

Oracleデータベース向け DellTM ストレージ 選択ガイドライン

デル テクニカル ホワイトペーパー

データベース ソリューション
エンジニアリング

作成: デル製品グループ Zafar Mahmood



本ホワイトペーパーは、情報提供のみを目的に執筆されており、誤字脱字や技術上の誤りには責任を負いません。本書の内容は執筆時現在のものであり、明示的または暗黙的を問わず、いかなる内容も保証いたしません。

EMCは、EMC Corporationの登録商標です。Intel、インテル、Xeonは、アメリカ合衆国およびその他の国におけるインテルコーポレーションおよび子会社の登録商標または商標です。Oracleは、Oracle Corporationの登録商標です。

本書では、マークや名前を届け出た実在のもの、もしくは、その製品のいずれかを参照するため、その他の商標、商号を使用している可能性があります。デルは、その他のマークや名称について、商標上の利権に対する要求に一切に応じません。

© 2008 Dell Inc. ©2008 デル株式会社 All rights reserved. (著作権所有)

Dell Inc. から書面による許可を得ずに本書を複製、転載することは、いかなる場合も禁止します。詳細は、デルにお問い合わせください。

目次

データベース向け ストレージ選択ガイドライン.....	4
要旨.....	4
はじめに.....	4
ファイバチャネルSANとIP SAN	4
Dell EqualLogic iSCSI SAN.....	5
Dell EMC CXシリーズのFC SAN.....	6
テスト方法.....	6
ORION (Oracle I/O Numbers)	6
テスト構成	7
テスト結果.....	8
まとめ.....	14
参考資料.....	14

データベース向け ストレージ選択ガイドライン

要旨

デルのエンタープライズ製品群は、様々なストレージ インタフェースとテクノロジーを取り揃えています。今や、この充実したストレージ製品構成により、中小企業から、大手エンタープライズに至るまで、あらゆるお客様のニーズにお応えできるようになりました。これらのストレージ製品には、以下が含まれます。

- SANストレージ: ファイバチャネル技術を使うFC SANか、iSCSI技術を使うIP SAN
- 直接接続ストレージ(DAS): SAS、または、SATAインタフェースを使用
- ネットワーク接続ストレージ(NAS): IPネットワーク経由で、ファイル レベルのデータアクセスを提供

デルは、複雑になりがちなデータセンタを簡素にする「ITのシンプル化」を提唱しており、その一環として、拡充し続けるストレージ製品群からデータベース用製品が容易に選べるよう、ガイドラインを作成することになりました。これらのストレージ選択ガイドラインを参照すれば、お客様のニーズに合わせて的確なストレージ技術が選べいただけます。本ホワイトペーパーは、データベースを運用されるデルのお客様が、ストレージ技術を選ぶときに役立つガイドラインをまとめたものです。これらのガイドラインは、デル データベース ソリューション エンジニアリング チームが、各種のストレージ製品上で実施した、ラボ テストやデータベース ワークロード シミュレーションから導き出しました。お客様がこれらのガイドラインを参考にすれば、適切なストレージ製品やテクノロジーが選択できるだけでなく、運用するデータベース ワークロードに合わせ、費用対効果を最大限に高めることができます。本書は、ストレージ管理者、IT管理者、システム コンサルタントの方を対象としており、その目的は、ストレージの選択により賢明な判断をしていただくことにあります。

はじめに

Oracle Real Application Cluster(RAC)などの典型的なクラスターリング データベース環境の場合、ストレージには、共有ブロック レベルでアクセスする必要があります。単一インタフェースの「直接接続ストレージ(DAS)」ソリューション(例: SCSI JBODアレイ)では、接続先ノードだけに限定されてしまうため、シングル ノードの非クラスターリング データベースでない限り、この要件を満たすことはできません。

- 1つの統一コンソールから、限られた管理能力のみ提供
- 機能性に制限
- ローカル ストレージの「孤島」が各所に形成されてしまい、資産の有効活用が困難
- スケーラビリティに制限

もう1つの低コスト ストレージ オプションとして挙げられる「ネットワーク接続ストレージ(NAS)」は、一般的な用途のファイルシステム共有に向いていますが、データベース環境には適していません。CIFSやNFSなどのファイル サービス プロトコルを使用するNASストレージ テクノロジーは、ファイルシステム レベルの共有とアクセス用に最適化されています。NASストレージは、クラスタ化したデータベース環境に共有ストレージ アクセスを提供するものの、ブロック レベルのアクセスに比べ、性能に劣ります。また、NASファイラーでエクスポートされるファイルシステムは、NASデバイス上のOSに依存するうえ、ファイルシステム キャッシュを迂回して共有ストレージに直接アクセスする「ブロック レベルのダイレクトI/O」も提供しません。ダイレクトI/Oは、Oracle® RACのような、並列アクセスのアクティブ-アクティブ型クラスターデータベース ソリューションにとって、データの不整合を避けるためにも必要不可欠です。さらに、NASソリューションには、次のような特徴があります。

- ブロックI/O処理に起因するファイルの分解/組み立てにより、性能にオーバーヘッドがかかってしまう
- エクスポートしたファイルシステム上では、サポートできないアプリケーションがある(一部のバックアップウィルス対策エージェントなど)
- メーカー独自の技術によって最適化されたNFSアクセスには、特別な配慮やツールが必要になるため、データベースのようなアプリケーションのサポートが制限される
- ウィルス感染/攻撃に弱い傾向がある
- 特別なバックアップ ソフトウェアのサポートが必要(NDMP)

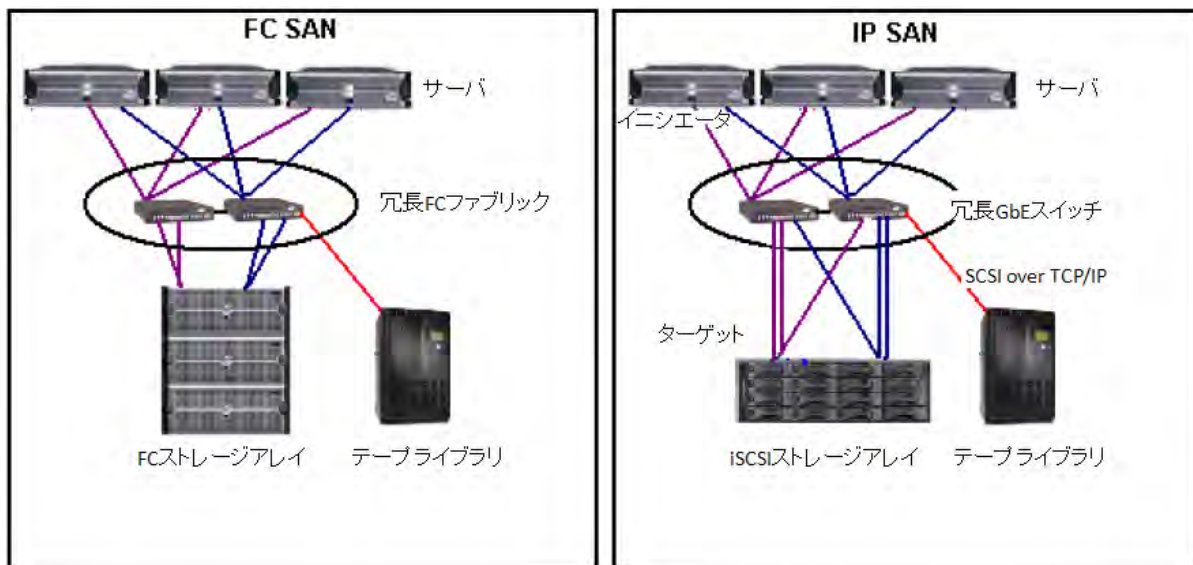
続いて「ストレージ エリア ネットワーク(SAN)」について説明すると、SANには、次の2種類のストレージ ソリューションがあります。

- ファイバチャネル(FC)プロトコルを使用するFC SAN
- iSCSIプロトコルを使用するIP SAN

これらのストレージ技術(特に、2種類のSANテクノロジー)については、以降で詳しく説明します。また、データベース用ストレージを選択するときの一般的なガイドラインは、エンジニアリング テストから導き出しているため、そのテスト結果についても説明します。

ファイバチャネル SAN と IP SAN

Dell | EMC® CX3-20Fなどのファイバチャネル モジュラーアレイでは、コントローラ モジュール(1組の冗長コントローラが搭載可)が、ディスクドライブを格納した複数のシェルフやディスクアレイ エンクロージャ(DAE)を管理します。このコントローラには、2つのフロントエンドFCコネクタ(ホストの接続用)と、2つのバックエンドFCコネクタ(ドライブシェルフの接続用)が付いています。接続するすべてのホストとドライブが、これらのコネクタと、コントローラ上のキャッシュメモリを使用します。SANに接続するサーバ数やドライブ数を増やせば増やすほど、サーバ1台あたりが利用できるバンド幅やパフォーマンスは低下していきます。この制限は、データベース サーバ用のストレージアレイを追加すれば解消できますが、既存のデータベースに、この増加分をストレージ容量として認識させるには、I/Oバスとボリュームをシステム管理者が手動で再構成しなければなりません。これらのファイバチャネル アレイは、スピンドル数を次々と追加していく「スケールアップ」型の拡張を提供するのが特長で、これにより、高性能なファイバチャネル インタフェースを存分に活用することができます。たとえば、Dell|EMC CX 3-20の場合、1台のストレージ プロセッサエンクロージャ(SPE)で、最大120個のスピンドルをサポートします。



一方、Dell EqualLogic™ PS5000XVのように、アレイのグルーピングをサポートするiSCSI SANで容量を増やすには、ドライブアレイ全体と、それに伴うコントローラ、キャッシュ、ホスト接続を追加して、SANで利用できるようにします。アレイのグルーピングとは、複数のストレージアレイを1つのグループにまとめて連動させる、ストレージアレイ機能のことです。これらのアレイがホストしているストレージボリュームは、グループ内のメンバー状況が変わるたびに、自動で再分配されます。これらのアレイは、アレイメンバーシップアーキテクチャと呼ばれる概念を採用しているため、「スケールアップ」ではなく、「スケールアウト」で必要なストレージ容量、IOPS、バンド幅を拡張する方が容易です。複数のアレイを1つのグループとして管理し、グループ内の利用可能な全ドライブにボリュームを分散させることで、性能向上も容量拡張もスムーズに行えます。Dell EqualLogic製品は、インターネットプロトコル(IP)上で稼働するiSCSIの特性を活かし、グループ全体を1つの仮想IPアドレスに回答させることでこの仕組みを実現しています。ファイバチャネルは、レイヤ2(第2層)プロトコルのため、FCアレイではこの技術が使えません。さらにiSCSIプロトコルは、双方向認証を使うCHAPと暗号化もサポートします。LUNマスキングは、iSCSIとファイバチャネルの両方でサポートします。

よくある誤解に、「4Gbpsファイバチャネルは、Gigabit Ethernet上で稼働するiSCSIより4倍高速」というものがあります。確かに理論上のバンド幅だけを見れば、4倍かもしれません。速度や実際のIOPS値(1秒あたりのI/O処理数)を比べると、こうはなりません。速度とバンド幅とは、根本的な違いがあります。物理的なデータリンクには、高速道路と同様、制限速度があります。そこで、ファイバチャネルのような高速ネットワークを、より車線の多い高速道路に喩えて考えてみましょう。車線が多い分、渋滞せずに多くの交通量を流すことができますが、それぞれの車線を見れば、移動時間はどれもほぼ同じです。つまり、FCデータリンクでも、Ethernetデータリンクでも、移動速度に変わりはありません。典型的なデータベース環境では、バックエンドストレージに十分なスピンドル数を搭載していないため、FCリンクがせっかく広いバンド幅を提供していても、それを限度一杯にまで活用していないのが現状です。

最も重要なのは、アプリケーションのI/O要件を明らかにし、スレーブ重視のアプリケーション(例:意思決定支援システム=DSS)なのか、それとも、IOPSや低レイテンシー重視のアプリケーション(例:OLTP)なのか判断することです。これらが最終的に、「導入・実装すべきなのはiSCSI SANか、それともFC SANなのか」を決める要因となります。どちらのSANであっても、データベースワークロードの種類に合わせて、最適にサイジングおよび構成することが可能です(詳細は後述)。たとえば、DSSアプリケーションに必要なスレーブが達成できないiSCSIアレイがあったとしても、Dell EqualLogicストレージアレイのようなiSCSI SANに新しいメンバーを追加し、ホストからのイニシエータ数やI/Oパス数を適切に拡張すれば、このボトルネックを解消することができます。特にOLTPのようなワークロードを扱うデータベースでは、TCO(総所有コスト)の低いiSCSIの強みが存分に活かれます。

Dell EqualLogic iSCSI SAN

多くのデータセンタでIP SANの採用が広がっていますが、その人気の秘密は、既存のEthernetインタフェースが活用できることにあり、それゆえ、現行インフラストラクチャにも統合しやすく、管理も容易だからです。管理しやすくTCOも低いIP SANは利点が多いのに、導入に二の足を踏むのは、データベース処理性能に疑問があるからです。テスト結果をまとめた後述のセクションでは、これらの疑問を払拭しており、IP SANまたはFC SANの選択に、よりの確かなガイダンスが得られます。

EqualLogic PSシリーズのiSCSIストレージアレイは、「ピアストレージ」と呼ばれるアーキテクチャを採用しています。このアーキテクチャでは、ストレージプール内の全アレイメンバーが「ピア(仲間)」として機能するように設計されているので、メンバー同士が連動しながらディスク容量を提供し、ワークロードを均等に分配することができます。PSシリーズのIP SANは、ストレージ容量が簡単迅速に拡張でき、拡張時にデータの可用性を損なうこともありません。SANにアレイメンバーを追加していくたびに、PSシリーズが迅速に拡張していき、追加するアレイ数に応じて直線的にパフォーマンスを伸ばすことができます。ストレージ管理者は、既存のストレージプールに新しいストレージアレイやメンバーを追加する際、既存のボリュームにデータブロックを再分配させることができます。この場合、新しいメンバースピンドルは自動的にプール内に取り込まれるため、マシンをダウンする必要はありません。

Dell EqualLogicストレージアレイは、ストレージの仮想化などの機能も提供しており、ストレージグループ内の全ストレージアレイメンバーを、1台のマネジメントシステムで統合管理することができます。ストレージの仮想化は、ストレージグループ、プール、メンバーという概念のもとに実現しているため、従来型ストレージアレイのようにデータボリュームをディスクにマッピングする必要がなく、全メンバーにわたってボリュームデータを最適に配置することができます。EqualLogic iSCSI SANは、これらの独自機能をサポートするため、特にOLTPワークロードを筆頭に、小〜中規模企業のデータベースストレージニーズに応えることができます（詳細は、後述のテスト結果を参照）。また、高スループットが必要なDSSワークロードでも、ストレージプールで利用するグループメンバー数と、データベースホストからのインシエータ数やI/Oパス数を注意深くサイジングすれば、Dell EqualLogic IP SANを活用することができます。Dell EqualLogic iSCSI SANには、その他にも次のようなメリットがあります。

- ストレージの仮想化：iSCSIツールは、ファイバチャネルより優れています。これは、FCの場合、それぞれに独自のワールドワイド名（WWN）が付けられており、物理ディスク、ポート、論理ボリューム、インシエータポートへのマッピングがいずれもハードコード化されているからです。TCP/IPに基づきiSCSIは、DNSが利用できるので、アドレスを仮想化することができます。
- 迅速なセットアップ：EqualLogicツールは、セットアップ時間の短縮など、数々の利点があります。EqualLogic SANは、ファイバチャネルSANよりはるかに素早くセットアップできます。
- 素早い習得：iSCSIは、広く普及した実績あるテクノロジーに基づいています。
- Ethernetインフラストラクチャ：一般にすべてのデータセンタには既にEthernetが構築済みでし、FC設備に比べて初期投資もはるかに低いことから、TCOも低くなります。

Dell EMC CX シリーズの FC SAN

Dell EMC CXファイバチャネルSANは、最大I/Oスループット、キャッシュサイズ、スピンドル数、冗長パス数、容量のサポート範囲がiSCSIより広がります。しかし、初期投資は決して安価ではなく、FC SANの管理には専門のスキルが欠かせません。また、冗長構成の場合、2台のFCスイッチと、各ノードにつき2個のHBAが必要になります。Dell EMC FC製品ラインは、MirrorView、AccessLogic、SnapViewなどの、より充実した管理ツールや付加価値ツールがご利用いただけるため、エンタープライズ向けのデータベースストレージとして位置付けられますが、価格はその分高額となります。Dell EMC FC SANが、エンタープライズデータベースソリューションに欠かせないストレージと見なされるのは、以降に示すような特長があるからです。

FCは現在、SANの導入に最も採用されているテクノロジーです。ストレージエリアネットワークの利点については長い間認識されてきましたが、実際にSAN（ネットワーク化した共有ストレージ）の普及が進んだのは、FCが登場してからです。

- 高い拡張性 — FCインタフェースは広範なアドレス能力を提供し、最大1,600万アドレスをサポートします。
- 高性能 — FCは、特にI/Oパフォーマンス処理を意図した設計となっており、現在、4 Gbpsという高スループットを達成するばかりか、近い将来、8 Gbpsの達成も確実視されています。さらにFCは、停滞知らずのクレジット式フロー制御により、宛先デバイスが達成し得る最大受信性能で、データを伝送することができます。
- 成熟した実績あるテクノロジー — FCの開発は1988年にまで遡り、高信頼性のインタフェース技術として長年、採用され続けてきました。長期の実績に裏打ちされた設計により、FCは、高性能エンタープライズデータセンタの標準テクノロジーとなっています。
- リンク速度の自動ネゴシエーションにより、投資を保護 — FCのリンク速度を自動認識・設定できるので、既存の1 Gbpsや2 Gbps環境に隔たりなく統合できますし、構成をアップグレードした環境では、すぐに性能を向上することができます。
- 長距離ケーブル — シングルモードのFCメディアは各種ありますが、たとえば、1,550 nmの長波長レーザー送信機に対応するものは、個々のデバイス間を、最大50,000メートルのケーブルで連結することができます。

FC SANの実装について全般的な効果や投資を検討するときは、潜在的な制限事項についても考慮してください。

- 習得が困難 — FC SANは、十分な予算や、設計/導入/管理に詳しい専任のIT担当者が確保できない企業にとって、敷居の高い技術となりがねません。
- 投資の負担 — FC SANの導入では、総所有コスト（TCO）が跳ね上がることがあります。これには、FC機材の購入コストだけでなく、運用管理、メンテナンス、スタッフのトレーニングにかかる費用も含まれます。

テスト方法

今回のテストでは、ORION（Oracle I/O Numbers Calibration）ツールを使用して、各種のデータベースI/Oワークロードを対象に、いくつかのストレージアレイ製品の性能とスケーラビリティを測定しました。

ORION (Oracle I/O Numbers)

ORIONとは、Oracleデータベースのディスクサブシステム特性をシミュレーションするツールです。OLTP、DSSや、これらを組み合わせたワークロードなど、Oracleデータベースの様々なI/Oパターンを再現することができます。

ORIONは、各種のI/Oワークロードをサポートしますが、今回は、スモール ランダムI/O、ラージ シーケンシャルI/O、ミックス(スモールとラージの組み合わせ)ワークロードをテストしました。

スモール ランダムI/Oワークロードは、ほとんどのI/O処理が乱読み込み、乱書き込みとなる、典型的なOLTP(オンライン トランザクション処理)アプリケーションを再現するもので、I/Oサイズは、データベースのブロックサイズ(通常、8 KB)と同じです。OLTPデータベース用のストレージ サブシステムでは、主に、IOPS(1秒あたりのI/O処理数)とI/Oレイテンシを使用して、性能を比較します。一方、ラージ シーケンシャルI/Oワークロードは、データ ウェアハウス(DW)アプリケーションに代表される、クエリーベースのビジネス インテリジェンス システムを再現したものです。DWアプリケーションは、複数の処理待ちI/O(サイズ=1 MB)から構成される、順次読み込み/順次書き込みストリームを生成する傾向があり、主な性能評価基準は、ストレージ サブシステム全体のデータ スループット(単位はMBPS=1秒あたりのメガバイト数)となります。

ORIONは、ワークロードの種類別に、様々なレベルのI/O負荷をかけ、MBPS、IOPS、I/Oレイテンシなどの性能を測定します。負荷レベルは、処理待ちの非同期I/O数で表されます。たとえば、スモール ランダム ワークロードでは、処理待ちのスモールI/O数で負荷レベルが決まり、ラージ シーケンシャル ワークロードの場合は、ストリームあたりの処理待ちラージI/O数で負荷レベルが決まります。

今回は、1台のCX3-20F、1台のPS5000XVアレイ、2台のPS5000XVアレイを対象に、3種類のORIONテストを使用しました。その3種類とは、次のとおりです。

- OLAPのフル テーブル スキャン: ラージ シーケンシャルI/Oの読み込み専用ワークロード
- OLAPのパッチ ロード: ラージ シーケンシャルI/Oの書き込み専用ワークロード
- 典型的なOLTP I/Oワークロード: スモール ランダムI/O(書き込み=80%、読み込み=20%)

各テストにかけた負荷レベルは、最大30個の処理待ちスモールI/Oか、シーケンシャル ストリームあたり最大8個の並列ラージI/Oです。いずれのテストも、1つあたり70 GBのロウ ボリュームを3個使用して実施しました。各テストで、Oracle ASM(Automatic Storage Management)によるストライピング効果をシミュレーションしています。テストは毎回、ストレージ サブシステムを接続する1台のノードから呼び出しました。

テスト構成

今回のテストでは、Dell EMC CX3-20FとDell EqualLogic PS5000XVを使用し、14ドライブのRAIDに、10個のボリュームを構成しました。両ストレージ エンクロージャとも、15,000回転のドライブを採用しています。IP SANソリューション(Dell EqualLogic PS5000XV)で使用したのは、SASディスクで、FC SANストレージで使用したのは、ファイバチャネル ディスクです。いずれのストレージ製品も、それぞれが提供するマルチパスI/O技術を利用することで、ホストから複数の冗長パスを構成しました。たとえば、CX3-20 SANの場合は、デュアルHBAと冗長FCスイッチを構成し、マルチパス ソリューションにEMCのPowerPathを使っています。また、PS5000XVの場合は、Linuxから標準提供されるDMデバイス マッパーを使って、ラウンドロビン(順繰り)方式のI/O負荷分散を構成しています。両ストレージとも、各70 GBのロウ ボリュームを3個使ってテストしました。

手順は、最初に1台のEqualLogic PS5000XVアレイでテストし、次に、もう1台のPS5000XVアレイを追加して、スケーラビリティを観察しました。IP SAN構成では、「iSCSI SANトラフィックを他のネットワークから分離する」という大前提のベストプラクティスに従うため、2台のDell PowerConnect™ 5424 Gigabit Ethernetスイッチを使って、ホストサーバとストレージ サブシステムを接続しています。この2台のGigabit Ethernetスイッチ間には、8 Gbpsのリンク アグリゲーション グループを構成することで、広バンド幅のスイッチ間ネットワークを実現しました。PS5000XVのテストは、フロー制御、スパニング ツリー プロトコル(STP)のPortFast機能、ロードキャスト&マルチキャスト ストーム制御、ジャンボ フレームを有効に、また、ユニキャスト ストーム制御とPC5424のiSCSI最適化機能を無効にしており、いずれも、ネットワークのベストプラクティスに従った設定です。

テストに使用したサーバはDell PowerEdge 2950 IIIで、それぞれ2基のデュアルコア インテル®Xeon® プロセッサ2.66 GHzと、32 GBのメモリを搭載しました。PS5000XVのテストでは、iSCSIトラフィックの送信用に、2個のインテル® PRO/1000 Gigabit NICポートを使って最初の2つのイニシエータを構成しています。ストレージ グループに他のメンバーを追加するスケーラビリティ テストでは、あと2つのイニシエータを追加構成してからテストを実施しました。

スモールI/Oを扱うテストケース(OLTPワークロードを処理するデータベース)はいずれも、iSCSIがFCに匹敵、または、上回る性能を達成しました。FCがiSCSIを上回ったのは、サイズが1MB以上の、ラージ シーケンシャルI/Oのみでした。Dell EqualLogicストレージ アレイでは、メンバーを追加するスケーラビリティ テストも実施しており、その結果、iSCSIに特有のスループット制限は、EqualLogicストレージアレイならではの能力を活用することで緩和できることが確認されました。ストレージ アレイとイニシエータを追加すると、I/Oスループットを向上することができ、その性能は、ファイバチャネルに匹敵します。これらのテスト手順とストレージ構成の詳細を、以降に示します。

ORIONテスト構成		
サーバ		1台のDell PowerEdge 2950 IIIサーバに次を搭載: - クアッドコア インテル® Xeon® プロセッサ2.66 GHz x 2 32 GB RAM 1 Gb インテル® PRO/1000 NIC ポート (iSCSIトラフィック用) - QLE2460 x 2 (FCテスト用)
外付ストレージ	PS5000XV	オリジナル構成 (1メンバー、2イニシエータ) - Dell EqualLogic PS5000XV (14ディスク) x 1 - RAID 10 (メンバーごとに2個のスペアディスク) 70 GBのボリューム x 3 テスト1 (2メンバー、2イニシエータ) - Dell EqualLogic PS5000XV (14ディスク) x 2 - RAID 10 (メンバーごとに2個のスペアディスク) 70 GBのボリューム x 3 テスト2 (2メンバー、4イニシエータ) - Dell EqualLogic PS5000XV (14ディスク) x 2 - RAID 10 (メンバーごとに2個のスペアディスク) 70 GBのボリューム x 3
	CX3-20	- Dell EMC CX3-20 (14ディスク) x 1 - RAID 10 70 GBのボリューム x 3 - PowerPath 5.1
iSCSIネットワーク		24ポートのDell PowerConnect 5424 Gigabit Ethernetスイッチ x 2 (iSCSI SAN用) 1Gbのスイッチ間リンク x 8 (リンク アグリゲーション グループを構成)
FCスイッチ		Brocadeファイバチャネル スwitch M4400 x 2
ソフトウェア		ORION: Oracle I/O Numbers – バージョン 11.1.0.0.0

テスト結果

注: すべての結果は、正規化しているため、実際の測定値ではありません。これらのデータを提供する目的は、ラボ環境における特定の構成で、各種のストレージテクノロジーおよびインタフェースを比較することのみにあります。したがって、ストレージ アレイやストレージ テクノロジーの最大性能を示したわけではありません。

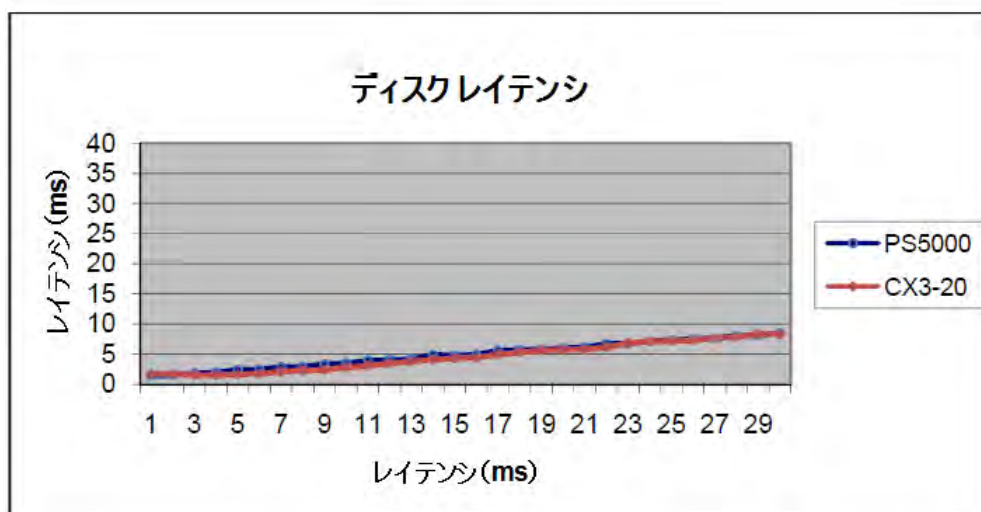
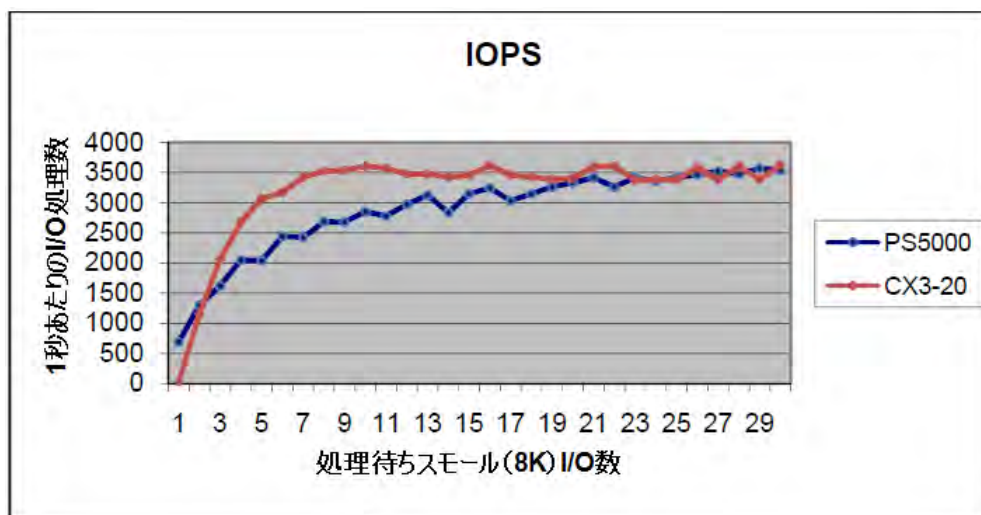
以降のテスト結果をご参照いただければ、各ストレージ テクノロジーで様々なデータベース ワークロードを処理したときの性能が明らかになり、これにまつわる疑問も解消されると思います。また、ストレージに関する一般的な推奨事項も紹介しました。今回のテストでは、Oracle ORIONベンチマーク ツールを使用して、各種のデータベース ワークロードをシミュレーションしており、具体的には、ほぼ同等に構成した2種類のストレージ アレイ上で、スモール ランダム リードライト (OLTP)、ラージ シーケンシャル ライト (バッチ ロード)、ラージ シーケンシャル リード (DSS)、スモール/ラージを組み合わせたランダム I/O (OLTPとDSSのミックス) を発生させました。これらのテスト結果は、見やすい表とグラフにまとめ、ワークロードごとの推奨事項も紹介しています。

本テストの目的は、次の疑問に答えることです。

- それぞれのストレージ製品は、各種のデータベース ワークロードに対し、どのような処理性能を示すか？
- iSCSIとFCの比較: iSCSIストレージ アレイは、データベース向けストレージとして実用に耐えるか？
- もしそうだとすれば、どんなデータベース ワークロードだったら、データベース I/O要件を損なうことなく、iSCSIストレージ アレイが有効活用できるか？
- スケーラビリティ: iSCSIベースのDell EqualLogic PS5000XVエンクロージャは、既存のストレージ プールにメンバーを追加することで、I/OスループットやIOPSを向上することができるか？

典型的なOLTPクエリーを処理するデータベース ワークロード

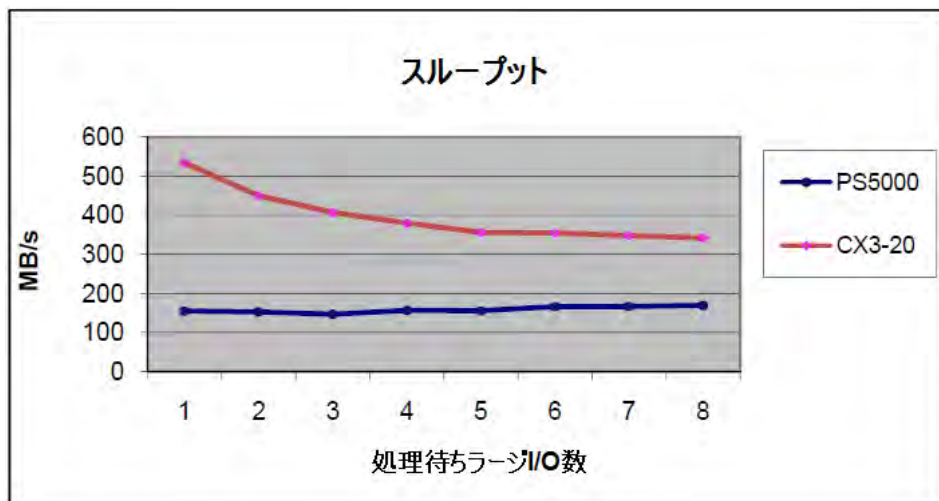
- ・ スモール(8K)ランダム I/O (読み込み=20%、書き込み=80%)



ここに示したとおり典型的なOLTPワークロードでは、iSCSIとファイバチャネルの両方とも、テスト全体で高性能を発揮しており、レイテンシも一貫して10ミリ秒未満に収まっています。トランザクション主体のワークロードを処理するデータベースでは、レイテンシを15ミリ秒未満に抑えることが推奨されます。上記のテストは、ストレージ プールに1つのPS5000XVアレイ メンバーを構成して実行したものです。

ランジ シーケンシャル リードを処理するデータベース ワークロード(フル テーブル スキャン)

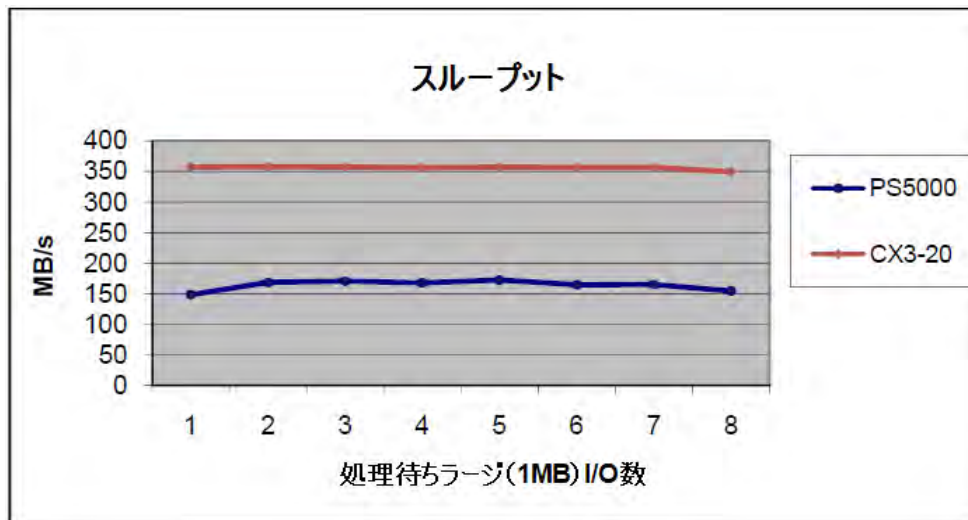
- ・ ランジ(1MB)シーケンシャル リード(書き込み=0%)



上記のグラフを見ると、大きな順次読み込みI/Oを処理するワークロード(データウェアハウスや意思決定支援システムなど)では、ファイバチャネルが圧倒的に優勢です。これらの処理では、高いI/Oスループットが求められますが、iSCSI SANでは、Gigabit Ethernetの能力に限界があるため、それがバンド幅の制限につながりました。このテストでは、1メンバーのPS5000を使用しました。後ほど、この構成にメンバーを追加して、iSCSIのスループット制限が克服できるか検証した結果を示します。

ランジ シーケンシャル ライトを処理するデータベース ワークロード(バッチ ロード)

- ・ ランジ(1MB)シーケンシャル ライト(読み込み=0%)



上記のグラフは、バッチ ロードを処理する典型的なデータベースのシミュレーションとして、大きな順次書き込みを実行したときの結果です。この場合、ファイバチャネルSANの方がはるかに高性能で、iSCSIソリューションのI/Oスループット性能とは、ほぼ100%もの差がつかしました。

次の結果は、ファイバチャネルとの差を縮めるため、Dell EqualLogicにメンバーを追加し、また、ホスト側にイニシエータを追加して測定したものです。多くのデータベース ソリューションは、たとえば、オンライン トランザクション処理を実行しながら、フル テーブル スキャン経由でレポート処理を行うデータベースのように、複数のワークロードを実行します。したがって、レポートや負荷の重いバッチ処理などは、ピーク時を避けて実行し、営業時間中のOLTP性能を最大限に伸ばすよう推奨します。このような工夫により、OLTP全体の性能が大幅に改善されます。

Dell EqualLogic PS5000XV のスケーラビリティ テスト:

前述のとおり、スタンドアロンのDell EqualLogic PS5000XVアレイは、OLTP環境で良好な性能を発揮していました。一方、OLAP/DSSを扱うシングル環境では、iSCSIスレーブが、iSCSIターゲットとイニシエータのバンド幅(GigE)に制限されてしまい、理論上、約80~90 MBPSが上限となります。しかし、機能性に優れたPS5000XVのようなiSCSIストレージは、標準構成の容量を拡張(スケールアウト)することで、このスレーブ制限を克服することができます。このスレーブ制限を解消し、広バンド幅が求められるDSSなどのワークロードでiSCSIを利用するには、複数のPS5000XVユニットを使い、それに合わせてイニシエータ数も追加してください。

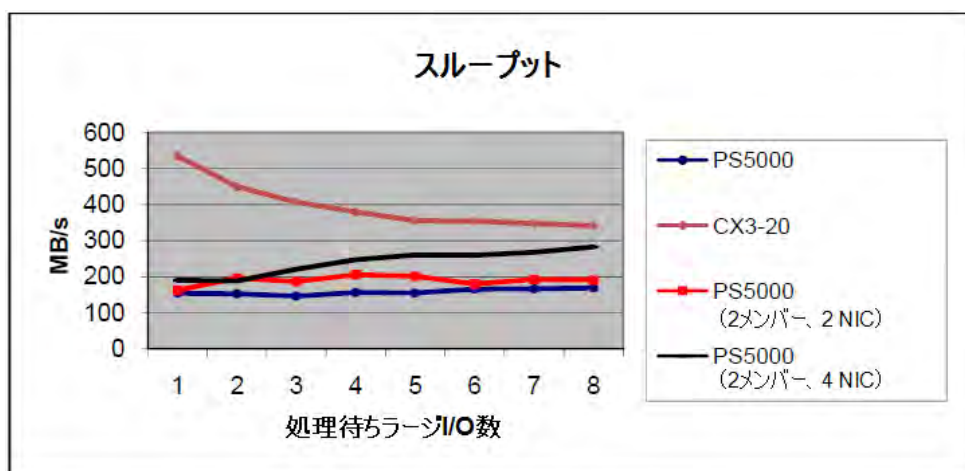
今回実施した2つのスケーラビリティ テストの詳細は、以下をご参照ください。

- ・ スケーラビリティ テスト1: もう1つのメンバーを追加して、同じテストを再実行しました。
- ・ スケーラビリティ テスト2: もう1つのメンバー追加し、さらに、あと2つのイニシエータを追加して、バンド幅を拡張しました。別の拡張方法として、PSエンクロージャの追加に伴い、もう1台のノードを追加することも可能です。

このテスト結果を、以下のグラフに示します。

ラージ シーケンシャル リードを処理するデータベース ワークロード(フル テーブル スキャン)

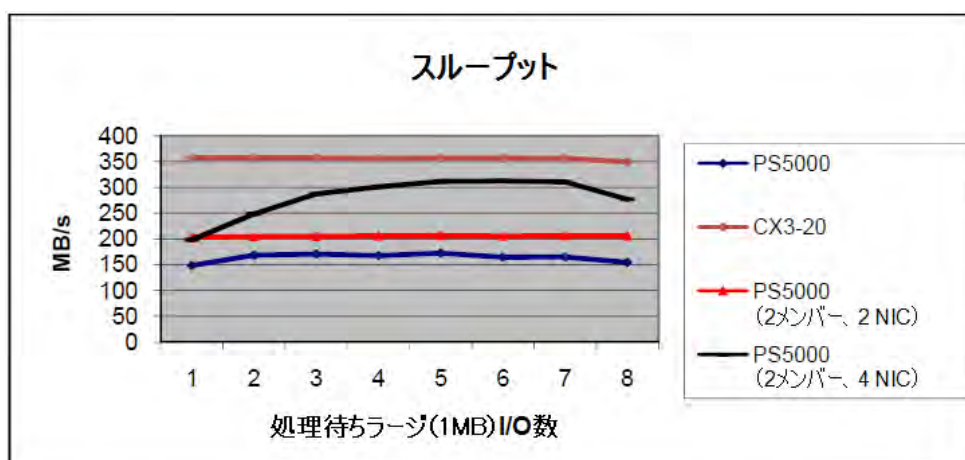
- ・ ラージ(1MB)シーケンシャル リード(書き込み=0%)



上記のグラフは、ファイバチャネル(CX3-20)とiSCSI SAN(PS5000XV)の性能を比較したもので、さらに、iSCSIアレイの場合は、拡張前と拡張後(iSCSIアレイにメンバーとイニシエータを追加)の性能も比較できるようになっています。このグラフが示すとおり、イニシエータも同時に追加しない限り、スレーブットは向上しませんでした。これは、メンバーの追加によって増えたスレーブット能力を維持するのに、十分なイニシエータ数が必要だからです。

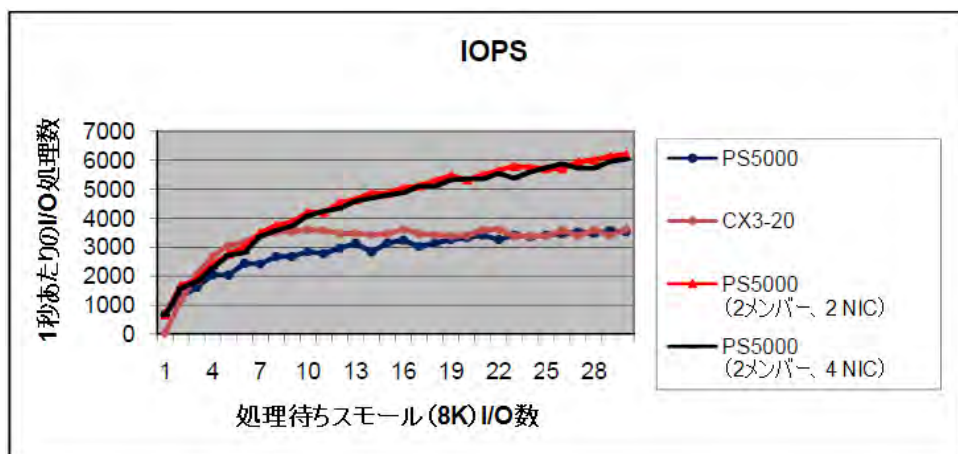
ラージ シーケンシャル ライトを処理するデータベース ワークロード(バッチ ロード)

- ・ ラージ(1MB)シーケンシャル ライト(読み込み=0%)



典型的な OLTP クエリーを処理するデータベース ワークロード

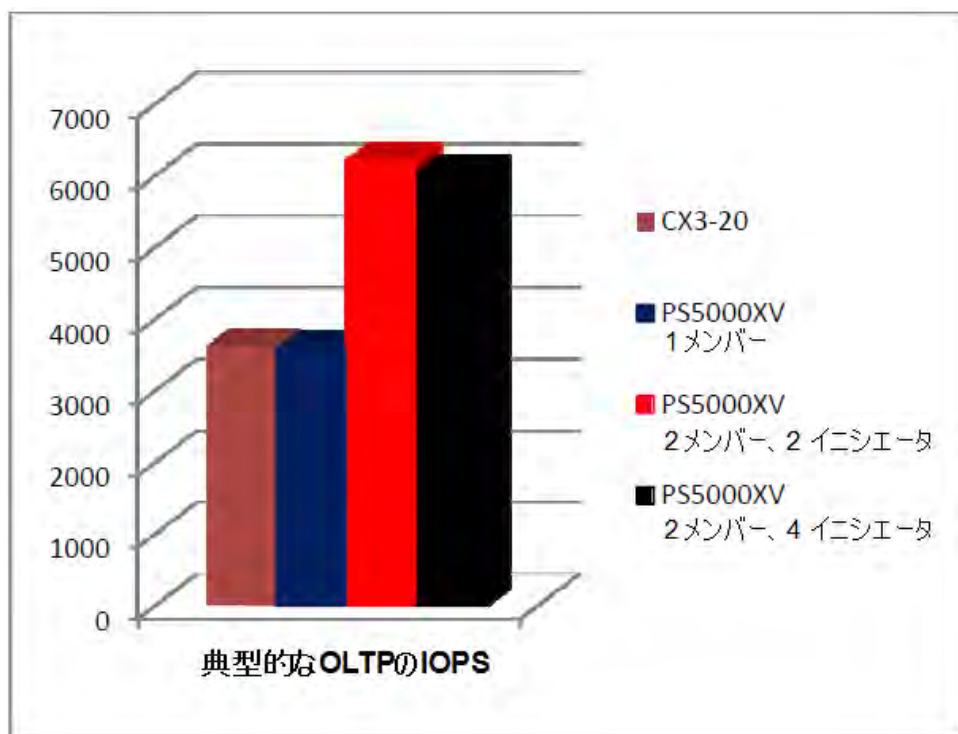
- ・ スモール(8K)ランダム I/O (読み込み=20%、書き込み=80%)

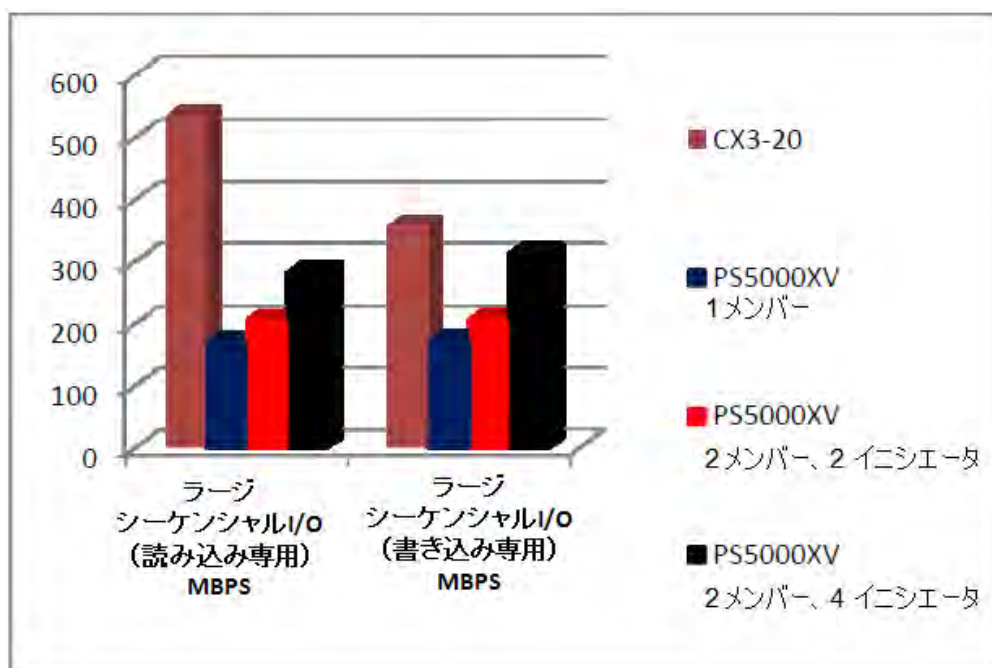


これまでのグラフから、興味深い傾向が読み取れます。iSCSIストレージ アレイ特有のスレーブ制限は、Dell EqualLogicストレージアレイのピア アーキテクチャを存分に活用することで — つまり、ストレージプールにアレイ メンバーを追加し、その分増えたiSCSIセッションを各ホストから呼び出せるようイニシエータも追加することで — 解消できることがわかります。

総評:

以降の図表は、これまでの比較結果をまとめたものです。





これまでのテスト結果を見ると、iSCSIストレージは、ボトルネックが(イニシエータのバンド幅ではなく)スピンドル数に起因するOLTP型のワークロードで、非常に優れた性能を達成しています。一方、ワークロードがDSSのように大きな乱読み込み/乱書き込みを扱い、大きな順次処理も伴う場合は、ファイバチャネル ストレージを使用すべきです(ただし、メンバーとイニシエータの追加によって、iSCSIのバンド幅制限が解消できる場合は、この限りではありません)。この種のワークロードでは、イニシエータとターゲットのバンド幅(=現状では、GigEインタフェース数)がボトルネックとなります。下表は、シミュレーションしたデータベース ワークロードの種類別に、推奨するストレージ テクノロジをまとめたものです。Dell EquiqlLogic PS5000XVでは、次のような傾向が見られました。

- － ホスト側に2つのイニシエータしかない場合、iSCSIのスループットは、最大で約2,200 MBPSとなります。このときターゲットは、約 $(3+3) \times 80 = 480$ MBPSというデータ転送速度が提供できますが、2つのイニシエータだけでは、最大200 MBPSのスループットしか維持できません。IOPSは、イニシエータNICを増やさなくても、直線的にスケーラビリティを伸ばすことができます。この件は、OLTPワークロードのテストで実証されており、2つのイニシエータで十分でした。
- － ストレージ メンバーを追加し、さらにホスト イニシエータを追加したときは、スループットがラージ シーケンシャル ライトでほぼ81%向上しており、ラージ シーケンシャル リードでも66%向上しています。
- － この結果から、DSSワークロードの場合、利用可能なバンド幅を拡張するには、グループ メンバーの追加だけでなく、イニシエータも増やすべきであることがわかります。

ワークロード		PS5000	CX3シリーズ
典型的なOLTP (書き込み=80%)	シングル エンクロージャ	✓✓	✓✓
	スケールアウト	✓✓✓	✓✓
ラージ シーケンシャル リードライト (DSSパッチ)	シングル エンクロージャ	✓	✓✓✓
	スケールアウト	✓✓	✓✓✓
最大スピンドル数		メンバーあたり16 ストレージプールあたり128	CX3-20=120 CX3-40=240 CX3-80=480

まとめ

クラスタリング データベースに適切な共有ストレージ インフラストラクチャを選択するのは、容易なことではありません。データベース処理性能は、ストレージ インタフェースとテクノロジーによって大きく変わるため、データベースのワークロード要件を入念に調べてから、決断を下す必要があります。本番環境へ移行する前に、ORIONなどの適格なテスト ツールや最適なサイジングを通して、ストレージ特性をしっかりと把握しておくことが大切です。また、スピンドル数、I/Oパス数、負荷分散テクノロジー、バックエンド ストレージ内のコントローラ数も注意深くサイジングし、バランスの取れたシステムを構築してください。システム構成に偏りがあると、そこからボトルネックが発生し、I/Oの制限や深刻な停滞を招くため、データベースはストレージ アレイの高い能力を存分に活かすことができません。ラボ結果を見ると、PS5000XV ストレージ サブシステムの性能は、一貫してアレイを加えることに直線的/ほぼ直線的に向上することがわかりました。Dell EMC CXシリーズのFC SANソリューションは、いかなるデータベース ワークロードであろうとニーズに応えられますが、Dell EqualLogicのIP SANソリューションも、IOPS重視であろうと、スループット重視であろうと、適切なグループ メンバー数とホスト インシエータ数を選択すれば、データベース ワークロードの種類を問わず、ニーズに応じてスケールアウトすることができます。このように、Dell EqualLogic PSシリーズのiSCSIストレージ システムが、ビジネスの成長に合わせてIP SANを拡張し、ストレージ性能を伸ばせるのは、特許取得済みのピア ストレージ アーキテクチャを採用しているためです。

参考資料

1. ORIONユーザーガイド:

<http://www.oracle.com/technology/software/tech/orion/index.html>

2. 「Inside the EqualLogic PS Series iSCSI Storage Arrays」(John Joseph, Eric Schott, Kevin Wittmer著、『Dell Power Solutions』2008年2月号より):

<http://www.dell.com/downloads/global/power/ps1q08-20080249-EqualLogic.pdf>

3. 『Interface decisions: SAS, FC, or iSCSI?』(作成: デル製品グループ):

http://www.dell.com/downloads/global/products/pvaul/en/fisc_i_fc_sas_interface.pdf