



ビッグデータアプリケーションを支えるフラッシュベースストレージ

第 1 部：ビッグデータストレージについて理解を深める

Scott Harlin

目次

- 1 はじめに
- 2 ビッグデータの概要
- 3 ビッグデータの使用例
- 4 ビッグデータ分析
- 5 ビッグデータにおける SSD のメリット
- 6 まとめ

1 はじめに

IT 管理者であれば、ビッグデータを中心とした重要なアプリケーションのために、フラッシュベースのソリッドステートストレージ（以降 SSD）の導入を検討していることだろう。ビッグデータとは、従来のデータベースやソフトウェア技術※1では処理できない、大容量かつ、頻繁に更新される構造化データおよび、非構造化データの総称で、企業の「データ・ディスカバリ」や「データ分析」※2のソースになる、デジタルデータの膨大なコレクションだ。ビッグデータの総量は、企業によって異なるため、正確に数値化することはできないが、一般的には数十テラバイトから数ペタバイトにおよぶ。

また分析的な観点から見ると、大量のデータやさまざまなデータ型を分析する能力は、隠されたデータパターンや未知の相関関係、企業の競争力を強化する情報などを明らかにし、よりよい意思決定や、収益を増加させる効果的なマーケティングにつながることになる。ちなみに McKinsey Global Institute（MGI）のリポートによると、ビッグデータにより生み出される企業のメリットには以下の5つがある※3：

- 透明度が高く使用頻度の高い情報の作成
- 正確かつ詳細な情報収集による、問題点の洗い出しとパフォーマンスの向上
- 顧客の詳細なセグメント化による、ニーズに合わせた製品・サービスの提供
- 高度な分析機能を活用したビジネスの意思決定
- 新製品や新サービスの品質改善

ホワイトペーパー第 1 部では、エンタープライズ向けストレージの導入についてより深く理解するため、ビッグデータアプリケーションに対する概念と特徴を紹介するとともに、SSD がビッグデータモデルにどのように適合するかを解説します。

ホワイトペーパー第 2 部（別文書）では、ビッグデータアプリケーションによるデータ駆動型分析に必要な、大規模データセットの超高速処理を実現する、OCZ のエンタープライズ SSD および、ソフトウェアソリューションについて解説します。
なお第 2 部で扱う OCZ Storage Solutions の製品は以下の通り：

- Intrepid 3000 SATA SSD シリーズ
- Z-Drive 4500 PCIe SSD シリーズ
- Windows Acceleration (WXL) ソフトウェア
- VXL 仮想化ソフトウェア
- ZD-XL SQL Accelerator
- StoragePeak 1000 Central Management

2 ビッグデータの概要

IDC の市場リサーチ※4 によると、ビッグデータは今後 5 年間で急速に成長し、世界的には数十億ドル規模に拡大するとされる。さらに重要なのは、ビジネス全体がビッグデータ駆動型へ移行しているという点だ。ちなみに IDC の定義によれば、ビッグデータとは、高頻度に取得・発見・分析した膨大かつ多種多様なデータから、経済的に意味のある値を導き出すために開発された新技術、またはアーキテクチャで、市場は以下の 3 つの主要セグメントで構成されている：

- 1.インフラストラクチャ：OCZ が提供している SSD などの外部ストレージや、サーバーコンポーネント（内部ストレージ、メモリ、ネットワークカードなど）、データセンターのネットワークインフラ（スイッチ、ネットワークコントローラ、物理層デバイスなど）、クラウドインフラサービスなど
- 2.ソフトウェア：情報管理ソフトウェアやデータ分析・ディスカバリソフトウェア、ビッグデータ向けアプリケーションなど



IDC では、ビッグデータ関連市場は、年平均成長率 27%で成長し、2012 年の 98 億ドルから 2017 年には 324 億ドルに達すると予想している。

3.サービス：コンサルティング、インテグレーションサービス、ストレージサービス、セキュリティサービス、ハードウェアやソフトウェアサポート、トレーニング、アウトソーシングなど、ビッグデータの実装に関連するサービス全般

IDC では、ビッグデータ技術/サービス市場は、2012 年の 98 億ドル から 2017 年には 324 億ドル規模まで拡大すると予測している。これは年平均成長率 27%、IT 市場の成長率の約 6 倍にあたる。

またビッグデータを定義する際には、非構造化データと多構造化データの情報が混在していることを、理解することが重要になる※5：

●非構造化データ：従来のデータベースやデータモデルでは分類できないデータ。メタデータ、ツイッターのつぶやき、SNS データなど、テキスト中心のデータがこれに当たる。非構造化データの大半（少なくとも 80%）は、組織のテキストファイルに存在しており、データ量が多いのが特徴。管理対象外のまま放置すると、ストレージコストがかさむだけでなく、万が一データを紛失した場合に、会社や事業運営に重大な問題が発生する可能性がある。

●多構造化データ：Web アプリケーション/トランザクション、ソーシャルネットワークなど、コンピュータシステムとユーザーの相互運用のため、様々なデータフォーマットやデータタイプで構成されるデータ。テキストデータや画像データ、トランザクション情報などを含む Web ログデータが、最も典型的な例としてあげられる。

このデータには、作業量やコスト面の問題から以前には発見できなかった、有用なパターンや情報が含まれている。現在のデータセンターでは、ビッグデータの 3V（Volume：量、Velocity：頻度、Variability：多様性）に対応する、費用対効果の高い手法が確立され、これらを実行できるようになってきた。以下にビッグデータの 3V について簡単に説明しておこう※6：

●Volume（容量）：データの総量

●Velocity（頻度）：サーバー・ストレージへのデータ入出力の頻度

（またはリアルタイムで処理されるデータストリーム数）

●Variety（多様性）：データタイプやソースの変化する範囲

3 ビッグデータの使用例

ビッグデータは、ソーシャルネットワーク、Web サーバーログ、トラフィックフローセンサー、衛星画像、ブロードキャストストリーム（オーディオデータと動画データの両方）、金融取引、インターネットダウンロード、文書検索、文書スキャン、エンジニアリング設計、GPS 追跡、自動車計測、市場データ、分析データなどの雑多なデータ、もしくは文字通り大きすぎたり、更新頻度が高すぎるなど、これまでの手法では対応できないデータの総称であることは、これまで説明した通り。ここでは、実世界の分野で、どのようにビッグデータが活用されているか紹介していこう ※7:



民間部門

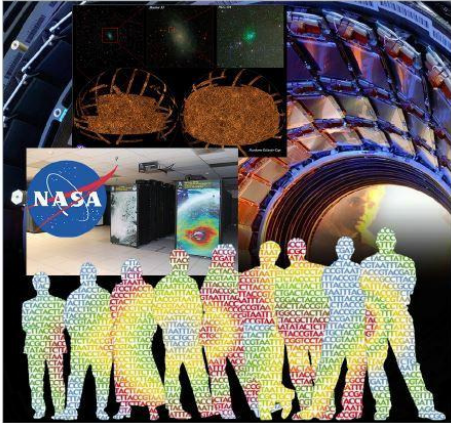
- Facebook では、1 日あたり 500 億枚のユーザー投稿画像を処理している
- ウォールマートでは、1 時間あたり 100 万以上の顧客トランザクションを処理している
- Amazon は数百万のバックエンド業務と、50 万以上のサードパーティクエリを 1 日で処理している
- eBay では、製品検索やユーザーのオススメ、マーチャンダイジングのため、7.5PB と 40PT のデータハウスを使用している



政府部門

- オバマ政権は 2012 年に、政府が直面する重要な課題に対してビッグデータを活用する「ビッグデータ研究開発イニシアティブ」を発表した
- オバマ大統領の再選運動でもビッグデータ分析は大きな役割を果たしている
- 米国連邦政府は、世界で最も強力なスーパーコンピュータ 10 台のうち 6 台を所有している
- 国家安全保障局（NSA）は、インターネット経由で収集した大量の情報を処理するため、データセンターの開発を進めている

科学研究部門



- 約 1.5 億個のセンサーを搭載する世界最強の粒子加速器（大型ハドロン衝突型加速器）では、1 秒当たり 4,000 万回のデータを素粒子物理学および高エネルギー物理学に解析するため、ビッグデータ分析を使用している
- スローン・デジタル・スカイサーベイ（SDSS）では、2000 年より1 日辺り 200GB、合計 140TB を超える天文データを収集している
- 解析に 10 年以上の時間が必要とされたヒトゲノム解析だが、現在ではビッグデータ分析を使用することで 1 週間たらずで完了することができる
- NASA 気候シミュレーションセンター（NCCS）では、32PB の気候観測データとシミュレーションデータを保存している

4 ビッグデータ分析

ビッグデータ分析の進歩により、日々の生活や将来に向けての開発における意思決定に、重要な影響をあたえるようになってきている。なお Bernard Marr 氏は、“The Awesome Ways Big Data is Used Today to Change Our World,”の記事において、以下のようなビッグデータ分析の使用例をあげている：

- 顧客決定と理解：ビッグデータ分析により、より良い顧客や彼らの行動、購買の好みを理解し、購買動向の予測モデルを作成することができる。これは、新たな高成長市場や新製品の特徴、機能、技術の開発につながる可能性があり、ビッグデータ分析の最も一般的な使用方法だ。
- ビジネスプロセスの理解と最適化：小売業者の予測モデルに基づいて、商品在庫の分析やモニタリングを行うことで、ビジネスプロセスを最適化する。また位置情報や RFID タグを通して商品や配送車を追跡し、実際のトラフィックデータと統合することで、納期を短縮できる経路を検索する、ルート最適化などもこの例に含まれる。
- 個人能力の定量化および最適化：スマートウォッチやブレスレットのような、ウェアラブルデバイスから収集したデータを使い、日々のカロリー消費量、活動レベル、睡眠パターンなど、個人の能力の定量化と最適化を行うことができる。また出会い系サイトの多くでは、加入者の最適なカップリングを見つけるためにビッグデータ分析を使用している。

ビッグデータ分析の活用例 10 項目：

- 顧客決定と理解
- ビジネスプロセスの理解と最適化
- 個人能力の定量化および最適化
- ヘルスケアおよび公衆衛生の改善
- スポーツ能力の改善
- 科学・研究分野の改善
- 機械やデバイスの最適化
- 安全保障や法的処置の改善
- 都市機能や国家機能の改善
- 金融市場の最適化

•ヘルスケアおよび公衆衛生の改善：ビッグデータ分析により、わずか数分で全DNAの文字列を解析できる他、新しい治療法の発見や、疾患のパターンを予測をすることができる。また未熟児の呼吸や心拍をモニターして、体調に大きな変化がおこる前に感染症を予測したり、インフルエンザの流行監視にも使用されている。

•スポーツ能力の改善：スポーツの分野では、傾向や一貫性を判断するために、ビデオデータのビッグデータ分析が使用されている。またボールやゴルフクラブ、テニスラケットなどにセンサーを取り付けることで、これまでとは違うアプローチによる改善も行うことができるだろう。さらに多くのスポーツチームでは、選手の栄養、睡眠、感情などを判断するため、ビッグデータ分析を活用している。

•科学・研究分野の改善：前述した大型ハドロン衝突型加速器は、科学・研究分野におけるビッグデータ分析の最たる例といえる。またスイスにある欧州原子核研究機構（CERN）の核物理研究室では、30PBのデータを分析するため、世界150のデータセンターに分散した数千台のコンピュータ（計65,000プロセッサ）を使用している。さらに石油やガスの採掘には、地震画像によるビッグデータ分析が用いられている。

•機械やデバイスの最適化：ビッグデータ分析を使用することで、機械やデバイスはよりスマートかつ自律的な動作が可能となる。例えば、Googleが開発した自動運転のトヨタ「プリウス」は、カメラ、GPS、強力なコンピュータやセンサーが装備され、人間が運転しなくても安全に走行することができる。

•安全保障や法的処置の改善：国家安全保障局（NSA）では、ビッグデータ分析により、サイバー攻撃の検出や防御、テロリストの計画を未然に防いでいる。さらにこれらのツールは、犯罪予測だけでなく、犯罪者の逮捕にもつながることがある。またクレジットカード会社では、不正行為の検出に活用している。

•都市機能や国家機能の改善：都市や自治体は、交通量の最適化のため、実際の交通状況や気象データ、ソーシャルメディアデータを含めたビッグデータ分析を行っている。交通インフラとこれらの手法を統合することで、電車の遅延に合わせてバスが

出発したり、交通量に合わせて信号を操作し、渋滞を最小限に抑えることができる。

ビッグデータアプリケーションでは、読取・書込のワークロードが混在しており、非常に低いレイテンシと高い IOPS性能が要求される。そのため HDD ではなく、エンタープライズ向けに設計された SSD が最適なソリューションになる。

●金融市場の最適化：ソーシャルメディアネットワークやニュースサイトの情報を使い、瞬時の意思決定ができるビッグデータ分析は、株の売買でもよく利用されている。ここまで、ビッグデータの一般的な活用方法について紹介してきたが、さらなるポテンシャルを秘めたビッグデータアプリケーションも多数存在しており、新たな活用方法がさらに生まれてくることだろう。

5 ビッグデータにおける SSD のメリット

そのデータの大きさや更新頻度、さらにデータベースアーキテクチャへの非適合もあり、ビッグデータから投資に見合う効果を得るには、IT部門は従来と異なるデータ分析手法を選択する必要がある。ビッグデータは、現在の企業や ITインフラに、大規模データセットを高速に処理できる環境を要求するとともに、データ駆動型分析を促進した。

ビッグデータアプリケーションでは、読取と書込が同時に処理されるため、非常に低いレイテンシと高い IOPS 性能が要求される。データアクセス時に、ヘッドを機械的に移動させる HDDでは、ディスク上の異なる場所に配置されたデータを処理する、ランダムアクセス性能は 100~300IOPS 程度しかなく、サーバー処理能力の数十万IOPS に比べると大きな隔りがある。このことから、ビッグデータアプリケーションや、増加を続けるサーバー処理に対応するのは難しく、エンタープライズ向けに設計された SSD が最適なソリューションになる。

HDD は、同一トラックに連続してデータが配置されている、シーケンシャルデータの読込・書込を前提とした設計が行われているが、最近の OSは、マルチプロセッサによる複雑なデータ処理を実行するため、ランダムアクセスが増加。HDD では性能が不足してきている。一方、SSD に採用されている NAND フラッシュは、高密度かつ、コントローラ自体にデータの位置情報が登録されているため、ランダムデータでも高速かつ低遅延にアクセスできる。さらに、ディスクや磁気ヘッドのような駆動部品が必要ないため、故障にも強いという特徴がある。

HDD に比べて高速な I/O 性能、様々なフォームファクタやインターフェイスへの対応、省電力性能、電源断時の対障害性の高さなどから、SSD はビッグデータアプリケーションを使用する上で、ポピュラーな選択といえる。

さらに詳しい情報について

OCZ Storage Solutions

6373 San Ignacio Avenue

San Jose, CA 95119 USA

TEL : 408-733-8400

E-Mail : sales@ocenterprise.com

Web サイト : <http://ocz.com/enterprise>

6 まとめ

大手市場調査会社によるストレージ業界の予測では、ビッグデータは急成長を続け、市場規模は今後 5 年間で数十億ドル規模になるとしている。また IDC の定義によれば、ビッグデータとは、高頻度に取得・発見・分析した膨大かつ多種多様なデータから、経済的に意味のある値を導き出すために開発された新技術、またはアーキテクチャとされる。

HDD に比べて高速な I/O 性能、様々なフォームファクタやインターフェイスへの対応、省電力性能、電源断時の対障害性の高さなどから、SSD はビッグデータアプリケーションを使用する上で、ポピュラーな選択といえる。またビッグデータは、巨大なデータセットを高速に処理できる環境を要求するとともに、データ駆動型分析を促進した。さらに従来の手法では処理できないほど、大容量かつ頻繁に更新される構造化・非構造化データで構成されるため、データ活動やフラッシュリソースをリモートで、管理・測定する機能には大きなメリットがある。

ホワイトペーパーの第 2 部「ビッグデータアプリケーションにおける OCZ SSD ソリューションの活用」では、OCZ が提供するビッグデータ向け SSD とストレージソリューションについて概説します。

参考データ

- 1 "What is Big Data" – a definition from Webopedia.com.
- 2 "What is Big Data," by Lisa Arthur, Forbes Magazine (2013 年 8 月 15 日)
- 3 "Big Data: The next frontier for innovation, competition, and productivity," by James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, McKinsey Global Institute (MGI), McKinsey & Company, (2011 年 5 月)
- 4 "Worldwide Big Data Technology and Services 2013-2017 Forecast," by Dan Vessel, Rob Brothers, Steve Conway, Matthew Eastwood, John Grady, Brian McDonough, Henry D. Morris, David Schubmehl, Mary Johnston Turner, Melissa Webster, Ashish Nadkarni, Christian Christiansen, Mukesh Dialani, Maureen Fleming, Tim Grieser, Rohit Mehra, Carl W. Olofson, Kuba Stolarski, Mary Wardley, Ali Zaidi, IDC, Report #244979 (2013 年 12 月)
- 5 "What is Big Data," by Lisa Arthur, Forbes Magazine (2013 年 8 月 15 日)
- 6 "What is Big Data," by Edd Dumbil, Strata-O'Reilly Media (2012 年 1 月 11 日)
- 7 "What is Big Data," – a definition from Wikipedia.com.
- 8 "The Awesome Ways Big Data Is Used Today To Change Our World," by Bernard Marr, LinkedIn.com (2013 年 11 月 13 日)

免責事項

OCZは、予告なく製品仕様や解説を変更することがあります。本文書は、情報提供のみを目的としており、技術的に不適切な記述や記載もれ、誤植を含むことがあります。本文書の性能評価結果は、OCZ 製品の性能測定システムで計測をしているため、ソフトウェアやハードウェア構成の違いにより、実際のパフォーマンスとは異なる可能性があります。OCZ では、本文書で参照している サードパーティ製ベンチマークや、Web サイトの設計・実装については一切関与しません。製品ロードマップ、コンポーネントや ハードウェアのリビジョン、新製品のリリース情報、ソフトウェア、ファームウェアなど、本文書に記載されている情報は変更される可能性があります。OCZは、本文書に記載されている情報のアップデートや訂正をする義務を負わないものとします。

OCZは、本文書に記述されている内容についての不正確な記述、誤り、欠落などに関して、一切責任を負わないものとします。

OCZは、あらゆる商品性の黙示保証もしくは特定目的との適合性を放棄し、本文書に記載されている情報に起因する損害について、直接的、間接的、付随的、結果的であることを問わず一切の責任を負わないものとします。

商標

© 2014 OCZ Storage Solutions, Inc. – A Toshiba Group Company. All rights reserved.

OCZ、OCZ ロゴ、OCZ XXXX、OCZ XXXXX、[製品名]、およびそれらの組み合わせは OCZ Storage Solutions, Inc. – A Toshiba Group Company の商標です。他のすべての製品名およびロゴは、参考情報として記載しており、それぞれの所有者の商標である場合があります。

ホワイトペーパー



ビッグデータアプリケーションにおけるOCZ SSDソリューション の活用

第2部：ビッグデータアプリケーションに要求される性能と管理方法を提供する


Scott Harlin

OCZ Storage Solutions, Inc. - A Toshiba Group Company

目次

- 1 はじめに
 - 2 Intrepid 3000 SATA SSDシリーズ
 - 3 Z-Drive 4500 PCIe SSDシリーズ
 - 4 OCZ WXLソフトウェアによるビッグデータの高速度
 - 5 OCZ VXLソフトウェアによるビッグデータの仮想化
 - 6 ZD-XL SQL Accelerator
 - 7 OCZ StoragePeak 1000による集中管理
 - 8 まとめ
-

1 はじめに



ホワイトペーパー第1部「ビッグデータストレージについて理解を深める」では、エンタープライズ向けストレージの導入についてより深く理解するため、ビッグデータアプリケーションに対する概念と特徴を紹介するとともに、SSDがビッグデータモデルにどのように適合するかを解説する。

HDDに比べて高速なI/O性能、様々なフォームファクタやインターフェイスへの対応、省電力性能、電源断時の対障害性の高さなどから、SSDはビッグデータアプリケーションを使用する上で、ポピュラーな選択といえる。そのデータの大きさや更新頻度、さらにデータベースアーキテクチャへの非適合もあり、ビッグデータから投資に見合う効果を得るには、IT部門は従来と異なるデータ分析手法を選択する必要がある。

ビッグデータアプリケーションでは、読取と書込が同時に処理されるため、非常に低いレイテンシと高いIOPS性能が要求される。そのためHDDではなく、エンタープライズ向けに設計されたSSDが最適なソリューションとなる。ホワイトペーパー第2部では、ビッグデータアプリケーションによるデータ駆動型分析に必要な、大規模データセットの超高速処理を実現する、OCZのエンタープライズSSDおよび、ソフトウェアソリューションについて解説する。なお第2部で扱うOCZ Storage Solutionsの製品は以下の通り：

- Intrepid 3000 SATA SSDシリーズ
- Z-Drive 4500 PCIe SSDシリーズ
- Windows Acceleration (WXL) ソフトウェア
- VXL仮想化ソフトウェア
- ZD-XL SQL Accelerator
- StoragePeak 1000 Central Management

それでは、早速各ソリューションがビッグデータアプリケーションに、どのように対応するのか解説していくことにしよう：

2 Intrepid 3000 SATA SSDシリーズ

標準的な2.5インチフォームファクタに対応する、Intrepid 3600 (cMLC) とIntrepid 3800 (eMLC) は、19nm MLC NANDフラッシュを採用。容量は100G、200GB、400GB、800GBがラインナップされる



OCZのエンタープライズ向けSATA SSDでは、最高の性能と容量を実現した「Intrepid 3000 SSD」シリーズは、HDDの代替として、ビッグデータアプリケーションには理想的な製品だ。NANDフラッシュには19nmのMLCタイプが採用され、最大容量は800GB。また高度なフラッシュマネジメント技術を実装する、「Everest 2」プラットフォームを採用することで、NANDフラッシュの寿命を高め、高い耐久性を実現している。

ビッグデータアプリケーションでは、大規模な読取・書込処理が混在するため、非常に低いレイテンシと高いIOPS性能が要求される。これに対応するため、「Intrepid 3000 SSD」シリーズでは、読込処理が中心のコストパフォーマンスモデル「Intrepid 3600」と、書込や混成処理が中心の「Intrepid 3800」の2種類のラインナップを用意している。

•Intrepid 3600 : オンラインアーカイブ、メディアストリーミング、ウェブブラウジングのような読込処理が中心のアプリケーションに最適な、コストパフォーマンスに優れたMLC NANDフラッシュを採用

•Intrepid 3800 : ビッグデータ、クラウドコンピューティング、オンライントランザクション処理 (OLTP)、仮想デスクトップインフラストラクチャ (VDI)、電子メールサーバー、解析処理のような、書込や混合処理が中心のアプリケーションに最適な高耐久eMLC NANDフラッシュを採用

「Intrepid 3600/3800」シリーズは、2.5インチフォームファクタに対応し、容量は100GB、200GB、400GB、800GBがラインナップされる。また全領域で書込・消去・再書込を行うことで、常に最適な状態を保ち、シーケンシャルアクセス、ランダムアクセスともクラス最高レベルの性能を実現している。なお各転送速度は以下の通り：

シーケンシャル読込 (128kブロック) 520MB/sec

シーケンシャル書込 (128kブロック) 470MB/sec

ランダム読込 (4Kブロック) 89,000 IOPS

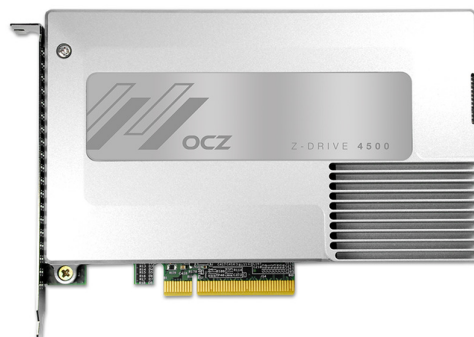
ランダム書込 (4Kブロック) 40,000 IOPS

「Intrepid 3000」シリーズは、データの圧縮率に関係なく一定のパフォーマンスを発揮できるように設計され、従来のOCZ製エンタープライズSATA SSDに比べて、持続4K書込は5倍、持続4K読込は2倍に向上している。また安定かつ低いレイテンシにより、効率的なI/O性能を実現。システムやストレージのボトルネックを軽減し、エンドユーザーの生産性や全体的なコンピューティングエクスペリエンスを改善する。

ちなみに4Kブロックの書込テスト結果では、従来のエンタープライズSATA SSDに比べて、I/Oレスポンスは約12倍に高速化。安定した低レイテンシを提供できることから、ビッグデータアプリケーションには最適な製品となるだろう。

3 Z-Drive 4500 PCIe SSDシリーズ

フルハイットまたはハーフハイットPCI-Expressに 対応する、「Z-Drive 4500 PCIe SSD」シリーズは、19nm MLC NANDフラッシュを採用。容量は800G、1.6TB、3.2TBがラインナップされる



サーバー側の構成を考えた場合、単一ドライブとしてPCI-Expressスロットに直接接続できる、OCZ「Z-Drive 4500 PCIe SSD」シリーズは、ケーブルが必要なSAS/SATAに代わる、有効な代替手段だ。「Z-Drive 4500 PCIe SSD」シリーズを使い、フラッシュストレージをホストに統合することで、サーバー処理能力に匹敵するIOPS性能を備えたローカルリソースを構築可能。これにより、データをホスト側に移動して、ホストリソースの処理性能や効率を改善できる他、SANストレージによって専有されているラックスペースや、消費電力を減らすことができる。

そのデータの大きさや更新頻度、さらにデータベースアーキテクチャへの非適合もあり、ビッグデータから投資に見合う効果を得るには、IT部門は従来と異なるデータ分析手法を選択する必要がある。

「Z-Drive 4500 PCIe SSD」シリーズは、最新ビッグデータアプリケーションのパフォーマンス要件を満たすよう設計され、従来のエンタープライズ向けPCI-Express SSDと比較しても、優れたパフォーマンスを実現している。NANDフラッシュは19nmのMLCタイプが採用され、容量は800GB、1.6TB、3.2TBが用意される。

またOCZ独自の「仮想コントローラ技術（VCA : Virtualized Controller Architecture）」を搭載。ホスト上からは1ドライブとして認識されるが、実際には内蔵された8つのコントローラを使い、動的にコマンドやプロセスを並び替えて並列処理することで、RAIDドライブのような効率的かつ高速な転送速度と、シングルドライブのような簡単な管

理方法を提供している。

この「仮想コントローラ技術 (VCA)」により、「Z-Drive 4500 PCIe SSD」シリーズは、MLCベースのエンタープライズ向けPCI-Express S SSDのなかでもトップクラスの持続性能を発揮する。なお業界標準のベンチマークによる転送速度結果は以下の通り：

シーケンシャル読込 (4kブロック) 2,900 MB/sec

シーケンシャル書込 (4kブロック) 2,200 MB/sec

ランダム読込 (4Kブロック) 252,000 IOPS

ランダム書込 (4Kブロック) 76,000 IOPS

高記憶容量、NANDフラッシュ採用による省電力性、高いIOPS性能、広い帯域幅などの特徴から、「Z-Drive 4500 PCIe SSD」は、読込・書込のI/Oが集中するビッグデータアプリケーションにとって理想的なソリューションとなるだろう。

4 OCZ WXLソフトウェアによるビッグデータの高速度化

「Intrepid 3000」および、「Z-Drive 4500 PCIe SSD」に対応する、フラッシュ管理・キャッシュソリューション「OCZ WXL (Windows Accelerator) ソフトウェア」では、Microsoft Windows Serverに低レイテンシのフラッシュボリュームや、HDDキャッシュを作成し、ビッグデータアプリケーションの処理速度を高速化する。

「Intrepid 3000」や「Z-Drive 4500 PCIe SSD」にホットデータをキャッシュすることで、アクセス時間やビッグデータの待ち時間、SANリソースへのアクセスを削減し、ストレージコストやレイテンシによるボトルネックを解消、ビッグデータアプリケーションの性能を向上させることができる。

ファイルサイズの小さいアプリケーションの場合、データをすべてフラッシュボリュームに格納することで処理性能を向上させる。またフラッシュボリュームに保存できない、サイズの大きなデータファイルは、OCZ独自キャッシュ技術により、インテリジェントに必要なデー

タのみを選択してキャッシュへと保存する。「OCZ WXLソフトウェア」は、最もアクセス頻度の高いデータをキャッシュして、SAN/DASシステムの性能とレイテンシを改善するように設計され、「Intrepid 3000」や「Z-Drive 4500 PCIe SSD」にホットデータをキャッシュすることで、アクセス時間やビッグデータの待ち時間、SANリソースへのアクセスを削減。ストレージコストやレイテンシによるボトルネックを解消して、ビッグデータアプリケーションの性能を向上させることができる。

キャッシュとして運用する場合、SANやHDDへのアクセスを監視し、独自のキャッシュポリシーに基づき、重要なデータのみを「Intrepid 3000」や「Z-Drive 4500 PCIe SSD」に保存することで、最大90%の外部トラフィックを削減する。キャッシュポリシーは、SSDの耐久性を損なうことなく、キャッシング効率を最大限に発揮するため、最も頻繁にアクセスされるデータのみを検出、それ以外のデータを除外する高度なキャッシュ・アルゴリズムを採用している。さらに「OCZ WXLソフトウェア」では、他のアプリケーションとキャッシュを共有するため、ホスト上からキャッシュボリュームを動的に配分することができる。

ビッグデータ解析作業に必要なデータを「Intrepid 3000」や「Z-Drive 4500 PCIe SSD」のキャッシュドライブに確実に保存するためには、キャッシュのウォームアップと分析メカニズムがとても重要な要件になる。

5 OCZ VXLソフトウェアによるビッグデータの仮想化

「Z-Drive 4500 PCIe SSD」に実装された仮想アクセス技術「OCZ VCL」では、内蔵コントローラがすべてのNANDフラッシュにランダム処理を分散することで、アプリケーションの性能を向上させている。さらに「Z-Drive 4500 PCIe SSD」と「OCZ VXLソフトウェア」を組み合わせることで、SSDリソースを複数の仮想マシンに分散して、ビッグデータのような主要アプリケーションの性能を最大限に発揮する、完全な仮想システムを構築することができる。

「VXLソフトウェア」では、エンド・ツー・エンド、高可用性 (HA)、

フォールトトレランス (FT)、動的仮想マシン移行などの種別に関係なく、すべてのハイパーバイザー型仮想クラスターの仮想マシンからアクセスできる、可用性の高いネットワークリソースとして「Z-Drive 4500 PCIe SSD」の仮想化ができる。そして仮想マシンが必要とするときにSSDソースを配布。不要になった段階ですぐに回収することで効率の良い処理を実現している。

複数の仮想マシンが、同時にピーク処理に達することは少ないことから、SSDを共有するこの仮想化手法は、仮想環境における最も費用対効果に優れる方法といえる。「Z-Drive 4500 PCIe SSD」によるSSD キャッシュは、同時実行されている仮想マシン数に関係なく、常に最適な効率で重要なデータを仮想マシンに提供。SANからのデータトラフィックを削減し、ビッグデータアプリケーションの仮想化を可能にする。

6 ZD-XL SQL Accelerator

フルハイトまたはハーフハイトPCI-Expressに対応する、「ZD-XL SQL Accelerator」は、19nm MLC NANDフラッシュを採用。容量は800G、1.6TB、3.2TBがラインナップされる



「ZD-XL SQL Accelerator」は、Microsoft SQL Serverで構築された大規模データベースの処理性能を改善するために、高速なPCI-Express SSDと、フラッシュボリュームやHDDキャッシュを作成できるチューニングアプリケーションで構成された統合ソリューションだ。高速なSSDに組み合わされるアプリケーションには、統計学的に最適化された独自キャッシュ機構や、ジョブや要求に合わせて必要なデータをキャッシュする動的ウォームアップスケジューラ、ウィザードベースのGUIを備えたデータベースキャッシュポリシーセットアップ等を搭載し、SQL Serverの処理を最適化できる。

「ZD-XL SQL Accelerator」は、SSDとキャッシュソフトウェアを密接に統合し、SQL Serverのストレージ性能や可用性、メンテナンス性を改善。SQL Serverバージョンは、2008 R2/2012に加え、今年4月1日にリリースされた最新バージョン2014に対応する。またフラッシュドライブからデータベースページを直接ロードする「Buffer Pool Extension」(BPE)もサポートされ、SQL Server 2014の性能を最大限に発揮。これにより、最新ビッグデータアプリケーションでもパフォーマンスを最適化することができる。

ビッグデータのような大規模なデータセットを分析し、企業にとって有益かつ重要な情報を引き出すことは、現在の企業ではとても重要な課題となっている。Panorama Softwareの「Necto Business Intelligence 3.0」のような高度なディスカバリツールと、強力なインメモリ技術を組み合わせることで、データ探索や高度な解析をわずか数分で完了することができる

「OCZ StoragePeak 1000」を使うことで、エンタープライズフラッシュリソースのクロスプラットフォーム監視、ネットワークサーバー接続、ストレージレイヤアプリケーションの集中管理、管理、保守、およびレポート作成を行うことができる。

OCZとPanorama Softwareが共同で開発しているソリューションを使うことで、ビジネス情報やデータベースに含まれるデータセットを、ビジネス駆動型探索やビジネス・インテリジェンスへと変換することができる。Microsoft SQL Server 2014に格納されたデータを、リアルタイムにビッグデータ解析することは、ビッグデータアプリケーションおよび、SSDの進化における重要なステップとなるだろう。現在開発しているソリューションは、最小限の労力で情報に基づいた意思決定ができるよう、インメモリエンジン、インタラクティブな自己分析、ユーザーが簡単にアクセスできる視覚情報やダイナミックダッシュボード、分析、視覚化、パフォーマンス追跡、協調動作、素早いデータ共有などの機能が実装されている。

「ZD-XL SQL Accelerator」は「Z-Drive 4500 PCIe SSD」と同等の性能を備えており、高記憶容量、NANDフラッシュ採用による省電

力性、高いIOPS性能、広い帯域幅などの特徴から、I/O集約型SQL Serverアプリケーションにとって理想的なソリューションとなるだろう。

7 OCZ StoragePeak 1000による集中管理

OCZが提供するビッグデータ向けソリューションのラストピースは、ネットワークアクセス管理システムとして開発された「OCZ Storage Peak 1000」だ。エンタープライズフラッシュリソースのクロスプラットフォーム監視、ネットワークサーバー接続、ストレージレイヤアプライアンスの集中管理、保守、およびレポート作成などの機能を搭載。IT管理者は、システムやストレージリソースの管理に必要な、システムやSSDの健全性などの情報や、ミッションクリティカルな処理を集中管理でき、データセンターの費用対効果を最大限に高めることができる。

「OCZ StoragePeak 1000」はネットワークを経由して、複数のホストシステムに安全に接続。IT管理者はWebベースの管理インターフェイスを使い、エンタープライズフラッシュリソースを管理できる。対応するホストOSはLinuxまたはWindowsで、扱いやすいWebベースの統合GUIにて、ドライブの詳細性能や信頼性、動作状況を確認できる。他、ユーザ設定可能な警報システムにより、システムやストレージの潜在的な障害を認識、早期に是正措置を取ることができる。

ちなみに「OCZ StoragePeak 1000」に用意されている機能は以下の通り：



- データセンタ全体のホストとSSDの状況を確認できる構造化グループビュー
- ホストやSSDからの警告やアラート表示
- シンプルかつ簡単なSSDのインストール、保守、管理
- 迅速かつ簡単なSSDの定期メンテナンス、ホストシステムのチェック、

ファームウェアアップデートや詳細レポート印刷などの管理タスクビッグデータは、従来のデータベースやソフトウェア技術では処理できない、大容量かつ、頻繁に更新される構造化データおよび、非構造化データで構成されているため、そのデータの活動やフラッシュリソースをリモートで管理・監視する機能は、ビッグデータアプリケーションではとても有効だ。

8 まとめ

OCZでは、ビッグデータアプリケーションを対象にしたSSDおよびストレージソリューションの完全なポートフォリオを提供している：

- 書込や混合処理が中心のビッグデータアプリケーションに対応する、高速かつ大容量なエンタープライズ向けSATA/PCI-Express対応SSD (「Intrepid 3800」 / 「Z-Drive 4500 PCIe SSD」)
- 書込や混合処理が中心のSQL Serverデータベースに対応する、高速かつ大容量なエンタープライズ向けPCI-Express対応ソリューション (「ZD-XL SQL Accelerator」)
- 読込処理が中心のシステムに最適な、大容量かつコストパフォーマンスに優れるSATA対応SSD (「Intrepid 3600」)
- Windowsアプリケーションのパフォーマンス改善 (「OCZ WXLソフトウェア」と「Intrepid 3000 SSD」または「Z-Drive 4500 PCIe SSD」)
- ハイパーバイザー型仮想マシンのパフォーマンス改善 (「OCZ VXLソフトウェア」と「Z-Drive 4500 PCIe SSD」)
- SSDの集中管理システム (「OCZ StoragePeak 1000」と「Intrepid 3000 SSD」または「Z-Drive 4500 PCIe SSD」)

データは前年比40%ずつ増え続け、この2年間で作成されたデータは世界中の90%にのぼることからも分かる通り、すべての企業はビッグデータについて完全に理解し、すぐにでも性能・分析・管理に対応するストレージ戦略を実施する必要がある。OCZ Storage Solutions-Toshiba Group Companyでは、このビッグデータストレージの厳しい要求に対応する、最適なソリューションを提供しています。

さらに詳しい情報について

OCZ Storage Solutions

6373 San Ignacio Avenue,

San Jose, CA 95119 USA

TEL : 408-733-8400

E-Mail : sales@oczenterprise.com

Webサイト : <http://ocz.com/enterprise>

免責事項

OCZは、予告なく製品仕様や解説を変更することがあります。本文書は、情報提供のみを目的としており、技術的に不適切な記述や記載もれ、誤植を含むことがあります。本文書の性能評価結果は、OCZ製品の性能測定システムで計測をしているため、ソフトウェアやハードウェア構成の違いにより、実際のパフォーマンスとは異なる可能性があります。OCZでは、本文書で参照しているサードパーティ製ベンチマークや、Webサイトの設計・実装については一切関与しません。製品ロードマップ、コンポーネントやハードウェアのリビジョン、新製品のリリース情報、ソフトウェア、ファームウェアなど、本文書に記載されている情報は変更される可能性があります。OCZは、本文書に記載されている情報のアップデートや訂正をする義務を負わないものとします。

OCZは、本文書に記述されている内容についての不正確な記述、誤り、欠落などに関して、一切責任を負わないものとします。

OCZは、あらゆる商品性の黙示保証もしくは特定目的との適合性を放棄し、本文書に記載されている情報に起因する損害について、直接的、間接的、付随的、結果的であることを問わず一切の責任を負わないものとします。

商標

© 2014 OCZ Storage Solutions, Inc. – A Toshiba Group Company. All rights reserved.

OCZ、OCZロゴ、OCZ XXXX、OCZ XXXXX、[製品名]、およびそれらの組み合わせはOCZ Storage Solutions, Inc. – A Toshiba Group Companyの商標です。他のすべての製品名およびロゴは、参考情報として記載しており、それぞれの所有者の商標である場合があります。