_bufcnt"=64 scope=memory' ;
sql ' alter system set "_backup_disk_bufsz"=1048576 scope=memory' ;
sql ' alter system set "_backup_file_bufcnt"=64 scope=memory' ;
sql ' alter system set "_backup_file_bufsz"=1048576 scope=memory' ;
allocate channel ch1 device type disk connect 'sys/passwd@exa-scan/bkup_db1' format  '/zfs/bkup1/%U';
allocate channel ch2 device type disk connect 'sys/passwd@exa-scan/bkup_db2' format '/zfs/bkup1/%U';
allocate channel ch3 device type disk connect 'sys/passwd@exa-scan/bkup_db3' format '/zfs/bkup1/%U';
allocate channel ch4 device type disk connect 'sys/passwd@exa-scan/bkup_db4' format '/zfs/bkup1/%U';
allocate channel ch5 device type disk connect 'sys/passwd@exa-scan/bkup_db1' format '/zfs/bkup1/%U';
allocate channel ch6 device type disk connect 'sys/passwd@exa-scan/bkup_db2' format '/zfs/bkup1/%U';
allocate channel ch7 device type disk connect 'sys/passwd@exa-scan/bkup_db3' format '/zfs/bkup1/%U';
allocate channel ch8 device type disk connect 'sys/passwd@exa-scan/bkup_db4' format '/zfs/bkup1/%U';
allocate channel ch9 device type disk connect 'sys/passwd@exa-scan/bkup_db1' format '/zfs/bkup1/%U';
allocate channel ch10 device type disk connect 'sys/passwd@exa-scan/bkup_db2' format '/zfs/bkup1/%U';
allocate channel ch11 device type disk connect 'sys/passwd@exa-scan/bkup_db3' format '/zfs/bkup1/%U';
allocate channel ch12 device type disk connect 'sys/passwd@exa-scan/bkup_db4' format '/zfs/bkup1/%U';
allocate channel ch13 device type disk connect 'sys/passwd@exa-scan/bkup_db1' format '/zfs/bkup1/%U';
allocate channel ch14 device type disk connect 'sys/passwd@exa-scan/bkup_db2' format '/zfs/bkup1/%U';
allocate channel ch15 device type disk connect 'sys/passwd@exa-scan/bkup_db3' format '/zfs/bkup1/%U';
allocate channel ch16 device type disk connect 'sys/passwd@exa-scan/bkup_db4' format '/zfs/bkup1/%U';
configure device type disk parallelism 16;
RECOVER COPY OF DATABASE
  WITH TAG 'incr_zfs';
BACKUP
  INCREMENTAL LEVEL 1
  FOR RECOVER OF COPY WITH TAG 'incr_zfs'
  DATABASE format '/zfs/bkup2/%U';
}
```

Oracle ZFS Storage ZS4-4におけるRecovery Managerのパフォーマンス・サイジング

次のグラフは、Oracle ZFS Storage ZS4-4 への Recovery Manager のバックアップおよびリストアの実行時に、Oracle 社内検証で取得された持続可能なスループットを示します。これらの速度は、従来型の Recovery Manager のバックアップ戦略の一部として、レベル 0 のバックアップを実行して収集しました。このグラフは、ディスク・シェルフが 2～8 つのさまざまな Oracle ZFS Storage ZS4-4 構成における、持続可能な最大スループットを示しています。

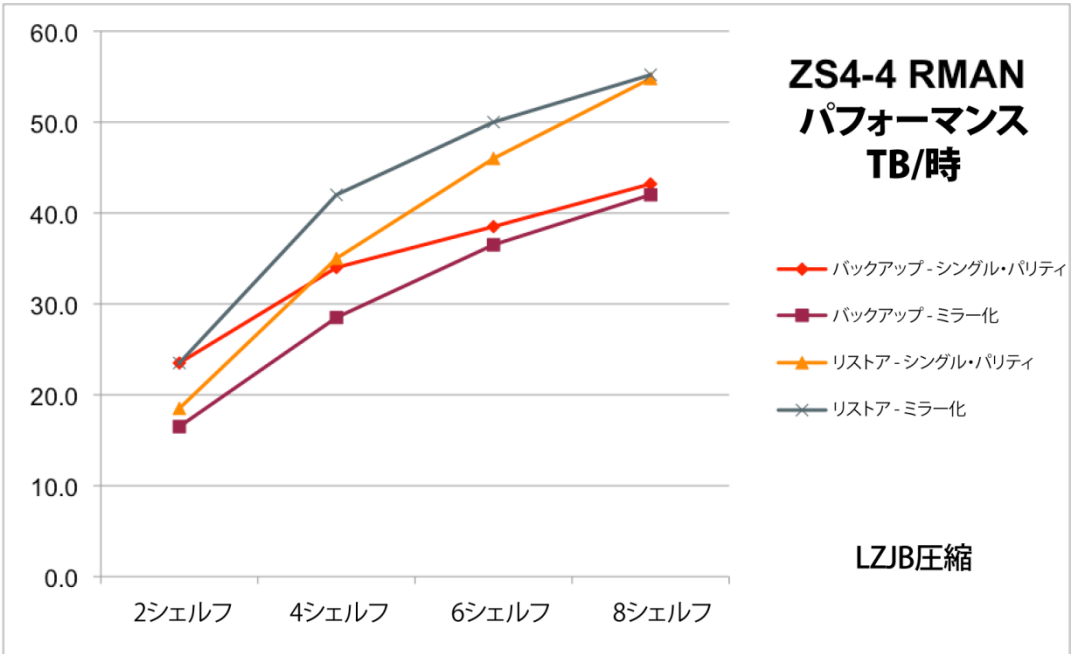

図9：Oracle ZFS Storage ZS4-4のZFS LZJB圧縮によるRecovery Managerのスループット

これらは、Oracle Database 12c、および販売受注スキーマにお客様のサンプル・データが移入されている、大規模な OLTP データベースを使用した場合の実際の結果です。ここでは、データベース・レベルで Advanced Row Compression を使用します。これが、OLTP ワークロードを実行中のお客様にとってベスト・プラクティスの推奨事項であるためです。これは、Recovery Manager のバックアップ操作の実行前、実行中、または実行後にライブ・トランザクション・ワークロードを実行する機能を備えた、完全に機能的なデータベースです。これらのスループット速度は、データベースまたは I/O ジェネレータ・テスト・ツールを使用すると、誤った結果を導く可能性があり、さらに実際のユースケースには間接的にしか適用できないため、この手法では取得しませんでした。また、低レベルのシステム・ベンチマークからは予測されません。

すべてのシナリオにおいて、持続可能なスループットは、ネットワーク層での物理 I/O を測定して決定されます。平均値は、長時間にわたるデータから収集しました。これらのグラフは、Oracle Database の最大バックアップおよびリストア速度を示しており、Oracle ZFS Storage ZS4-4 が、実際の Exadata バックアップ環境において、これらの速度が持続可能であることを証明しています。持続可能な最大リストア速度は 55TB/時で、最大バックアップ速度は 42TB/時であることを示して

います。これらの速度はすべて、80:20 DATA/RECO 分割および DATA ディスク・グループの標準 ASM 冗長性を備えた、Exadata Database Machine X5-2 フルラックにより達成できます。パフォーマンス・データは、Oracle Recovery Manager の操作時に、Exadata および Oracle ZFS Storage ZS4-4 の両方がアイドル状態の間に収集しました。テスト環境は、このホワイト・ペーパーで示されているベスト・プラクティスに従って構成しました。

次のグラフは、ZFS 共有レベルの圧縮を使用せずに得られた、持続可能なスループットを示しています。

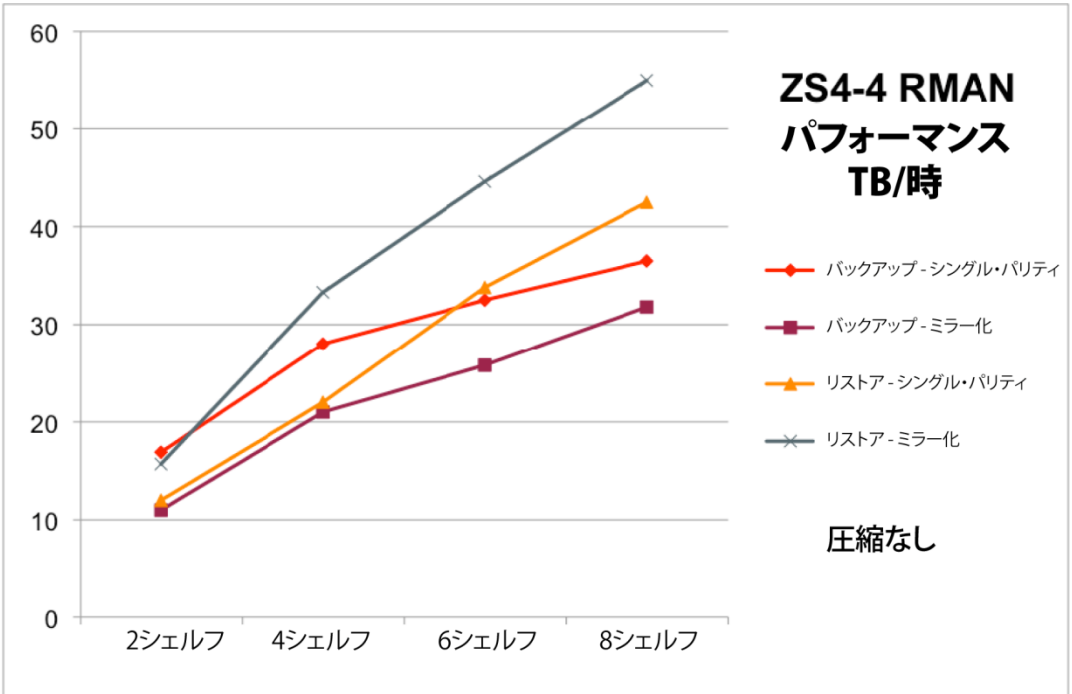

図10：Oracle ZFS Storage ZS4-4のZFS圧縮を使用しないRecovery Managerのスループット

ミラー化ストレージからのリストアなどの特定のシナリオでは、ネットワークのスループットは影響を受けません。また、8 つのディスク・シェルフにより、55TB/時を達成しているケースもあります。バックアップのパフォーマンス全体において、HDD 利用率または SAS 帯域幅が制限要因であるシナリオでは、ZFS 共有で LZJB データ圧縮を使用することにより、バックエンドのボトルネックを軽減できるため、ネットワーク全体のスループットが向上します。

次の最後のグラフは、単一の Oracle ZFS Storage ZS4-4 コントローラのみでバックアップまたはリストアを実行する際に、持続可能な最大の Recovery Manager スループットを示しています。これらのレベル 0 バックアップおよびリストア操作では、1 つの共有のみを使用しました。共有を所有できるのは、一度に 1 つのコントローラのみであるため、これらの Recovery Manager ワークロードを実行するリソースは、単一のコントローラを基にしました。また、これらのジョブは、CPU、メモリ、ネットワーク、PCIe、および SAS リソースの半分を使用して実行されました。パフォーマンス・データは、シェルフが 1～4 つのストレージ・プール構成を使用して収集しており、ミラー化およびシングル・パリティ・プロファイルの両方が含まれています。

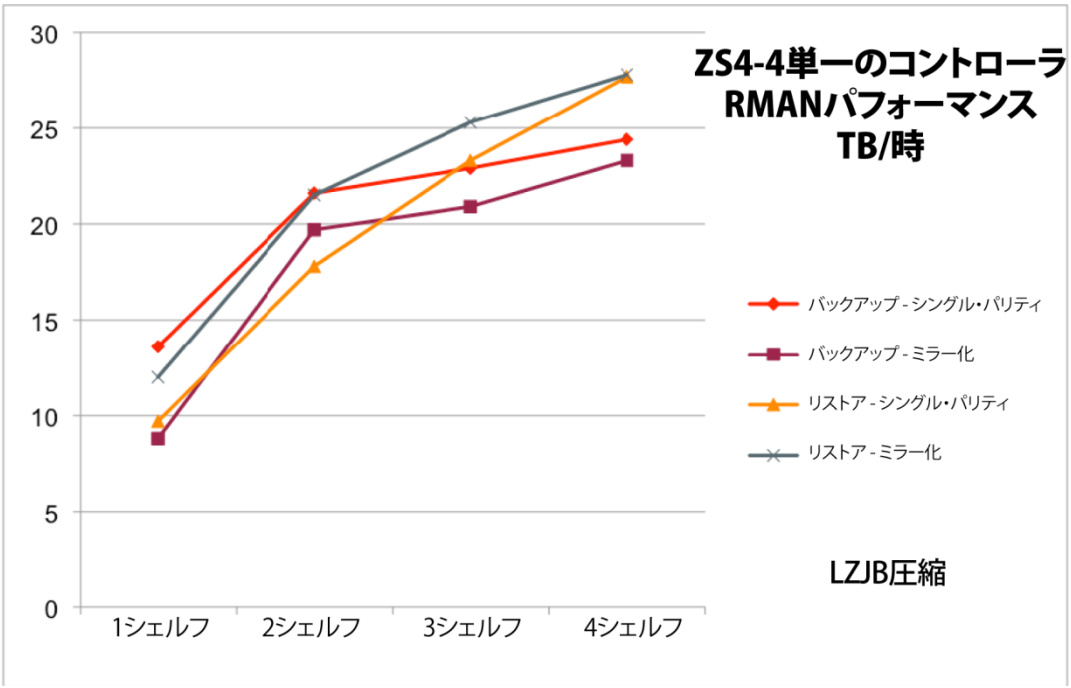

図11：Oracle ZFS Storage ZS4-4の単一のコントローラでのRecovery Managerスループット

Recovery Manager のワークロードによるパフォーマンス・サイジングでは、もっとも一般的に展開されている、ミラー化およびシングル・パリティ・ストレージ・プロファイルに注目しました。ダブル・パリティでも、Recovery Manager のワークロードによるシングル・パリティと同様のパフォーマンス特性が得られます。パフォーマンスは多少低くなりますが、通常、シングル・パリティの場合の 5～10%以内の範囲です。

Oracle ZFS Storage ZS3-2におけるRecovery Managerのパフォーマンス・サイジング

次のグラフは、Oracle ZFS Storage ZS3-2 への Recovery Manager のバックアップおよびリストアの実行時に、Oracle 社内検証で取得された持続可能なスループットを示します。これらの速度は、従来型の Recovery Manager のバックアップ戦略の一部として、レベル 0 のバックアップを実行して収集しました。このグラフは、ディスク・シェルフが 2～8 つのさまざまな Oracle ZFS Storage ZS3-2 構成における、持続可能な最大スループットを示しています。

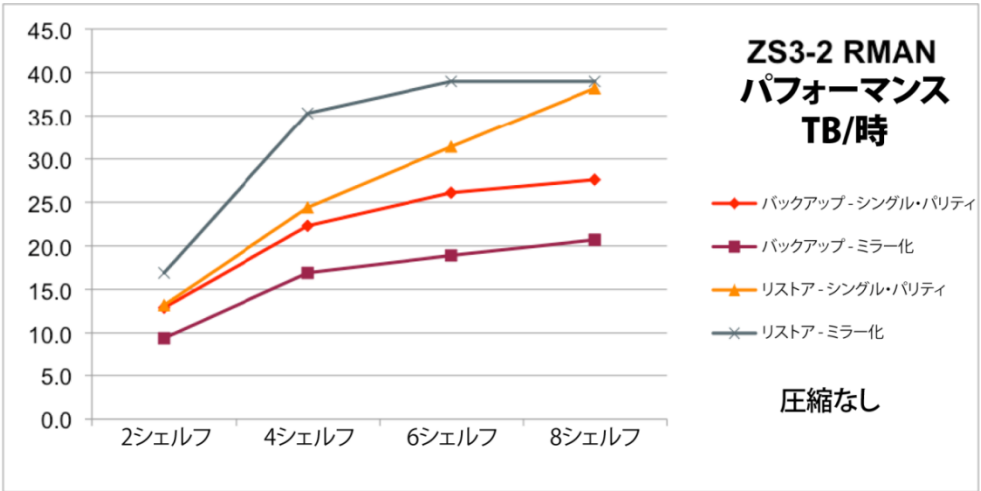

図12：Oracle ZFS Storage ZS3-2におけるRecovery Managerのスループット

これらは、Oracle Database 12c、および販売受注スキーマにお客様のサンプル・データが移入されている、大規模な OLTP データベースを使用した場合の実際の結果です。ここでは、データベース・レベルで Advanced Row Compression を使用します。これが、OLTP ワークロードを実行中のお客様にとってベスト・プラクティスの推奨事項であるためです。これは、Recovery Manager のバックアップ操作の実行前、実行中、または実行後にライブ・トランザクション・ワークロードを実行する機能を備えた、完全に機能的なデータベースです。これらのスループット速度は、データベースまたは I/O ジェネレータ・テスト・ツールを使用すると、誤った結果を導く可能性があり、さらに実際のユースケースには間接的にしか適用できないため、この手法では取得しませんでした。また、低レベルのシステム・ベンチマークからも予測されません。

すべてのシナリオにおいて、持続可能なスループットは、ネットワーク層での物理 I/O を測定して決定されます。平均値は、長時間にわたるデータから収集しました。これらのグラフは、Oracle Database の最大バックアップおよびリストア速度を示しており、Oracle ZFS Storage ZS3-2 が、実際の Exadata バックアップ環境において、これらの速度を持続可能であることを証明しています。持続可能な最大リストア速度は 39TB/時で、最大バックアップ速度は 28TB/時であることを示しています。これらの速度はすべて、DATA ディスク・グループの標準 ASM 冗長性を備えた、Exadata Database Machine X5-2 フルラックにより達成できます。パフォーマンス・データは、Oracle Recovery Manager の操作時に、Exadata および Oracle ZFS Storage ZS3-2 の両方がアイドル状態のときを使用して収集しました。テスト環境は、このホワイト・ペーパーで示されているベスト・プラクティスに従って構成しました。

前のグラフにおける Recovery Manager のパフォーマンス・サイジング処理は、各コントローラに 2 つの SAS HBA が取り付けられている、Oracle ZFS Storage ZS3-2 で実施されました。ただし、一部の ZFS Storage ZS3-2 では、各コントローラに単一の SAS HBA のみが展開されています。これらの環境では、バックアップおよびリストアのスループットにおいて、SAS 帯域幅が制限要因になる場合があります。次のグラフは、SAS 制約付き構成による影響を示しています。

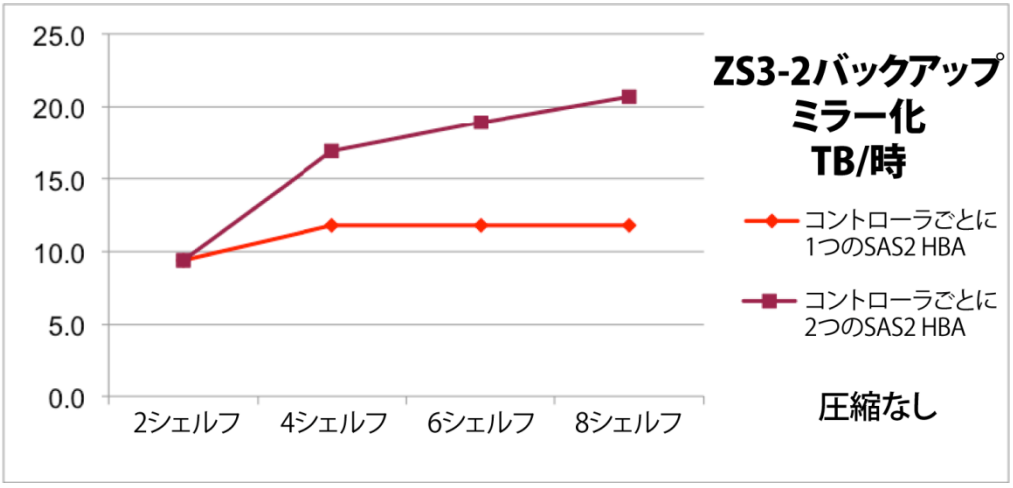

図13：複数のSAS HBA構成による、Oracle ZFS Storage ZS3-2におけるRecovery Managerのバックアップ・ミラー化スループットの比較

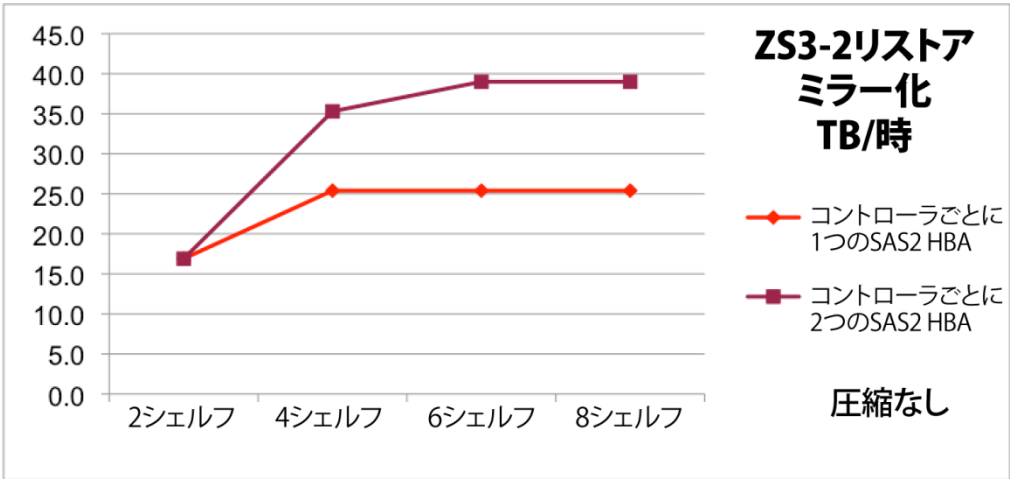

図14：複数のSAS HBA構成による、Oracle ZFS Storage ZS3-2におけるRecovery Managerのリストア・ミラー化スループットの比較

ミラー化ストレージへのバックアップは、バックエンド・ディスクへの帯域幅が 2 倍になるため、SAS 制約事項に関するもっとも困難な処理です。前のグラフに示したように、リストア操作では帯域幅は膨張しませんが、SAS は制限要因のままです。シングル・パリティでは、幅の狭い 3+1 ストライプを利用しますが、パリティが原因で書込み時の帯域幅が 33%膨張します。これは、ミラー化の場合と比較して、SAS 帯域幅での負担は少ないですが、単一の HBA 構成が即座に飽和してしまう可能性があります。クラスタ化 Oracle ZFS Storage ZS3-2 が、4 つ以上のディスク・シェルフで構成されている場合、各コントローラで 2 つの SAS HBA を使用することを推奨します。

Recovery Manager のワークロードによるパフォーマンス・サイジングでは、もっとも一般的に展開されている、ミラー化およびシングル・パリティ・ストレージ・プロファイルに注目しました。ダブル・パリティでも、Recovery Manager のワークロードによるシングル・パリティと同様のパフォーマンス特性が得られます。パフォーマンスは多少低くなりますが、通常、シングル・パリティの場合の 5～10%以内の範囲です。

結論

Oracle Exadata Database Machine において、適切なバックアップ・ソリューションを見つけることは、困難な問題です。コストの高い代替手法を使用しても、投資に対する利益は低く、高パフォーマンスの環境もサポートできません。競合製品は柔軟性がなく、お客様のニーズのすべてに対応できるわけではありません。

Oracle ZFS Storage Appliance は、Oracle Exadata Database Machine 上のミッション・クリティカルなデータを保護するための理想的なソリューションであることが証明されました。強力な各種機能と Oracle 間のカスタム統合を組み合わせることにより、Oracle Recovery Manager の幅広いバックアップ戦略のオプションが使用可能になります。これらのオプションにより、サード・パーティ製のソリューションでは対抗できないような、他に類を見ないパフォーマンスと柔軟性が実現されます。

優れたリストア・スループットは、もっとも厳格なリカバリ時間目標の場合でも対応できます。アーカイブ・ログは多重化されるため、リカバリ・ポイントは 20 分以下で提供されます。Oracle Intelligent Storage Protocol、Hybrid Columnar Compression、および Oracle Direct NFS は、Oracle Database を保護する際にそれぞれ独自のメリットを提供します。ネイティブの InfiniBand サポートは、高スループットの Exadata インフラストラクチャとシームレスに統合されます。

Oracle ZFS Storage Appliance を使用した Exadata バックアップ・ソリューションでは、データ保護に関するメリットの他に、Oracle Database の外部の非構造化データに対する、低コストで高パフォーマンスなストレージの実現、開発およびテスト環境のプロビジョニングで理想的なスナップショット・クローンを作成するソリューションの実現など、多数のメリットが存在します。このように、Oracle ZFS Storage Appliance が、Exadata Database Machine を保護するための最適なソリューションである理由が簡単に理解できます。

お問い合わせ窓口
Oracle Direct
TEL 0120-155-096
URL oracle.com/jp/direct

ORACLE

CONNECT WITH US

blogs.oracle.com/oracle
facebook.com/oracle
twitter.com/oracle
oracle.com

Integrated Cloud Applications & Platform Services

Copyright © 2015, Oracle and/or its affiliates. All rights reserved.
本文書は情報提供のみを目的として提供されており、ここに記載される内容は予告なく変更されることがあります。本文書は一切間違いがないことを保証するものではなく、さらに、口述による明示または法律による黙示を問わず、特定の目的に対する商品性もしくは適合性についての黙示的な保証を含め、いかなる他の保証や条件も提供するものではありません。オラクル社は本文書に関するいかなる法的責任も明確に否認し、本文書によって直接的または間接的に確立される契約義務はないものとします。本文書はオラクル社の書面による許可を前もって得ることなく、いかなる目的のためにも、電子または印刷を含むいかなる形式や手段によっても再作成または送信することはできません。
Oracle および Java は Oracle およびその子会社、関連会社の登録商標です。その他の名称はそれぞれの会社の商標です。
IIntel および Intel Xeon は Intel Corporation の商標または登録商標です。すべての SPARC 商標はライセンスに基づいて使用される SPARC International, Inc.の商標または登録商標です。AMD、Opteron、AMD ロゴおよび AMD Opteron ロゴは、Advanced Micro Devices の商標または登録商標です。UNIX は X/Open Company, Ltd.によってライセンス提供された登録商標です。0615
Oracle ZFS Storage Appliance による Exadata Database Machine の保護
2015 年 7 月
著者：Greg Drobish

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