



Oracle ホワイト・ペーパー
2013年1月

Exadata Smart Flash Cacheの機能とOracle Exadata Database Machine

フラッシュ・テクノロジーとExadata Database Machine.....	2
Oracle Database 11g：フラッシュ向けに最適化された初のデータベース.....	3
Exadata Smart Flash Cacheハードウェア	5
Exadata Storage Server SoftwareとFlash Cacheハードウェア.....	6
Exadata Smart Flash Cache：データベース・オブジェクトのためのフラッシュ	6
Exadata Smart Flash Logging：データベース・ロギングのためのフラッシュ	9
Exadata Smart Flash Cacheが提供するミッション・クリティカルな可用性	11
結論	11

フラッシュ・テクノロジーとExadata Database Machine

Oracle Exadata Database Machineは、オンライン・トランザクション処理（OLTP）、データウェアハウス（DW）、統合された複合データベース・ワークロードなど、すべてのデータベース・アプリケーションに対して、卓越したパフォーマンスとスケーラビリティをもたらします。これを実現する重要な要素の1つがExadata Smart Flash Cacheハードウェアであり、インテリジェントなOracle Exadata Storage Server Softwareと組み合わせることで機能します。Exadata Storage Server SoftwareのExadata Smart Flash Cache機能は、データベース・オブジェクトをフラッシュ・メモリにインテリジェントにキャッシュします。これにより、パフォーマンスの低い機械的なディスクI/O操作から、高速なフラッシュ・メモリ操作に移行することが可能となります。Exadata Storage Server Softwareは、データベース・ログのI/Oを高速化するExadata Smart Flash Logging機能も提供しています。Exadata Smart Flash Cacheは、Oracle Exadata Database Machineの中核をなすテクノロジーです。Oracle Exadata Database Machineは、1秒あたりのランダムI/O操作数（IOPS）で最大150万を実現するとともに、Exadataストレージ内のデータを最大100GB/秒でスキャンします。

Oracle Database 11g :

フラッシュ向けに最適化された初のデータベース

オラクルのExadata Smart Flash Cache機能は他に例を見ない独自の機能です。Exadataフラッシュ・ストレージは単にディスクを置き換えるだけではありません。Exadataソフトウェアが、いつ、どのようにフラッシュ・ストレージを使用するかを決定し、データ・キャッシングの協調戦略の一環として、最適な方法でデータベースにフラッシュを組み込みます。Exadataストレージをスケールアウトすると、フラッシュ・パフォーマンスがもたらすメリットをアプリケーションでも享受できます。従来のストレージ・アレイは内部やネットワークに多数のボトルネックを抱えているため、フラッシュによる利点が妨げられています。フラッシュをストレージ・アレイに追加することはできても、潜在的なパフォーマンスはアプリケーションにまで及びません。

Exadata Database Machineは、フラッシュとの間に最大100GB/秒の帯域幅を実現しています。これはその他のソリューションと比べて飛躍的に高いパフォーマンスであり、しかも非圧縮データに対する数値です。この帯域幅をHybrid Columnar Compressionおよびオフロード処理と組み合わせると、実効帯域幅は格段に向上します。従来のストレージ・アレイのディスクを集約した帯域幅は、アレイ・コントローラで処理可能な帯域幅を上回ります。ほとんどのストレージ・アレイで、アレイ・コントローラの数GB/秒の帯域幅がボトルネックとなります。これらのシステムにフラッシュを追加しても、帯域幅は向上しません。このボトルネックの重大度が高まるだけです。従来のストレージ・アレイでは、ビジネスの要件に対応するために優れた順次的な帯域幅を必要とするような、データウェアハウスや、OLTPバッチおよびレポートのワークロードにおいて、フラッシュによる効果は得られません。

従来のストレージ・アレイは、単純なディスクによるランダムIOPS処理に向いています。従来のストレージ・アレイがランダムIOPSに対応できるのは、大部分のOLTPシステムではランダムIOPSが比較的少ないためです。しかし、通常のディスクと比べて桁違いに高速なフラッシュがシステムに追加されると、従来のストレージ・アレイがボトルネックになります。たとえば、ハイエンドのストレージ・サブシステムは、約12万IOPS程度で実行されています。一方、Exadata Database Machineは、データベース・レベルで最大150万IOPSを達成しています。Exadataソフトウェアは、フラッシュとディスクから同時にスキャンして帯域幅を最大化できます。これは、フラッシュでの150万IOPSが、ストレージ・サーバーやネットワークを介してデータベース・サーバーにおいても実現されることを意味します。ストレージ・アレイはすでにボトルネックとなっているため、フラッシュを追加しても状況は悪化する一方です。

オラクルでは、フラッシュ・ディスクの代わりにフラッシュPCIeカードを使用しています。フラッシュ・ディスクを既存のストレージ・サブシステムに追加する場合、その他を変更する必要がないため作業は簡単ですが、テクノロジーが持つ潜在能力を有効に発揮することはできません。ディスク・コントローラやディスク・ディレクタは、フラッシュ・ディスクが提供するパフォーマンスに対応するように設計されていないためです。オラクルのソリューションではフラッシュPCIeカードを使用しているため、パフォーマンスの低いディスク・コントローラによってフラッシュのパフォーマンスが阻害されることはありません。Exadataストレージでは、フラッシュ・カードごとに1.33GB/秒近いスループットが実現されており、それぞれのExadata Storage Serverで4つのカードに対してこのパフォーマンスが直線的に向上します。従来のストレージ・アレイの場合、システムにフラッシュ・カードを追加することはできません。ディスク・コントローラによる制約をなくすには、アーキテクチャを変更する必要があります。

Oracle Exadata Storage Serverにはスマート・フラッシュ・キャッシュが直接実装されています。Exadata Smart Flash Cacheでは、頻繁にアクセスされるデータは非常に高速なフラッシュ・ストレージに格納されますが、大部分のデータは非常にコスト効率の高いディスク・ストレージに格納されます。この処理は、ユーザーの手を煩わすことなく自動的に実行されます。Oracle Flash Cacheは、再利用されることのないデータやキャッシュに適さないデータを巧みに見分けてキャッシングを回避します。また、Oracle DatabaseとExadataストレージでは、データベースの表や索引、またはセグメントのレベルで特定のデータをフラッシュに保持するよう、ユーザーが指定することもできます。フラッシュへの表の出し入れは簡単なコマンドで実行でき、フラッシュ・ディスクを含んだ従来のストレージのように、別の表領域やファイル、またはLUNに表を移動させる必要はありません。

Exadata Smart Flash Cacheは、ログ書き込み/I/Oの待機時間を短縮するためにも使用でき、データベースのロギングにより発生する可能性のあるパフォーマンス・ボトルネックを解消します。ユーザー・トランザクションのコミット時間は、ログ書き込みの待機時間に大きく影響を受けます。また、領域管理や索引分割などのパフォーマンス・クリティカルなデータベース・アルゴリズムの多くもログ書き込みの待機時間に大きく影響を受けます。今日のExadataストレージは、ディスク・コントローラでバッテリー・バックアップ式DRAMキャッシュを使用することでログ書き込みを高速化します。通常、ディスク・コントローラ・キャッシュへの書き込みは非常に高速ですが、ディスク/I/Oが多い期間は低速になることがあります。Smart Flash LoggingでExadataストレージのフラッシュ・メモリを利用することで、ログ書き込みが高速化されます。

フラッシュ・メモリにおける書き込み待機時間の平均値は非常に良好ですが、この平均よりも1~2桁も速度が低下するような遅延の異常値がたびたび発生します。Exadata Smart Loggingの考え方は、REDO書き込みをフラッシュ・メモリとディスク・コントローラ・キャッシュの両方に対して同時に実行し、その2つのいずれかが最初に完了した時点で書き込みを完了するというものです。これによって文字どおり、Exadataにこの2つのうち良い方が取り入れられます。Smart Flash Loggingによりパフォーマンス・クリティカルなデータベース・アルゴリズムを高速化することで、ユーザー・トランザクションの応答時間を削減し、さらにI/O集中型ワークロードの全体的なデータベース・スループットも改善します。

ExadataにはHybrid Columnar Compressionが導入されており、過去に実現されたあらゆるテクノロジーをはるかに凌ぐレベルの圧縮が可能です。これにより、大量のデータの保管コストを大幅に削減できるばかりでなく、データのスキャン速度を向上できるなど、数多くの利点がもたらされます。また、Exadataのフラッシュ・テクノロジーと組み合わせることで大きな相乗効果が得られます。データを10倍以上に圧縮できるため、10倍を超えるデータをフラッシュに格納できます。つまり、このフラッシュは、同じフラッシュ機能を持つその他のどのような製品よりもはるかに効果的に機能します。

オラクルのフラッシュ・テクノロジーはExadataのエンド・ツー・エンド・アーキテクチャと緊密に統合され、データベース・ストレージ階層（DRAM、フラッシュ、ディスク）に完全に統合されます。ユーザーが手動で管理および最適化しなくてはならないような付加的なアクセラレータではありません。Exadata Smart Flashテクノロジーは、データベースやシステム管理者に負荷をかけずに、もつとも要求の厳しいアプリケーションに必要とされる帯域幅とIOPSを実現します。

Exadata Smart Flash Cacheハードウェア

Exadataシステムでは、フラッシュ・ディスクではなく最新のPCIフラッシュ・テクノロジーを採用しています。低速のディスク・コントローラやディスク・ディレクタの背後ではなく、高速PCIバスに直接フラッシュを配置することで、PCIフラッシュのパフォーマンスが大幅に向上します。Exadata Storage Serverには、4つのPCIフラッシュ・カードが搭載されており、フラッシュ・メモリの総容量は1.6TBです。また、フル・ラックのExadata Database Machineには56個のPCIフラッシュ・カードが搭載されており、そのフラッシュ・メモリは22.4TBです。Exadata X3に使用されるフラッシュ・モジュールは、標準的なデータベースのデータを使用する場合に、10年以上の耐久性が期待されます。このソリッド・ステート・ストレージは、Exadataストレージのパフォーマンスを大幅に向上します。データベースの読取りと書込みを自動的にキャッシュし、フル・ラックのX3データベース・マシンにおいて、100万回/秒の8Kフラッシュ書込みと150万回/秒の8Kフラッシュ読取りを実行できます。



Sun Flash Accelerator F40 PCIeカード

Exadataのきわめて高いパフォーマンスを実現する重要な要素の1つがExadata Smart Flash Cacheハードウェアであり、インテリジェントなOracle Exadata Storage Server Softwareと組み合わせることで機能します。Exadata Storage Server SoftwareのExadata Smart Flash Cache機能は、データベース・オブジェクトをフラッシュ・メモリにインテリジェントにキャッシュします。これにより、パフォーマンスの低い機械的なディスクI/O操作から、高速なフラッシュ・メモリ操作に移行することが可能となります。Exadata Storage Server Softwareは、データベース・ログのI/Oを高速化するExadata Smart Flash Logging機能も提供しています。Exadata Smart Flash Cacheは、Oracle Exadata Database Machineの中核をなすテクノロジーです。Oracle Exadata Database Machineは、1秒あたりのランダム8K I/O操作数 (IOPS) で最大150万を実現するとともに、Exadataストレージ内のデータを最大100GB/秒でスキャンします。

さまざまなExadata X3構成のデータベース・レベルにおけるExadata Smart Flash Cacheのパフォーマンスについて、次の表に示します。

	Exadata Database Machine X3-8 および X3-2 フル・ラック	Exadata Database Machine X3-2 ハーフ・ラック	Exadata Database Machine X3-2 クォーター・ラック	Exadata Database Machine X3-2 1/8ラック
Exadata Smart Flash Cache	22.4TB	11.2TB	4.8TB	2.4TB
ディスク・データ帯域幅 • 高パフォーマンスSAS • 大容量SAS(データ圧縮なし)	最大 25GB/秒 18GB/秒	最大 12.5GB/秒 9.0GB/秒	最大 5.4GB/秒 4.0GB/秒	最大 2.7GB/秒 2.0GB/秒
フラッシュ・データ帯域幅 • 高パフォーマンスSAS • 大容量SAS(データ圧縮なし)	最大 100GB/秒 93GB/秒	最大 50GB/秒 46.5GB/秒	最大 21.5GB/秒 20GB/秒	最大 10.7GB/秒 10GB/秒
データベース・フラッシュ・キャッシュの8K読取りIOPS	最大1,500,000	最大750,000	最大375,000	最大187,000
データベース・フラッシュ・キャッシュの8K書込みIOPS	最大1,000,000	最大500,000	最大250,000	最大125,000
データベース・ディスクのIOPS • 高パフォーマンスSAS • 大容量SAS	最大 50,000 28,000	最大 25,000 14,000	最大 10,800 6,000	最大 5,400 3,000

Exadata Storage Server SoftwareとFlash Cacheハードウェア

Exadata Flashハードウェアを使用するExadata Storage Server Softwareのおもな機能は2つあり、これらにより、Exadata Database MachineはOracle Databaseを配置するための非常に高速なシステムとなります。おもな機能の1つは、Exadata Smart Flash Cacheです。これは、フラッシュ内のアクティブなデータベース・オブジェクトをステージングする機能を提供します。もう1つは、Exadata Smart Flash Loggingです。これは、データベース・ロギングという重要な機能を高速化するものです。さらに、Oracle Databaseを配置するにはミッション・クリティカルな耐久性が必要ですが、Exadata Storage Server SoftwareをOracle Databaseと組み合わせることでこの耐久性を実現します。

Exadata Smart Flash Cache：データベース・オブジェクトのためのフラッシュ

Exadata Smart Flash Cacheは、Exadata Database Machine内で頻繁にアクセスされるデータを自動的にキャッシュするメカニズムを提供します。Exadata Smart Flash Cacheは非常に多くのランダム読取り/書込みを処理できるライトバック・キャッシュであり、これによってDatabase Machineに配置されたOLTPアプリケーションの応答性が向上します。

自動化されたExadata Smart Flash Cache管理

Exadata Smart Flash Cacheでは、頻繁にアクセスされるデータは非常に高速なフラッシュ・ストレージに格納されますが、大部分のデータは非常にコスト効率の高いディスク・ストレージに格納されます。この処理は、ユーザーの手を煩わすことなく自動的に実行されます。Oracle Flash Cacheは、再利用されることのないデータやキャッシュに適さないデータを巧みに見分けてキャッシングを回避します。

読取り要求または書き込み要求が、データベースからExadata Storage Serverに送信される際、再読込みされる可能性の高いデータとキャッシュすべきデータに関する情報がリクエストに追加されます。データベースから送信された情報に基づき、Exadata Storage Server Softwareは再読込みされる可能性の高いデータ、キャッシングに値するデータ、キャッシュする必要のないデータをインテリジェントに決定します。表や索引のランダム読取り/書き込みが実行されると、それに続く読取りが発生する可能性が高いため、通常これらはキャッシュされ、フラッシュ・キャッシュからデータが提供されます。一般的に、スキャンや順次読取りされた表はキャッシュされません。これは、順次アクセスされたデータに対して再読取りが発生する可能性が低いからです。

キャッシュすべきでないデータを把握することは、キャッシュのパフォーマンスの潜在能力を發揮するために非常に重要です。たとえば、バックアップを書き込む場合や、ブロックのミラー・コピーに書き込む場合、このソフトウェアはこれらのブロックをキャッシュしません。これらのブロックが再読込みされることは当面ないため、貴重なキャッシュ領域を割く理由はないからです。このような可視性を持ち、システム上で実行されるすべてのI/O操作の性質を把握しているのは、Oracle DatabaseとExadata Storage Server Softwareだけです。I/Oスタック全体がくまなく把握されているため、Exadata Smart Flash Cacheハードウェアを最大限に利用し、もっとも妥当性の高いデータのみをキャッシュに格納できます。

Exadata Smart Flash Cacheソフトウェア (11.2.2.3.2) はデータベースの読取り/書き込みの自動キャッシュを実装しており、書き込みIOPSが20倍、データ帯域幅が33%向上します。このソフトウェアは、X3 Database Machineフル・ラックで100万回/秒の8K SQLフラッシュ書き込みIOPSを実現します。22.4TBのフラッシュ容量とともに、この新しい書き込みキャッシュ・テクノロジーによって、もっとも要求の厳しいOLTPデータベースでも超高速パフォーマンスを確保できます。

この新しいExadata Smart Flash Cacheソフトウェアには以前までの世代のDatabase Machineとの下位互換性があり、以前までの世代のシステムに対する書き込みキャッシュが可能です。X2およびV2のDatabase Machineでは最大50万回/秒の8K SQLフラッシュ書き込みを実行でき、既存のユーザーに対してパフォーマンスの向上を透過的にもたします。さらに、Exadata Smart Flash CacheはExadata Storage Serverが再起動された後も持続し、ウォームアップ期間はありません。

1つのExadata Database Machineへの複数のデータベースの統合は、コスト削減につながるソリューションです。Exadata Storage Server Software 11.2.2.3以降では、Exadata I/O Resource Manager (IORM) を使用して、Database Machineで稼働するそれぞれのデータベースについて、フラッシュの使用を有効または無効に切り替えることができます。この機能によって、もっともパフォーマンス・クリティカルなデータベース向けのフラッシュを予約できます。

また、Oracle DatabaseとExadataストレージではオプションで、データベースの表や索引、またはセグメントのレベルで特定のデータをフラッシュに保持するよう、ユーザーが指定することもできます。表はフラッシュ内で保持でき、従来のストレージやフラッシュ・ディスクのように、別の表領域やファイル、またはLUNに表を移動させる必要はありません。

ユーザーによるExadata Smart Flash Cache管理

キャッシュの使用と管理を手動で行うには、次の2つの方法があります。1つは、フラッシュ・キャッシュにオブジェクトを固定する方法です。もう1つは、フラッシュから論理ディスクを作成して、フラッシュ・ディスク上にオブジェクトを永続的に配置する方法です。

フラッシュ・キャッシュへのオブジェクトの固定

Exadata Storage Server SoftwareとOracle Databaseでは、どのデータベース・オブジェクトをキャッシュするかを優先的に指定する機能が提供されています。たとえば、オブジェクトをキャッシュ内に固定して、常にキャッシュしておくことができます。また、キャッシュしないように指定することもできます。このコントロールは、新しいストレージ句属性であるCELL_FLASH_CACHEによって実現されます。この属性は、データベースの表、索引、パーティション、およびLOB列に対して割り当てることができます。

CELL_FLASH_CACHE属性には、次の3つの値のいずれかを設定できます。DEFAULTを指定すると、このオブジェクトに対して使用されるキャッシュは、前項で説明したとおりの方法で自動的に管理されます。NONEを指定されたオブジェクトは、どのような場合もキャッシュされません。KEEPを指定すると、そのオブジェクトがキャッシュ内に読み込まれた後にそのまま保持されます。

たとえば、CUSTOMERS表のページをExadata Smart Flash Cacheに読み込んだ後にそのまま保持するには、次のコマンドを使用します。

```
ALTER TABLE customers STORAGE (CELL_FLASH_CACHE KEEP)
```

このストレージ属性を表の作成時に指定することもできます。

この場合、Exadata Storage ServerはCUSTOMERS表のデータをキャッシュし、KEEPが指定されていないその他の表がエージングによりキャッシュから削除されても、そのCUSTOMERS表のデータを保持し続けます。通常、CUSTOMERS表は多数のExadata Storage Server上に分散されているため、それぞれのExadataセルのフラッシュには、表の一部がキャッシュされます。通常は、KEEPが指定されたオブジェクト自体より多くのフラッシュ・キャッシュが利用可能な状態にあります。この場合、表全体が長期的にキャッシュされます。

デフォルトの動作では、順次スキャンされたデータはフラッシュ・キャッシュに格納されませんが、KEEPが指定されている場合はこれに当てはまりません。KEEPが指定されたオブジェクトが、オフロードされたSmart Scanを介してアクセスされている場合、このオブジェクトはキャッシュに保持され、キャッシュからスキャンされます。フラッシュ・キャッシュが優れているもう1つの点は、キャッシュに保持されているオブジェクトがスキャンされると、Exadataソフトウェアによって、フラッシュとディスクの両方からデータが読み取られることです。これにより、各ソースから単独にデータを読み取った場合よりも優れた総計スキャン速度が達成されます。

フラッシュ・キャッシュを使用したフラッシュ・ディスクの作成

Exadataセルがインストールされると、デフォルトでフラッシュがFlash Cacheとして（またはSmart Logging用に）割り当てられ、ユーザー・データはデフォルトのキャッシュ動作で自動的にキャッシュされます。

また、キャッシュの一部は論理フラッシュ・ディスクとして予約し使用することもできます。このようなフラッシュ・ディスクは、その他すべてのExadataセル・ディスクと同様に取り扱われますが、実際はキャッシュ内に常駐し、非揮発性ディスクとして保存されます。それぞれのExadataセルでフラッシュ・ディスク用に確保される領域は、16のセル・ディスク（フラッシュ・カードあたり4つのセル・ディスク）に割り当てられます。これらのフラッシュベースのセル・ディスク上にはグリッド・ディスク（物理セル・ディスク上の論理ディスク）が作成され、このグリッド・ディスクは自動ストレージ管理（Oracle ASM）ディスク・グループに割り当てられます。ここでのベスト・プラクティスは、各Exadataセルで同じサイズのフラッシュをフラッシュ・ディスク向けに確保し、通常のExadataグリッド・ディスクに対する構成と同様に、Oracle ASMディスク・グループをExadataセル間で均等に分散する方法です。こうすることで、フラッシュI/OロードがExadataセルおよびフラッシュ間で均等に分散されます。

パフォーマンスの高いこれらの論理フラッシュ・ディスクは、頻繁にアクセスされるデータを格納するために使用できます。論理フラッシュ・ディスクの使用に際しては、事前に計画を立てて、表領域を格納する領域を適切に確保しておく必要があります。また、メディア・リカバリが必要な場合は、従来のディスクに保存されたデータと同じように、フラッシュ・ディスク上のデータもバックアップする必要があります。ほとんどの状況で、この機能を使用しても、この方法でのフラッシュの設定と管理にかかるオーバーヘッドに対して得られる利点は限定的です。しかし、このオプションは、書込み集中型のワークロードで、ディスク書込み率がディスクで処理できる量を上回るようなケースに有効となる場合があります。

Exadata Smart Flash Logging：データベース・ロギングのためのフラッシュ

OLTPワークロードでは、データベース・ログ書込みの応答時間が短いことが非常に重要です。データベース管理者（DBA）は可用性を高めるためにREDOログ・グループとミラー化されたログ・ファイルを構成しますが、パフォーマンスの低いディスクがREDOログの書込み待機時間とシステムのパフォーマンスに悪影響を及ぼすことがあります。もっともパフォーマンスの低いディスクへの書込みが完了するまでのログの書込み待機時間が発生するためです。また、ディスク・ドライブ自体でも、パフォーマンスにおいてたびたび“一時的な停止”が発生することがあります。これらのパフォーマンスの悪化がデータベース・パフォーマンスに大きく影響する可能性があります。さらに、フラッシュ・テクノロジーにも、消去サイクルや摩耗平準化によって、同じようなパフォーマンス上の一時的な停止が発生する可能性があります。これらを含む問題に対処する手段がExadata Database Machineで提供されています。

Smart Flash Logは、OLTPワークロードに対応するためのExadataの独自機能であり、Exadataがハードウェア/ソフトウェア・スタック全体での設計の改善によってデータベース・パフォーマンスを最適化していることを表すもう1つの好例でもあります。Smart Flash Loggingを使用するには、Exadata Storage Softwareバージョン11.2.2.4以降と、Oracle Databaseバージョン11.2.0.2（バンドル・パッチ11適用）またはOracle Databaseバージョン11.2.0.3（バンドル・パッチ1適用）以降が必要になります。

Smart Flash Loggingは、Exadata Database Machineのフラッシュ・ハードウェアを利用します。Smart Flash Loggingは単にREDOログをフラッシュ内に配置するだけではなく、二重化、ミラー化されたログ・ファイルも作成して重要な目的のために使用します。フラッシュベースのログをREDOログ・グループに追加するだけでは、これまでに説明した問題、つまり、ディスクでもフラッシュでも、パフォーマンスの低いデバイスへのログ書込みの待機時間がデータベースで発生するという問題は解決されません。

また、最大の可用性を得るためには、ユーザーがREDOログ・グループとミラー化されたログ・ファ

イルを設定する必要があります。Exadata Smart Flash Loggingは、このような構成に対する変更を加えずに、待機時間の短いログ書込みの利点を享受できるように設計されています。基本的に、この機能はデータベース・ソフトウェア構成や、同様に重要となるデータベース・リカバリに対して透過的です。

Smart Flash Loggingは次のように機能します。ExadataはREDOログの書込み要求を受け取ると、ディスク上のREDOログとフラッシュ・ハードウェアに確保されている小さな領域に対してパラレルに書込みを実行します。これらの書込みのいずれかが完了したときに、データベースに対して即座に処理の完了を通知します。ログをホストしているディスク・ドライブの応答時間が長い場合は、Exadata Smart Flash Cacheのログ書込みの応答時間がより短くなります。逆に、Exadata Smart Flash Cacheの応答時間が（摩耗平準化アルゴリズムなどの理由で）一時的に長くなっている場合は、ディスク・ドライブの応答時間がより短くなります。このようなアルゴリズムによってREDO書込みの応答時間の問題が解消され、全体的なデータベースのパフォーマンスが改善されます。

Exadata Smart Flash CacheはREDOデータの永続的な保存領域としては使用されません。あくまで、REDO書込みの応答時間を高速化する目的で使用される一時保存領域です。Exadata Smart Flash Cacheは、REDOデータが安全にディスクに書き込まれるまでデータを保管するためのキャッシュです。Exadata Storage Serverには相当量のフラッシュ・ストレージが内蔵されています。データベースのロギング用に小さな領域が割り当てられており、残りの領域はユーザー・データのキャッシングに使用されます。REDOログのサイジング、二重化、ミラー化に関するベスト・プラクティスや構成は、Exadata Smart Flash Loggingを使用する場合も変わりません。Smart Flash Loggingがクラッシュ・シナリオとリカバリ・シナリオのすべてを処理するため、REDOログを使用した通常のデータベースのリカバリに必要な作業以外の特別な追加管理作業は必要ありません。エンドユーザーの立場から見れば、システムは完全に透過的に動作しており、フラッシュがREDOの一時保存領域として使用されていることを考慮する必要はありません。動作上の違いは、REDOログ書込みの待機時間が常に短くなることだけです。

デフォルトでは、512MBのExadataフラッシュがSmart Flash Loggingに割り当てられます。それぞれのExadataセルに1.6TBのフラッシュがあることと比較すると、小さな投資で大きなパフォーマンス上の効果を得ることができます。このデフォルトの割当ては、ほとんどの状況で十分なサイズです。フラッシュによるREDO書込みの回数と頻度や、Smart Flash Logging用に割り当てられているフラッシュ領域が不十分であるなどの理由で処理できなかったREDO書込みを示す統計情報が管理されます。REDO生成率の高いデータベースの場合や、多数のデータベースを1つのExadata Database Machine上に統合している場合は、Smart Flash Loggingに割り当てるフラッシュのサイズを拡大する必要があります。また、統合環境向けに、Database Machineで稼働するそれぞれのデータベースについて、Smart Flash Loggingの使用を有効または無効に切り替えられるようにExadata I/O Resource Manager (IORM) が機能拡張されています。この機能を使用して、もっともパフォーマンス・クリティカルなデータベースに対してフラッシュを確保できます。

Exadata Smart Flash Cacheが提供するミッション・クリティカルな可用性

Exadata Smart Flash Cacheにはきわめて信頼性の高いハードウェアが使用されていますが、障害が発生する可能性のないハードウェアはありません。フラッシュ・キャッシュを4つのPCIeカード間で分散することで、このリスクは多少軽減されます。いずれかのフラッシュ・カードに障害が発生すると、Exadata Storage Server Softwareは障害を自動的に検出し、障害が発生した部分のフラッシュ・キャッシュをオフラインにします。このプロセス中、Exadataセルは処理を続行し、残りのキャッシュからデータを提供します。このことから、I/O要求を処理するフラッシュ・キャッシュが減ることでパフォーマンスが低下する可能性はあっても、システムは中断することなく実行され、データ損失も発生しません。したがって、Exadataセルをオフラインにしてフラッシュ・カードを交換するために都合の良いタイミングが来るまで、障害が発生したフラッシュ・カードの交換を延期できます。カードを交換すると、Exadata Storage Server Softwareは自動的に新しいカードの存在を検出し、追加されたフラッシュ・キャッシュの使用を自動的に開始します。障害の発生したフラッシュ・カード内に、ディスクにまだ書き込まれていない"ダーティ"ブロックがあった場合、Exadata Storage Server SoftwareはOracle ASMとともに、他のストレージ・セルからミラー・コピーを自動的に取得して、データの最新コピーをリカバリします。

論理フラッシュ・ディスクがフラッシュ上に作成されている状態で、いずれかのフラッシュPCIeカードに障害が発生した場合も、Exadata Storage Server SoftwareとOracle ASMによって障害による影響が最小限に抑えられます。フラッシュ・カードに障害が発生すると、そのカード上にある4つのフラッシュ・セル・ディスクは自動的にオフラインになり、これらのディスクに対するI/Oは別のExadataセル上のフラッシュに保管されたミラー・エクステンツから提供されます。いずれ、Oracle ASMの再バランス化が実行され、影響を受けていないフラッシュ・ディスク間でデータのミラー分割が完了します。障害の発生したカードが交換されると、フラッシュ・ディスクは自動的にOracle ASMディスク・グループに再追加され、再バランス化が実行されて通常の構成に戻ります。

結論

Exadata Smart Flash Cacheは、Exadata Database MachineのOLTP機能を背後から支えます。もっとも要求の厳しいデータベース・アプリケーションにも応える、これまでにないIOPSを実現し、ウェアハウスやレポート作成アプリケーションに対して2倍以上のスキャン速度を達成し、重要なデータベース・ロギング機能も特別にサポートします。Exadata Smart Flash Cacheを通じて、どのデータをキャッシュすべきで、キャッシュをどのように自動管理するかを把握したOracle Databaseは、最初で唯一のフラッシュ対応データベースです。



Exadata Smart Flash CacheとOracle Exadata Database Machine

2013年1月

著者：Ron Weiss、Mahesh Subramaniam

共著者：Caroline Johnston、Juan Loaiza、

Kesavan Srinivasan、Kodi Umamageswaran

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

Copyright © 2013, 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、その内容に誤りがないことを保証するものではなく、また、口頭による明示的保証や法律による黙示的保証を含め、商品性ないし特定目的適合性に関する黙示的保証および条件などのいかなる保証および条件も提供するものではありません。オラクル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。

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。Intel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。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。1010

Hardware and Software, Engineered to Work Together