

Windows Server 2012とInfiniBandで、 ファイル入出力を効率化、計算効率向上を実現



Version11はWindowsServer2012対応



Windows Server 2012 およびWindows 8 は、Windows のファイル共有プロトコルであるサーバーメッセージブロック(SMB) の最新バージョン、SMB 3.0 を採用しています。SMB はレガシWindows との互換性や非Windows との相互運用性を提供するため、長い間、レガシなプロトコル(SMB 1.0) のままでした。Windows Server 2008 およびWindows Vista では、再設計されたSMB 2.0 が採用され、トラフィックの削減とファイル転送の高速化が行われ、Windows Server 2008 R2 とWindows 7 のSMB 2.1 でさらなるファイル転送の高速化がなされました(SMB 1.0 の約3.5 倍)。SMB 3.0 では、以下に示す新機能が搭載され、パフォーマンス、スケーラビリティ、可用性、セキュリティ、運用のすべての面で強化されています。SMB 3.0 の新機能のメリットは、Windows Server 2012 ベースのファイルサーバーに対する、Windows 8 クライアントおよびWindows Server 2012 上のアプリケーションから利用できます。

SMB マルチチャンネル

クライアントとサーバーの両方が、複数のネットワークアダプターを持ち、ファイルアクセスに複数のパスが使用可能である場合、SMB マルチチャンネルにより、ネットワーク帯域幅の集約による高速化と、冗長化によるフォールトトレランスを実現します。

SMB ダイレクト(SMB over RDMA)

RDMA (Remote Direct Memory Access) 対応のネットワークアダプター(InfiniBandなど) を使用することで、リモートディスクへのアクセスが大幅に高速化され、ローカルディスクに匹敵するI/Oパフォーマンスを実現できます。※Windows8未対応

SMB スケールアウト(スケールアウトファイルサーバー)

フェールオーバークラスター上にアクティブ/アクティブ構成の高可用性ファイルサーバーを展開できます。スケールアウトファイルサーバーは、クライアントの負荷分散とネットワーク障害時の継続的なアクセスを可能にします。

SMB 透過フェールオーバー

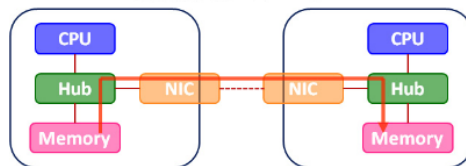
スケールアウトファイルサーバーの共有フォルダーに対するアクセスにおいて、ノードの障害やメンテナンス時に、自動的に別のノードに再接続する機能です。

SMB ディレクトリリリース

ブランチオフィスの応答時間を改善する機能です。遅延の大きいWAN リンクを介したアクセスにおいて、ディレクトリ情報とファイルのメタデータのキャッシュを提供します。



RDMAのイメージ



MPI (メッセージパッシングインタフェース)

高い伝送効率を実現するチャネルアーキテクチャ

PCI ExpressのロードストアアーキテクチャはCPUが接続するI/Oデバイスと直接送受信するため、遅いI/Oデバイスの使用時や接続数が多くなるとオーバーヘッドが大きくなる。

チャネルアーキテクチャではCPUは必要なデータをメモリに書き込んだらあとは別のタスクに移行できる。I/Oデバイスは自分の速度に合ったペースでメモリからデータを取得する。

I/Oデバイスの数が増えてもオーバーヘッドは小さいため、大規模システムにも対応可能。

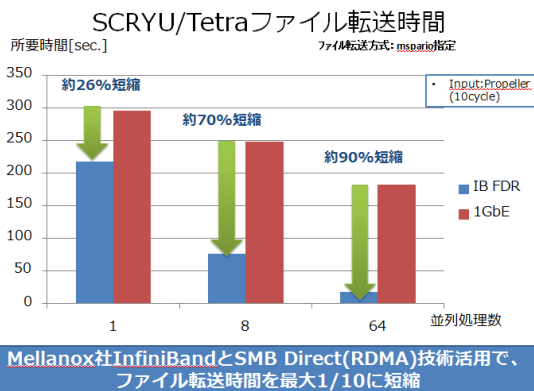
メラノックス社のConnect-XシリーズHost Channel Adapter用コントローラチップが提供する様々なオフロード機能により、さらに効率の良い、レイテンシーの低い通信が実現できます。

RDMA (Remote Direct Memory Access)

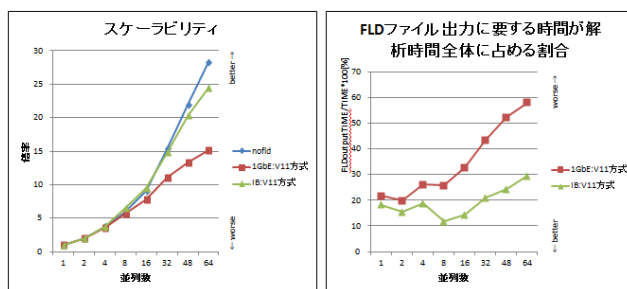
独立したシステムのメインメモリ間、システムとストレージ間のデータ転送を効率化するテクノロジー。メモリ間のデータ転送時にはホストCPUの割り込みがほとんど発生しないため、ホストCPUのオーバーヘッド低下に加え、データ転送時のレイテンシー短縮にもつながる。

10ギガビットEthernetを利用する場合、システム構築や運用管理がInfiniBandよりも容易という利点を持つ。Mellanox社NICでは、Ethernet上でRDMA機能を実現するRoCE (RDMA over Converged Ethernet)もサポートしており、HPC以外の分野での導入が進んでいる。

従来性能との比較



STREAM: 1GbEとIB(SMBdirect)の比較



- IB(SMBdirect)ではFLDファイル出力無しの場合に近いスケーラビリティを見せる

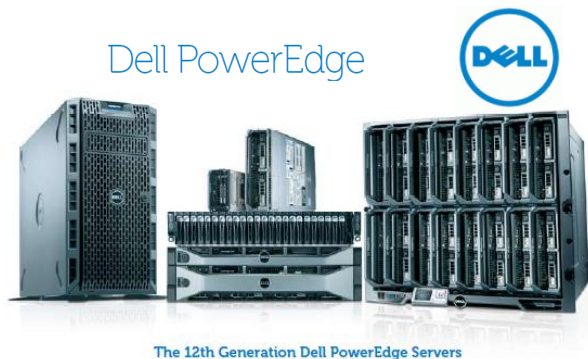
ベンチマーク実施環境

- FILESERVER**
 - MACHINE: Dell PowerEdge R720
 - OS: Windows Server 2012 + HPC Pack 2012
 - CPU: Intel Xeon E5-2665@2.40GHz
2CPU16core(HTTon-32core)
 - MEM: DDR3-1600MHz 8GB*12
 - NIC: Intel GbE I350c rNDC, Intel GbE I350-3
 - IB: Mellanox ConnectX-3 (FW:2.11.500,Driver:4.401.4223.0)
 - SSD: DELL_P320h-MTFDGL350SAH * 4台
- COMPUTE-NODE (4nodes)**
 - MACHINE: Dell PowerEdge R620
 - OS: Windows Server 2012 + HPC Pack 2012
 - CPU: Intel Xeon E5-2690@2.90GHz,2CPU16core(HTTOff)
 - MEM: DDR3-1600MHz 8GB*12
 - NIC: Intel GbE I350c rNDC, Intel GbE I350-T4
 - IBHCA: Mellanox ConnectX-3 (FW:2.11.500,Driver:4.401.4223.0)
- InfiniBandSwitch: Mellanox SX6036**
- ソフトウェア**
 - SCRYU/Tetra V11 beta
 - SCTsolver_Bx64net.exe: 6111.20300.20130730
 - sctsol_Dx64net.exe: 6511.22301.20130725
 - STREAM V11 beta
 - Stsolver_Bx64net.exe: 1111.20300.20130730
 - sctsol_Dx64net.exe: 1511.22300.20130726
 - その他
 - MPI:MSMPI, SMB: v3.0

上記ベンチマーク環境
OS(ソフトウェア)
ハードウェア

Windows Server 2012+HPC-Pack2012
DELL PowerEdgeシリーズ + Mellanox FDRシステム

SMB Direct検証システムと各種製品の御紹介



Rack : R910 R820 R815 R720 R720xd R715 R620 R520 R515 R420 R415 R320 R310 R210II
Tower: T620 T420 T320 T310 T110II
Blade : M1000e / M910 M915 M820 M620 M520 M420
Cloud : C6145 C8220 C8220X C6220 C5220 C410x

Intel社製最新CPU(Ivy-Bridge)を搭載した各種サーバを御用意しております。



SMB Directに対応したInfiniBandならびにEthernetスイッチならびにアダプタをご用意しております。

各メーカーホームページ情報

デル株式会社
<http://www.dell.co.jp/>
メラノックステクノロジーズジャパン株式会社
<http://www.mellanox.co.jp/>

連絡先ならびに資料作成

株式会社HPCソリューションズ
03-5640-7858
info@hpc-sol.co.jp

GPU仮想化を利用したアプリケーション処理の効率化

Windows Server 2012+SMB Directは、計算におけるファイル入出力の効率向上以外にも有用になります。昨今話題となっている画像処理を必要とするアプリケーション処理専用端末の仮想化システムのバックボーンとして十分に利用できます。

各処理端末にデータを転送するのではなく、ローカルと同じ使い勝手、同じ性能を提供するSMB Directは、ファイルサーバ上のデータを高速で読み取り、アプリケーション処理に利用できます。

※なおSMB DirectとGPU仮想化を同時に利用すること将来対応予定となっております。

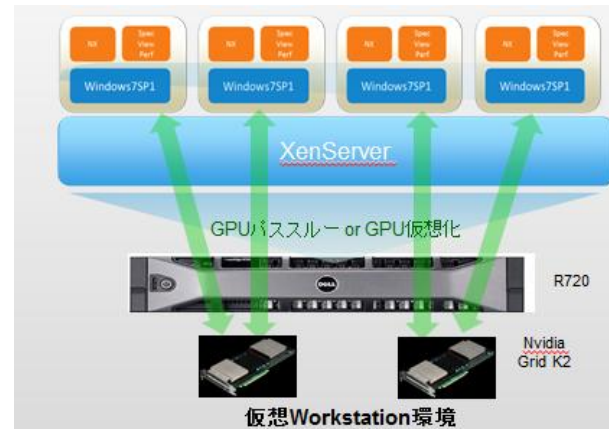
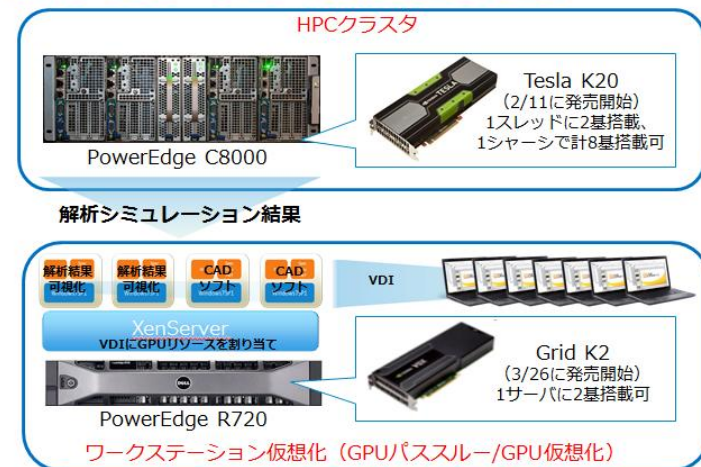
優位点

- ・クラスタの計算結果データをローカルへ北へする時間が不要となる。
- ・ファイルがサーバとローカルの両方に存在し、ディスク容量を圧迫していたことが改善される。
- ・ファイルのバックアップを行っている場合は対象が絞られる。
- ・ローカルにファイルが存在しないのでセキュリティが高まる。
- ・アプリケーションに応じて複数のワークステーションを物理的に配置する必要がなくなる。

実現するためには

- ・アプリケーション処理をサーバ上で行うことが必要があり、NVIDIA社提供Grid K1/K2と仮想化技術を組み合わせたGPU仮想化/GPUパススルー技術を活用。

解析シミュレーション可視化処理との連携



NVIDIA Grid K1/K2

	GRID K1	GRID K2
GPUの数	4 x ハイエンドKepler GPU	2 x ハイエンドKepler GPU
総NVIDIA CUDA コア	768	3072
総メモリサイズ	16 GB DDR3	8 GB GDDR5
最大出力	130 W	225 W
ボードの長さ	10.5"	10.5"
ボードの高さ	4.4"	4.4"
ボードの幅	デュアルスロット	デュアルスロット
ディスプレイIO	なし	なし
補助電源	6ピンコネクタ	8ピンコネクタ
PCIe	x16	x16
PCIe の世代	Gen3 (Gen2と互換性あり)	Gen3 (Gen2と互換性あり)
冷却機構	パッシブ	パッシブ

各メーカーへの問い合わせ情報

デル株式会社
<http://www.dell.co.jp/>
 メラノックステクノロジーズジャパン株式会社
<http://www.mellanox.co.jp/>
 NVIDIA
<http://www.nvidia.co.jp/>

連絡先ならびに資料作成

株式会社HPCソリューションズ
 03-5640-7858
info@hpc-sol.co.jp

参考構成資料

Dell PowerEdge™ C8000シリーズ



大規模コンピューティング用の構築ブロック

PowerEdge™ C8000シリーズを使用すれば、1つのシャーシ、1つのラック、さらにはスケールアウト環境全体から最大限の生産性を引き出せます。

卓越したインテルのパフォーマンス

効率性と汎用性に優れたインテル® Xeon® プロセッサ E5-2600 製品ファミリーを活用して、スケールアウト環境を強化してください。C8000シリーズの演算ノードには、パフォーマンス、搭載機能、およびコスト効率のバランスを最高水準で実現した、これら次世代プロセッサが2基搭載されています。

スマートな拡張

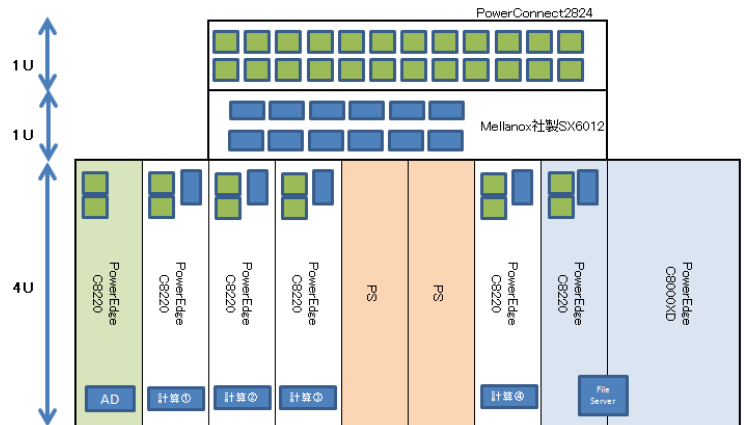
効率性と柔軟性に優れた冗長共有インフラストラクチャシャーシによって、パフォーマンスや拡張性を増進することなく、スペースと予算の問題を解消できます。

Mellanox SX6012



SX6012はQSFP+1枚にて12portを有するFDR(56Gbps)対応のSX6012FとQDR(40Gbps)相当のFDR-10対応のSX6012Tを御用意しております。
また、マネジメント機能を内蔵しないSX6005F、SX6005Tも御用意しております。

ラックマウント6Uサイズに認証サーバ (ActiveDirectory), ファイルサーバ (Disk装置)、計算サーバ4台、インフラスイッチ、LAN接続スイッチを搭載できます。CPU、メモリ、HDD容量をご自由に選択できます。さらには、計算サーバの1台にグラフィックス仮想化用カードを追加することも可能です。参考価格は、担当営業にご相談ください。その他、ラック、UPS、ケーブル、マウス、長期保証などもそろえております。



各メーカーへの情報

デル株式会社
<http://www.dell.co.jp/>
 メルノックステクノロジーソリューションズ株式会社
<http://www.mellanox.co.jp/>

連絡先ならびに資料作成

株式会社HPCソリューションズ
 03-5640-7858
info@hpc-sol.co.jp