Hewlett Packard Enterprise

Case Study

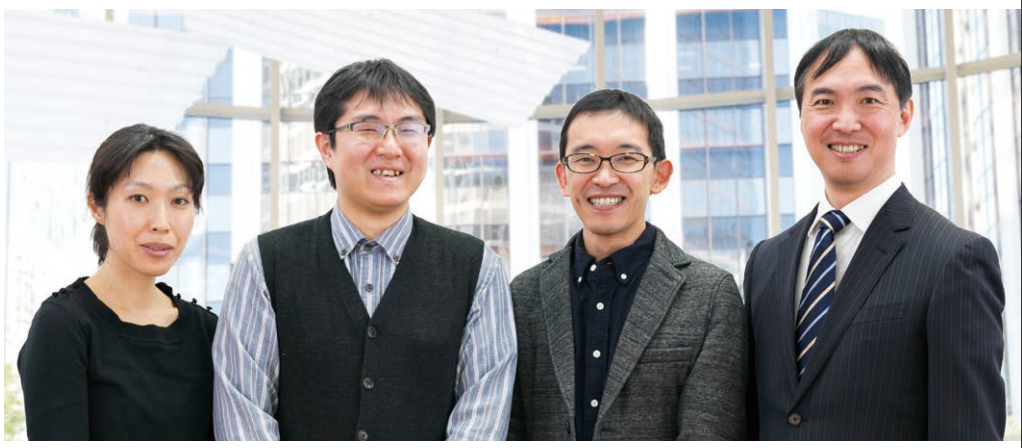
NTT ソフトウェアイノベーションセンタが、不揮発性メモリNVDIMMを活用したデータベース高速化を検証

NVDIMM搭載HPE ProLiant DL360サーバーを使用し PostgreSQL更新処理においてSSDの最大40倍の高速化を実証

“HPE Scalable Persistent Memoryのようなエンタープライズクラスの製品を、業界に先駆けて提供し続けるHPEには、サーバーベンダーとしての底力を感じます”

—日本電信電話株式会社
NTT ソフトウェアイノベーションセンタ
分散処理基盤技術プロジェクト
グループリーダ
主幹研究員　日高 東潮 氏

NTT



目的

次世代ストレージデバイスと目される不揮発性メモリ（NVDIMM）と、最新のライブラリPersistent Memory Development Kit（PMDK）の適用によるPostgreSQLのパフォーマンス向上。

アプローチ

PostgreSQLのストレージI/O処理の改善にNVDIMMとPMDKを適用し、PostgreSQL更新処理およびチェックポイント処理における高速化の効果を計測する。

ITの効果

- いち早くNVDIMMに対応したHPE ProLiant DL360サーバーを検証プラットフォームに採用することで、BIOSなどを含め安定した不揮発性メモリ利用環境における技術検討を迅速に推進
- PostgreSQLに対して計32GBのNVDIMMとPMDKの適用の組み合わせを適用し、従来のストレージデバイス適用時と比較してHDDの約170倍、SSDの約40倍のトランザクション処理性能向上を確認
- 性能改善のためWALの更新処理へPMDKを適用することにより、NVDIMMの単体利用時と比較して最大80%のトランザクション処理性能向上を確認

ビジネスの効果

- 不揮発性メモリとPMDKおよび、PostgreSQLの組み合わせにより高速トランザクション処理を可能とするOSSソリューションが実現可能であることを実証
- NVDIMMとPMDKの組み合わせにより、I/O負荷の高いアプリケーションに対し、少ない改造で高速化することを実証
- 低速なストレージI/Oとキャッシュを前提にしたプログラム構造をシンプル化できる根拠を提示

日本電信電話株式会社
NTT ソフトウェアイノベーションセンタ
分散処理基盤技術プロジェクト
グループリーダ
主幹研究員
日高 東潮 氏

日本電信電話株式会社
NTT ソフトウェアイノベーションセンタ
分散処理基盤技術プロジェクト
主任研究員
石崎 晃朗 氏

日本電信電話株式会社
NTT ソフトウェアイノベーションセンタ
分散処理基盤技術プロジェクト
研究主任
一柳 淑美 氏

日本電信電話株式会社
NTT ソフトウェアイノベーションセンタ
分散処理基盤技術プロジェクト
研究員
毛受 崇 氏

NTT ソフトウェアイノベーションセンタが画期的な検証を実施した。テーマは、次世代のストレージデバイスと目される「不揮発性メモリ（NVDIMM）」と、その能力を引き出すライブラリ「Persistent Memory Development Kit（PMDK）」を組み合わせたPostgreSQLの高速化である。結果は驚くべきもので、最大でHDDの約170倍、SSDの約40倍を実測した。検証環境に採用されたのは、業界に先駆けて大容量NVDIMM搭載を可能にしたHPE ProLiant DL360サーバーである。

チャレンジ

不揮発性メモリNVDIMMと その性能を引き出すシステムライブラPMDK

世界屈指の通信事業グループを率いる日本電信電話株式会社（NTT）。NTTは、持株会社としてNTTグループ全体の経営戦略をリードするとともに、基礎的研究開発において重要な役割を担う。現在、サービスイノベーション総合研究所、情報ネットワーク総合研究所、先端技術総合研究所と大きく3群の総合研究所を擁し、それぞれの傘下に専門分野に特化した研究所を編成している。

「私たちが所属するソフトウェアイノベーションセンタは、オープンソースソフトウェア（OSS）による基盤開発を中心に、ソフトウェア技術の研究からプラットフォームの開発・運用・保守に至るまで一元的に取り組んでいます。IoTやAIを活用した革新的なサービスの創出を支え、TCOの削減に結びつく研究開発とともに、NTTグループの事業会社がOSSを安定的に利用するための開発やサポートにも注力しています」と、分散処理基盤技術プロジェクト グループリーダ 主幹研究員の日高東潮氏は説明する。

独自の技術開発に強みを持ってきたNTTグループが、OSSを軸とするコ・イノベーションに舵を切ったのは2000年代初頭。2012年に設立されたNTT ソフトウェアイノベーションセンタは、多数のコミッターやコアデベロッパーを輩出するなど、OSSコミュニティへの幅広い貢献で知られている。PostgreSQLは同センタが注力するOSS製品のひとつだ。

「分散処理基盤技術プロジェクトでは、基盤システムソフトウェアをより競争力の高いソリューションとして活用するための研究開発に取り組んでいます。その一環として、超高速ストレージデバイスとして注目される『不揮発性メモリ（NVDIMM）』と、最新のシステムライブラリ『Persistent Memory Development Kit（PMDK）』を組み合わせたPostgreSQLの高速化に臨みました」（日高氏）

分散処理基盤技術プロジェクトのチームが検証に着手しようとしたのは2016年6月。この時点で、不揮発性メモリ（NVDIMM）を搭載可能なサーバーの製品化を計画していたのはHPEだけだった。検証プラットフォームに採用されたのは、計64GBのNVDIMMを搭載するHPE ProLiant DL360サーバーである。

ソリューション

HPE Persistent Memory搭載 HPE ProLiant サーバーを検証環境に採用

HPE ProLiantサーバーで利用できる不揮発性メモリ製品「HPE Persistent Memory」は、業界に先駆けて2016年4月から提供されている。業界標準のNVDIMM-N技術をベースに、メモリならではの高速アクセスと永続的にデータを保持するストレージの特長を兼ね備えたソリューションだ。「HPE Smartストレージバッテリー」を利用し、サーバー本体の電源が失われても、最大16基のHPE Persistent Memoryに対してフラッシュデバイスで安全にデータを保護することができる。

「私たちのチームでは、オペレーティングシステムや仮想化技術など、基盤システムソフトウェアに関わる研究テーマの一環として『データストアの高速化』を掲げています。AIやIoTを存分に活用するための基盤システムには、より高速なデータストアが欠かせません。不揮発性メモリ（NVDIMM）と、そのパフォーマンスを最大化するシステムライブラリ（PMDK）が突破口になると考えました」と、分散処理基盤技術プロジェクト主任研究員の石崎晃朗氏は話す。

チームがNVDIMM向けに最適化された「Non-Volatile Memory Library（NVMライブラリ）」に着目したのは2015年のことだった。インテルが開発を主導するNVMLは、2017年12月にPMDK（Persistent Memory Development Kit）と名称を変更している。*http://pmem.io/

「HPEからNVDIMMを実装する商用サーバーが登場し、不揮発性メモリの本格的な普及を確信しました。私たちはさっそくHPE ProLiant DL360サーバーを採用して、NVDIMMとPMDKによる検証に着手しました。これまで蓄積してきたPMDKのノウハウを、I/O負荷の高いPostgreSQLに適用することでどれだけの効果が得られるか、具体的な数値として示したかったのです」（石崎氏）

「NVDIMMとPMDKによるPostgreSQL高速化の効果」を数値化するこの検証が、アプリケーション開発者やOSSコミュニティへ与える影響は計り知れない。

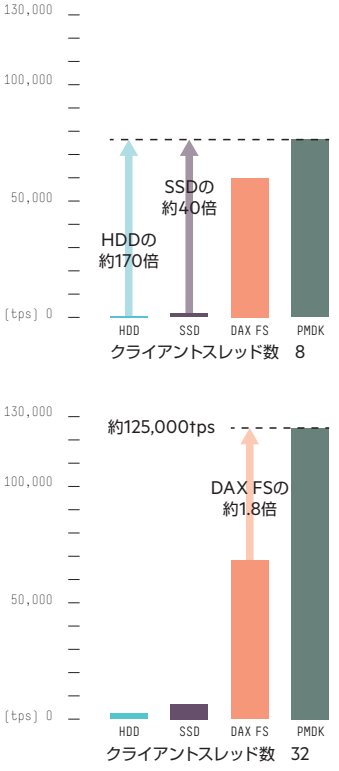
「従来のアプリケーションは『ストレージI/Oが低速である』ことを前提に、高速なメモリへの

HPE Persistent Memory(32GB)搭載
HPE ProLiant DL360 Gen9

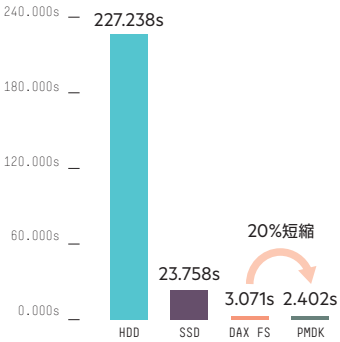
検証環境
HPE ProLiant DL360 Gen9
Xeon E5-2667 v4(3.20GHz)×2(HT 無効)
DRAM:DDR4-2400 32GiB ×2
NVDIMM:DDR4-2133 32GiB ×2(4枚ごとに32GiBへインタリーブ)
OS:Ubuntu 16.04
Linux kernel:4.12
PMDK:65e8122(master@ Aug 30, 2017 改良版)
PostgreSQLベース:00f6d5c (master@Aug 29, 2017)

検証項目　32GBのNVDIMMをデータストアとして検証を実施
①更新処理の性能改善:WAL(Write-ahead logging)へのPMDKの適用
②チェックポイント処理の性能改善: RelationへのPMDKの適用

挿入データサイズ　1KBの計測結果



チェックポイント処理時間



キャッシュなど様々な工夫を施してきました。不揮発性メモリ(NVDIMM)の実用化によって、そうしたプログラミングのパラダイムが一変する可能性を秘めています」(日高氏)

SSDとの比較で最大約40倍
HDD比で最大約170倍のWAL高速化を実証

従来のHDD/SSDには、ランダムよりもシーケンシャルにアクセスした方がより高いI/O性能を引き出せるという特徴がある。OSSのリレーショナルデータベースPostgreSQLでは、データ更新時にストレージI/Oがボトルネックになることを回避するため、ストレージに書き込むデータをDRAM内でバッファリングし、ストレージにシーケンシャルにデータを書き込む機能を備えている。

このように、データベースのような更新負荷が高いシステムでは、DRAM内でデータをバッファリングし、シーケンシャルに書き込むテクニックは一般的だ。だが低速なHDDと違い、NVDIMMでは、ページキャッシュなどを一度DRAMに乗せてからNVDIMMに保存するような手順にはムダがある。アクセス方法を変えることで、PostgreSQLの劇的な高速化が図られる可能性は高い。

「PMDKでは、OSやファイルシステムを介さずダイレクトにNVDIMMへアクセスします。この特長を活かすアプローチとして、PostgreSQLにおける更新処理の性能改善、チェックポイント処理の性能改善の2つをテーマにNVDIMMへのPMDK適用の効果を測定しました」(石崎氏)

①更新処理の性能改善:WAL (Write-ahead logging)へのPMDKの適用

WALは、トランザクションがコミットされたことを保証するために、変更されたログファイルを先行してディスクに書き込む処理。ログファイルの同期を低遅延で行うためにシーケンシャルに書き込まれる。

②チェックポイント処理の性能改善:RelationへのPMDKの適用

チェックポイントは、ログ内の情報を反映するためにあるタイミングのフルセットのデータを永続化させる処理。チェックポイントによって、全てのデータファイルがディスクに書き出された状態になる。

「PostgreSQLでは、同期書き込み処理のレスポンス遅延がトランザクション性能に大きな影響を及ぼします。WALの処理にNVDIMMとPMDKを適用し、ログの書き込み時間を短縮できれば、全体の性能向上につながると考えました」と、①更新処理の性能改善を担当した分散処理基盤技術プロジェクト 研究主任の一柳淑美氏は話す。

検証結果は驚くべきものとなった。NVDIMMとPMDKの組み合わせは、NVDIMMとDAXファイルシステムの組み合わせに対し、実に1.8倍というスループットを発揮したのである。SSDとの比較で最大約40倍、HDDに対しては実に最大約170倍もの高速化が実証された。

「②チェックポイント処理は、トランザクションの設計やワークロードの状況によって、一度に大量のディスク書き込みが発生することがあります。ここにNVDIMMとPMDKを適用すれば高い効果が得られると考えました」と話すのは、分散処理基盤技術プロジェクト 研究員 毛受崇氏である。

この検証でも圧倒的な効果が明らかになった。NVDIMMとPMDKは、HDDやSSDと桁違いの時間短縮を達成し、NVDIMM+DAXファイルシステムよりも約20%短縮された。

「WALへの適用で更新スループットを80%向上、Relationへの適用でチェックポイント処理時間を20%短縮という結果を出すことができました。PMDKは、DAXファイルシステムを大きく上回る高速化が可能で、プログラミングの工夫やアプリケーションの改修でさらに大きな改善が期待できます。PostgreSQLの高速化には、まだ大きなポテンシャルがあるのです」(一柳氏)

「今回の検証では、PostgreSQLのアプリケーションロジックを変えずに、書き込み処理部分にわずかなコード変更を加えるだけでPostgreSQLにPMDKを適用することができました。NVDIMM本来の性能を引き出すためには、ソフトウェアの内部に手を入れることが必要不可欠です。今後は有識者と連携して内部コードの改善を検討し、OSSコミュニティへの提案を進めていきたいと考えています」(毛受氏)

Case study
日本電信電話株式会社

業界
通信

ソリューション概略

導入ハードウェア
- HPE ProLiant DL360サーバー
- HPE Persistent Memory（NVDIMM）

導入ソフトウェア
- Persistent Memory Development Kit（PMDK）
- PostgreSQL

“先進的にNVDIMMを取り入れられているHPE ProLiantサーバーには、インメモリコンピューティングの推進において、今後も大きな役割を担っていただきたいと思います。HPEの先進的なハードウェアテクノロジーが、私たちの研究開発をさらに加速させてくれることを期待しています”

日本電信電話株式会社 NTT ソフトウェアイノベーションセンタ 分散処理基盤技術プロジェクト
グループリーダ 主幹研究員　日高 東潮 氏

ベネフィット

インメモリコンピューティングから
メモリドリブンコンピューティングへ

HPEが提供する不揮発性メモリ「HPE Persistent Memory」は、8GBおよび16GB NVDIMM製品として提供されている。さらに、1TBという大容量を利用可能にする「HPE Scalable Persistent Memory」も発表されている。汎用的なDRAMにバッテリーを組み合わせて、不揮発性メモリのように利用できるパッケージ製品だ。

「HPE Scalable Persistent Memoryのようなエンタープライズクラスの製品を、業界に先駆けて提供し続けるHPEには、サーバーベンダーとしての底力を感じます。個人的な見解ですが、サーバーが4TB以上の不揮発性メモリを搭載できると、オンライントランザクション環境に適用できるユースケースが大きく広がると考えています。ソフトウェアの最適化との両輪で一気に普及が進むでしょう」と日高氏は言う。

データベース性能への要求の高まりはとどまるところを知らない。パフォーマンスチューニングはいまだに大きな課題のひとつだが、根本的に解決される日は近いかもしれない。

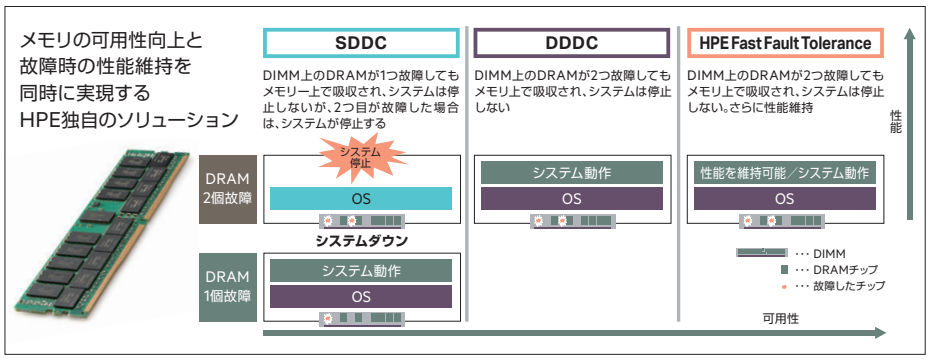
「ストレージI/Oのボトルネックを考える必要のない世界が、目前まで迫っています。データベースエンジニアはパフォーマンスチューニングから解放され、アプリケーション開発はキャッシュやバッファを考慮する必要のないシンプルなものになるでしょう。一昔前のシンプルなアプリケーションに戻るということかもしれません」と一柳氏は話す。

ハードウェアやデバイスの進化がソフトウェアの進化を促す関係は、これからも繰り返されていく。イノベーションが加速する中、その利用価値を見極めて、いち早く最新テクノロジーを活用することがいっそう重要になるだろう。

最後に、日高氏が次のように話して締めくくった。

「今後、不揮発性メモリの普及などの先に来るインメモリコンピューティングという大きな潮流の中、HPEが『The Machine』で提唱しているメモリドリブンコンピューティングのように、様々な技術が提言されると思います。そのような中で先進的にNVDIMMを取り入れられているHPE ProLiantサーバーには、インメモリコンピューティングの推進において、今後も大きな役割を担っていただきたいと思います。HPEの先進的なハードウェアテクノロジーが、私たちの研究開発をさらに加速させてくれることを期待しています」

詳しい情報
HPE ProLiant DL360サーバーについてはこちら
www.hpe.com/jp/proliant



お問い合わせはこちら
カスタマー・インフォメーションセンター
0120-268-186（または03-5749-8279）
月曜日～金曜日 9:00～19:00
（土曜日、日曜日、祝日、年末年始、および5月1日お休み）

日本ヒューレット・パッカード株式会社
〒136-8711 東京都江東区大島 2-2-1

ぜひご登録ください

Hewlett Packard Enterprise

CHS00006-01　記載事項は個別に明記された場合を除き2018年1月現在のものです。