インテル® Optane™ DC SSDを活用した高信頼構成を日立製作所が徹底検証！

オンメモリーと比較しても遜色のない高速性かつ高信頼なHiRDB環境を実現

膨大なデータを常に活用し続けるには、データベースの高速化、高信頼化が欠かせない。前回の検証において日立製作所は、ノンストップデータベース「HiRDB」を用いてインテル® Optane™ DC SSDを検証し、その性能の高さを実証した。今回は、HiRDBのHAクラスタ構成にインテル® Optane™ DC SSDを活用し、高信頼システムに耐えうるかどうかの検証を行った。

少ないメモリーでバッファヒット率が低くても高スループットを実現

近年、企業が保有するデータは爆発的に増加し続けており、ITシステムにはレスポンスの高速化が求められている。とくに電子マネーの小口取引処理など、大量の小さなトランザクションを高スループットで処理しなければならない場面では、データベースのパフォーマンスがITシステムのレスポンスに直結する。

日立製作所ではこうしたニーズに応え、純国産のRDBMS「HiRDB」を提供している。株式会社日立製作所の熊谷 昌大氏は、「とくに金融や公共系のミッションクリティカルな基幹系システムでは、お客様から高性能・高信頼に対するご要望を多くいただきます。これにお応えするため自社開発のノンストップデータベース“HiRDB”をご提供しています」と話す。

株式会社日立製作所
サービスプラットフォーム事業本部 IoT・クラウドサービス事業部
データマネジメント本部 DB部
部長
熊谷 昌大 氏

前回の記事において日立製作所では、インテル® Optane™ DC SSDを搭載した「日立アドバンストサーバ HA8000V」にてHiRDBを動作させ、性能検証を実施した。インテル® Optane™ DC SSDは、3D XPoint™ メモリーメディアと、システム・メモリー・コントローラー、インターフェイス・ハードウエア、ソフトウエアIPを独自に組み合わせた「インテル® Optane™ テクノロジー」を搭載するSSDだ。　検証では、従来型※NVMe SSD比で2～3倍、オンメモリー時の7割以上のスループットを達成。インテル® Optane™ DC SSDをHiRDBに適用することで、少ないメモリーでバッファヒット率が低くても高スループットが実現できることを実証し、コストを抑えながらもデータベース高速化が可能であることが分かった。

※ NAND型S　SD（1TB RI SC2 2.5型 NVMe DS ドライブ）

「インテル® Optane™ DC SSDは、低レイテンシー（低遅延）かつ安定したI/O性能、高速な書き換え処理、高耐久性、最適なコストが特徴です。前回の検証では実際のアプリケーションにおいてもそれらの効果が大きいことが分かりました」と語るのは、インテル株式会社の樋口 裕磨氏だ。

インテル株式会社
技術本部
フィールド・アプリケーション・エンジニア
樋口 裕磨 氏

スタンバイ側へ書き込む際の遅延がDB高信頼化の課題

ITシステムの実運用においてデータベースに求められるのは高速性だけではない。高信頼であることも求められる。従来の高信頼システムはサーバー障害に備えてHAクラスタ構成を組むことで信頼性を担保してきた。しかし、このHAクラスタ構成にインテル® Optane™ DC SSDを適用することでまったく新しい高信頼システムを構築できると熊谷氏は解説する。

「従来のHAクラスタ構成では共用ディスクを使用してデータベースファイルを引き継いでいました。インテル® Optane™ DC SSDは内蔵型のSSDです。サーバー2台にインテル® Optane™ DC SSDを内蔵し、ローカルのデータファイルに書き込む際に、リモート側のデータファイルも同時に書き込み、データファイルを複製することで、高信頼で高速なHAクラスタ構成を実現できると考えました」（熊谷氏）。

今回のHAクラスタ構成では、リモート（スタンバイ）側に書き込む際の遅延が課題となる。そこで日立製作所では、インテル® Optane™ DC SSDを活用したHiRDBのHAクラスタ構成の性能評価を実施した。

「データファイルの二重書きによるHAクラスタ構成では、二重書きにかかる処理時間をいかに抑えるかが重要です。遅くては実用に耐えません。ここでポイントとなるのが“NVMe Over Fabrics”とそれを支える“100GbEネットワーク”です。この構成をHA8000V上で実現しました」と話すのは、株式会社日立製作所の金子 勇氏である。

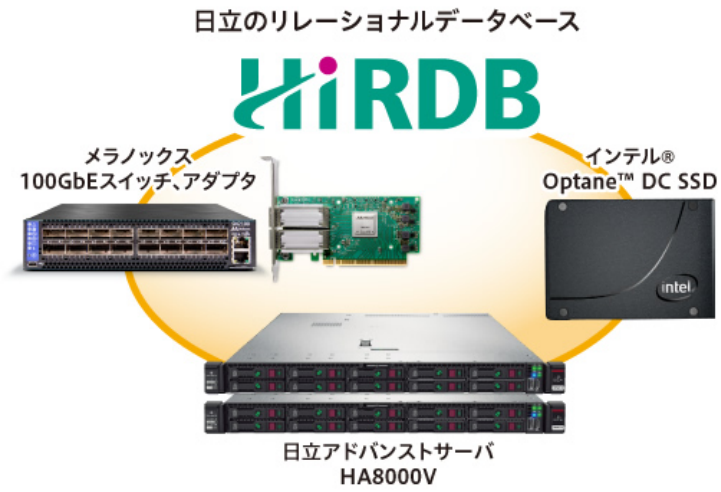
株式会社日立製作所
ITプロダクツ統括本部 ハードウェア開発本部
プロダクツ第2設計部
主任技師
金子 勇 氏

サーバーにインテル® Optane™ DC SSDを内蔵して単につなげるだけでは、HiRDBのHAクラスタ構成においてその性能を最大限には発揮できない。ネットワークがボトルネックとなるからだ。そこで、2台（アクティブとスタンバイ）のHA8000Vを結ぶネットワークにはNVMe over Fabricsおよび100GbEネットワークを利用する。NVMe over Fabricsは、高速なリモートアクセスを実現するSSDストレージの接続プロトコルだ。

サーバー、ストレージ、ネットワークの最新技術を結集

今回、インテル® Optane™ DC SSDを活用したHAクラスタ構成での性能評価を行うにあたり、使用したソリューションは図1の通りである。

●図1 性能評価でのソリューション構成



HiRDBを動作させるサーバーは、「日立アドバンストサーバHA8000V/DL360 Gen10」。優れた処理性能と高可用性を備え、システムの高集積化を実現する1Uサイズのハイパフォーマンスサーバーだ。最新世代のCPUをサポート、選択できるSSDの種類も大幅に拡充し、高速I/OやGPUなど最新技術も積極的に採用している。

なお今回の性能評価では、2018年10月リリースのHiRDB Version 10に実装される「複製ディスク機能」を利用している。複製ディスク機能は、HiRDBのデータファイルを二重書きするための機能だ。

ネットワークアダプタはメラノックス テクノロジーズ社（以下、メラノックス）製の「Mellanox ConnectX-5」をHA8000Vに装着。サーバー間を結ぶスイッチは同じくメラノックス製の「Mellanox SN2100 100GbE Switch」を利用する。

メラノックス テクノロジーズ ジャパン株式会社の小宮 崇博氏は、「メラノックスはもともとハイパフォーマンスコンピューティング領域におけるネットワークソリューションをビジネスの中心としていましたが、近年は100GbEネットワーク領域にもビジネスを拡大し、現在非常に伸びています。今回の性能評価では低遅延であることがとても重要になります。メラノックスのネットワーク製品はこうした低遅延が求められる分野で多く使われています」とメラノックス製品の特徴を語る。

メラノックス テクノロジーズ ジャパン株式会社
パートナー営業部 シニア ソリューション アーキテクト
小宮 崇博 氏

シングル構成比で約9割、オンメモリー時の7割以上のスループットを達成

インテル® Optane™ DC SSDを活用したHAクラスタ構成での性能評価を行う目的について熊谷氏は、「HiRDBが採用される案件では、HAクラスタ構成を適用するケースが多いので、インテル® Optane™ DC SSDを適用したHAクラスタ構成が実用に耐えうる性能なのかを検証しました」と語る。

今回の検証では比較のためシングル構成とHAクラスタ構成の2構成を準備している。シングル構成ではインテル® Optane™ DC SSDを搭載したHA8000V/DL360 Gen10でHiRDBを動作させる。

HAクラスタ構成ではインテル® Optane™ DC SSDを搭載した2台のHA8000V/DL360 Gen10をそれぞれアクティブとスタンバイとし、HiRDBを動作させている。なお比較のため、2台のHA8000V/DL360 Gen10に従来型のNVMe SSDを搭載してHAクラスタ構成を構築した環境も用意している。HAクラスタ構成のネットワークは、2台のサーバーの間にMellanox SN2100 100GbE Switchを設置し、ネットワークアダプタはMellanox ConnectX-5を装着している。（図2 評価環境参照）

●図2 評価環境

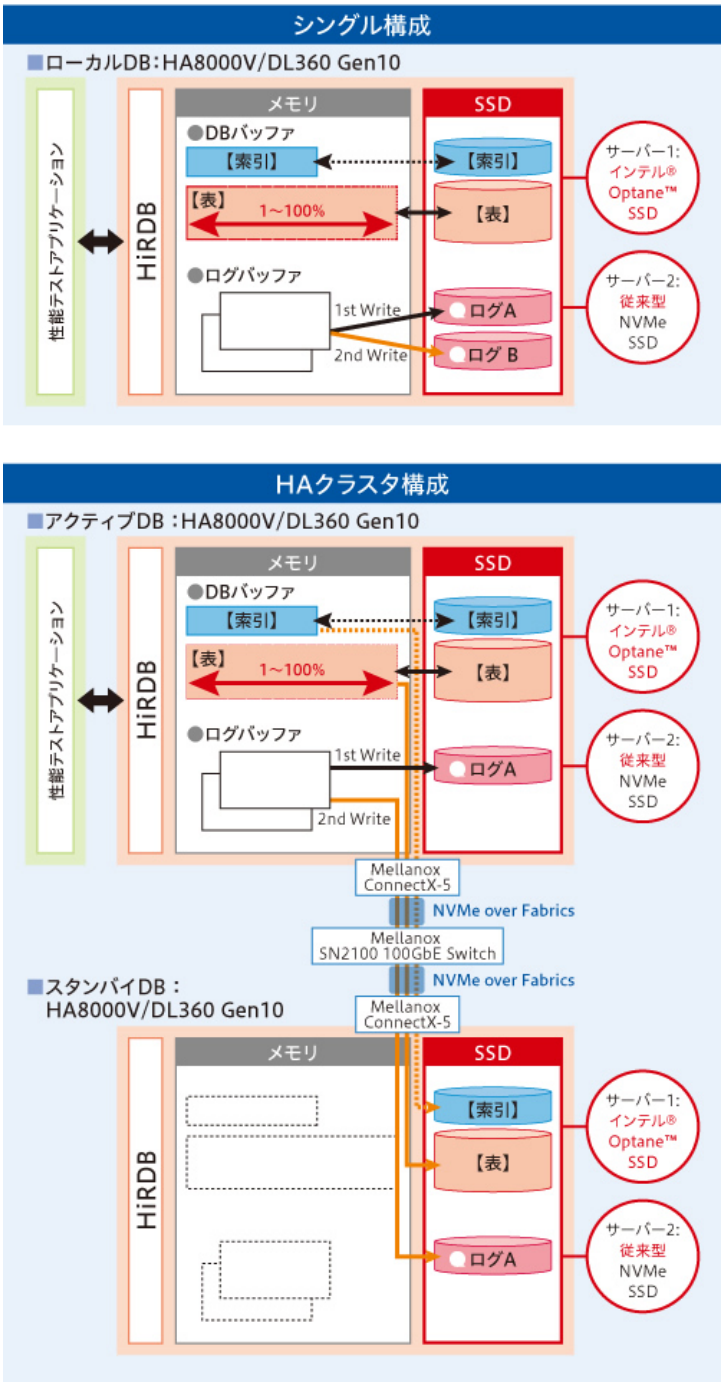
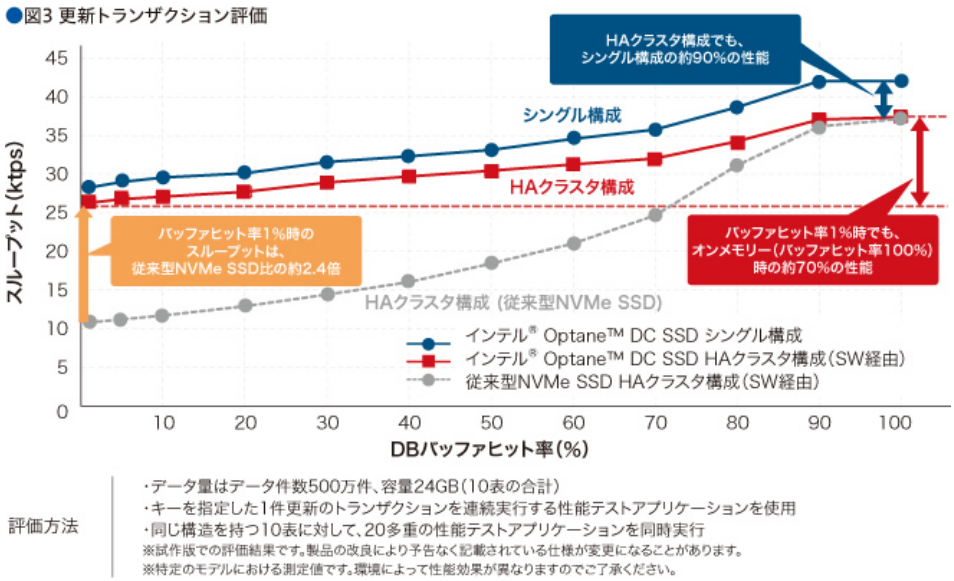


図3の更新トランザクション評価グラフは、同じ構造を持つ10表に対して、20多重の性能テストアプリケーションで更新トランザクションを連続実行した評価結果だ。



「HAクラスタ構成でもシングル構成の約90%のスループットという結果でした。これは100GbEネットワークとNVMe over Fabricsの効果により遅延が最低限に抑えられている効果が大きいと考えられます。従来型NVMe SSDでHAクラスタ構成を組んだ場合と比較して、インテル® Optane™ DC SSDでのHAクラスタ構成では、約2.4倍のスループットを達成。これは前回の検証での結果と同様で、インテル® Optane™ DC SSDの効果です。すべての表データがDBバッファに常駐しているオンメモリー時と比較して約70%という結果ですが、これは十分に実用に耐えうる性能です」と熊谷氏は評価を語る。

熊谷氏は検証からさらに一つの考察を見いだした。バッファヒット率1%時でもスループットが高いということは、系切り替え直後の立ち上げ性能が速いことになる。

「DBバッファに頼ったシステムですと、万一の際に系が切り替わった直後は、DBバッファにデータが常駐していないため、データを読み込むまでの間、サービスはトップスピードで動作しません。それに対してDBバッファに頼らないインテル® Optane™ DC SSDのシステムでは、切り替わった途端にトップスピードでサービスが開始できます。これは大変有効なことだと思います」と熊谷氏。

HiRDBをさらに進化させ、より信頼性の高い社会インフラ基盤を提供

今回の性能評価を受け、インテルの樋口氏は次のように語る。「ソリューション全体の中でどこかにボトルネックがあると、全体としての性能は頭打ちになってしまいます。今回は日立様に、当社やメラノックス様の製品特性を十分に理解した上で、性能を十分に発揮できる構成を作っていただきました。その結果も高信頼性と高性能を両立する優れたものでした。日立様の製品性能向上に当社の製品が寄与できたことは非常にうれしく思います」（樋口氏）。

ネットワークを担ったメラノックスの小宮氏は、「NVMe over Fabricsという技術はアプリケーションを、ひいてはビジネスを変えると以前から言われていましたが、今回の検証でその有用性を見ることができました。しかもインテル® Optane™ DC SSDにより、実用的以上と言える性能でHAクラスタ構成が実現できたのは良かったです。今回のように新しい技術を組み合わせて、新たなソリューションを生み出すことは素晴らしいことだと思います」と語った。

日立製作所ではこの秋、今回の性能評価でも利用した複製ディスク機能など、さまざまな機能追加が実施された新バージョン「HiRDB Version 10」をリリースする。さらに今回の性能評価をもとに、インテル® Optane™ DC SSDおよびメラノックスのネットワークを組み合わせたHiRDBのソリューションの提供を予定している。

「これまでの性能検証の過程で、HiRDB、HA8000Vとインテル® Optane™ DC SSD、100GbE、NVMe over Fabricsを組み合わせた環境で、高い処理性能を引き出すノウハウを蓄積することができました。今後は、このノウハウをソリューションとしてお客様にご提供していきます」 と熊谷氏は展望を語る。

HiRDBは、社会インフラを支えるデータベースとして、今後も進化を続けていく。

■関連URL
日立アドバンストサーバHA8000Vシリーズ ▶ http://www.hitachi.co.jp/ha8000v/
インテル® Optane™ メモリー ▶ https://www.intel.co.jp/content/www/jp/ja/products/memory-storage/solid-state-drives/data-center-ssds/optane-dc-p4800x-series.html
メラノックステクノロジーズ社 製品 ▶ http://jp.mellanox.com/page/products_overview

株式会社日立製作所
http://www.hitachi.co.jp/

ノンストップデータベース HiRDB
http://www.hitachi.co.jp/hirdb/
