

High Frequency Trading, ビッグデータ分析を支えるIT

―日本の金融業界におけるGPUコンピューティング―

古橋 勇作　日本ヒューレット・パッカード

金融システムへの情報技術活用事例

HFT（High Frequency Trading）は，日本語では超高頻度取引と呼ばれ，株やオプションなどの証券取引において，コンピュータ技術を駆使して通常の取り引きでは考えられないようなスピードで売買を繰り返す手法である．HFTにおける重要なポイントは，単に高速な技術を取り入れるのではなく，対象システムに最適な技術要素を組み合わせ，いかに1番速いスピードを持ったシステムで，最も有利な条件で取り引きを成立させることができるかにある．この競争は年を経るごとに過激さを増し，数年前は秒単位だった取引スピードはミリ秒単位となり，ついにはマイクロ秒単位の世界に突入している．2011年夏にはシンガポール証券取引所が世界最速の執行スピードとなる90マイクロ秒のローレイテンシー（低遅延）システムをリリースした．1マイクロ秒の差が億単位の利益を生み出すため，証券各社はたった1マイクロ秒のスピード改善のために，導入後，数カ月しか経っていないシステムを惜しみなくリプレースしている．たとえリプレースコストがその都度数千万円に達しても，一番速いスピードで取り引きができれば一瞬で投資を回収することができるため，常識では考えられないライフサイクルでシステムを刷新し続けているのである．マイクロ秒単位の競争で生き残るために必要な技術要素は多岐にわたる．それらの技術要素のうち，今回はネットワークの高速化のために重要な技術であるInfiniBand，データアクセス高速化の要となるNAND Flash製品としてIOアクセラレータおよびViolin Memory，ビッグデータへの高速アクセス技術であるVertica，リスク分析高速化の重要な技術であるGPUなどを後続の章で触れていきたい．

ローレイテンシーを実現する技術要素

■ InfiniBandの特長

ローレイテンシーを追求するための重要な考慮点として，ネットワークのオーバヘッドの極小化が挙げられる．従来のイーサネットによるTCP/IP通信では，プロトコルにかかわる処理のオーバヘッドが大きく，マイクロ秒単位のローレイテンシー化の障壁となっていた．InfiniBandは2000年頃に主要ハードベンダにより業界団体が設立され，統一規格として成立した通信プロトコルで，当初は学術研究分野やスーパーコンピュータなどのHPC（ハイパフォーマンスコンピューティング）分野での採用が中心だった．現在は，広帯域，ローレイテンシーに秀でるネットワークとして，高速なノード間通信での採用が進んでいる．2000年代後半からは金融業界でもHFTなどを始めとして採用分野が広がっている．現在は第3世代となるQDR（Quad Data Rate）が主流であるが，2011年末から第4世代のFDR（Fourteen Data Rate）が登場し，PCIe Gen3（PCI Express 3.0）に対応したこともあり，実質的な性能は2倍近くまで向上している．またFDRは，40Gbイーサネットと同じQSFP＋インタフェースを利用している（図-1参照）．2012年から2013年にかけて，InfiniBandとイーサネットを相互通信

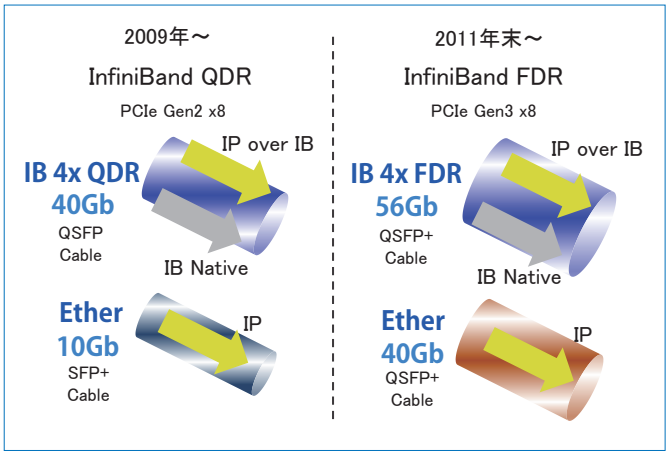


図 -1　現在主流の InfiniBand 規格

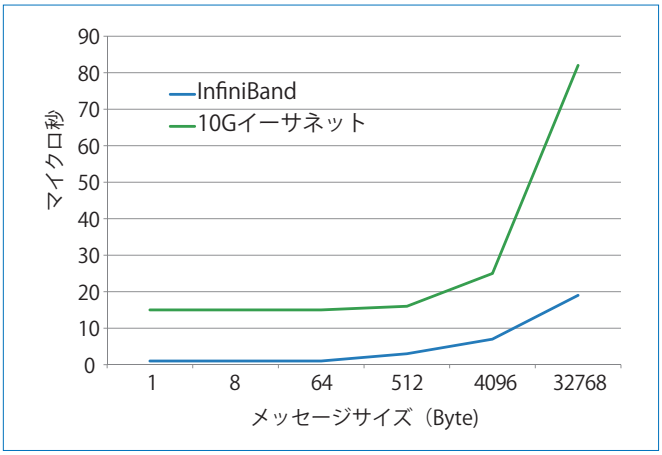


図 -2　イーサネットと InfiniBand の比較

可能なスイッチが登場する予定で，IP over IB 通信（IPoIB：InfiniBand で IP 通信を可能にするためのプロトコル）を利用してイーサネットと InfiniBand 間の通信が容易に実現可能となる．

InfiniBand を利用することで，図 -2 のような非常に低いレイテンシーの通信が可能となる．

■ InfiniBand が高速な理由

InfiniBand が従来のイーサネットに比べ高速な理由の 1 つとして，RDMA（Remote Direct Memory Access）通信が挙げられる（図 -3 参照）．通常のデータ転送処理では，アプリケーション側でデータを保存しているメモリ領域から，通信処理用のメモリ領域にコピーをする必要があり，このコピー処理は OS が CPU を利用して処理するため，オーバヘッドが存在していた．これに対し RDMA 処理は，アプリケーションのメモリ領域から直接転送先ノード側のアプリケーションのメモリ領域へコピーをする．

■ NAND Flash

NAND Flash とは一般的に知られている SAS（Serial Attached SCSI）や SATA（Serial Advanced Technology Attachment）インタフェースを持つ SSD（Solid State Drive）でも採用されているテクノロジーで，電子的に消去・書き換えすることが可能な不揮発性半導体デバイスである．特長としては，従来のハードディスクのような回転体や磁気ヘッドを持たず，シーク時間や回転待ち時間が発生しないため，非常に高速であることが挙げられる．容量あたりの単価は，従来のハードディスクに比べ高価であるが，価格あたりの性能単価としては従来のハードディスクよりも安くなっている（表 -1 参照）．

HP IO アクセラレータは NAND Flash 業界で有名な Fusion-io 社の OEM 製品である．PCIe インタフェースに直結するタイプであるため，SAS や SATA インタフェースのバス幅よりも高速な通信が可能となる（表 -2 参照）．書き込みスピードが高速かつ長寿命（10 年程度）であることも特長である．

金融業界で非常に重視される単体での可用性も加味した Violin Memory という製品も注目されている．この製品は HP IO アクセラレータと同様に PCIe インタフェース直結の NAND Flash であるが，加えて以下のような特長を持つ．

- ファイバチャネルの SAN（Storage Area Network）環境に接続して通常のディスクとして認識することが可能
- ハードウェア RAID 機能により 1 枚の NAND Flash の故障によってサービス停止の影響を受けない
- ホットスワップ機能によりサーバ無停止でのメンテナンスが可能
- ホットスペア機能により，複数の NAND Flash

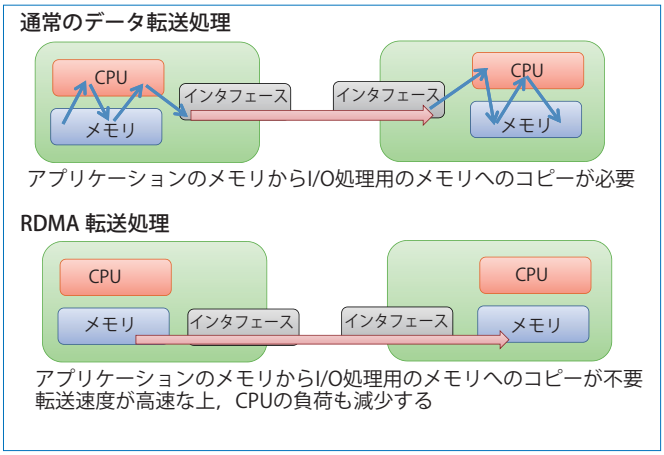


図 -3　RDMA（Remote Direct Memory Access）とは？





	従来型SASディスク	IOアクセラレータ	Violin Memory
数	700	**4**	**1**
容量	102TB	**9.64TB**	**5TB**
IOPS	252,000 IOPS	**260,000 IOPS**	**250,000 IOPS**
性能あたりの価格	¥151	**¥17**	**¥80**

表 -1　IOPS（1 秒間の IO 数）あたりの価格

を内部で自動的に予備の NAND Flash に置き換えることが可能

ビッグデータ関連技術の金融市場への活用事例

■ ビッグデータ活用事例

ビッグデータという用語は現在 IT における多くの分野で利用されている．ビッグデータの定義はさまざまだが，1 つには蓄積されるデータの量が非常に大きいという特長が挙げられる．金融市場において昨今課題となっているのは，HFT やアルゴリズムトレード，ローレイテンシー技術の進化によって急増するマーケットデータの扱いである．オプション取引のマーケットデータについては，2005 年では 1 秒あたりのメッセージ数が 10 万件程度であったが，2009 年時点で 200 万件に届く勢いで急増している．そのため，従来型のデータベースでは，時系列明細レベルでの保存期間が非常に限定的になっている．米国の大手投資銀行では，1 日に 500GB，1 カ月で 20TB に達しているマーケットデータを 1 年間保存するために列指向アーキテクチャによって高圧縮でデータ保存が可能な分析データベース Vertica を採用している．これによって投資手法を過去のデータで確認するバックテスティングやアルゴリズムの最適化，リスク分析，SOR（Smart Order Routing：複数市場に跨る商品について最良の条件を満たす市場で自動的に注文を執行する仕組み）などに活用されている．Vertica は，このほかに大手銀行が不正検知のシステムで活用している．

	1205GB MLC	2410GB MLC Duo
Read Bandwidth (1MB)	1.5 GB/s	3.0 GB/s
Write Bandwidth (1MB)	1.3 GB/s	2.5 GB/s
Read IOPS (Sequential. 512B)	445,000	892,000
Write IOPS (Sequential. 512B)	535,000	935,000
Read IOPS (Random. 512B)	143,000	285,000
Write IOPS (Random. 512B)	535,000	725,000
Read Access Latency	68 μs	68 μs
Write Access Latency	15 μs	15 μs
PCI インタフェース	PCI-Express 2.0 x4	PCI-Express 2.0 x8

表 -2　IO アクセラレータのスペック

■ 列指向データベースの特長

列指向データベースでは，従来型の行単位でデータを保存するデータベースとは異なり，列単位でデータを保存するため，同一データが連続する確率が高く，高い圧縮率を提供できる．同一データが存在しないタイプの列であっても，さまざまな圧縮アルゴリズムのうち，最適なパターンを自動的に選択して高圧縮率（50% ～ 90%）を維持する（図 -4 参照）．たとえば日付などのタイプは RLE アルゴリズム，顧客番号などユニークな数列を持つタイプは Delta Encoding，取引データなどランダムな数値の連続するタイプでは LZO アルゴリズムなどが採用される．

また列指向のもう 1 つの特長として圧倒的なアクセススピードの速さが挙げられる．従来型の行指

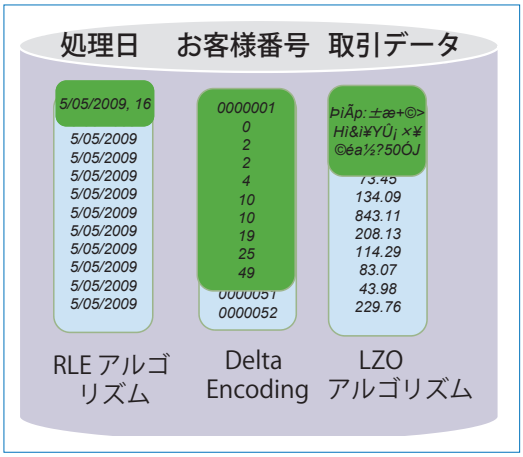


図 -4　Vertica の圧縮パターン

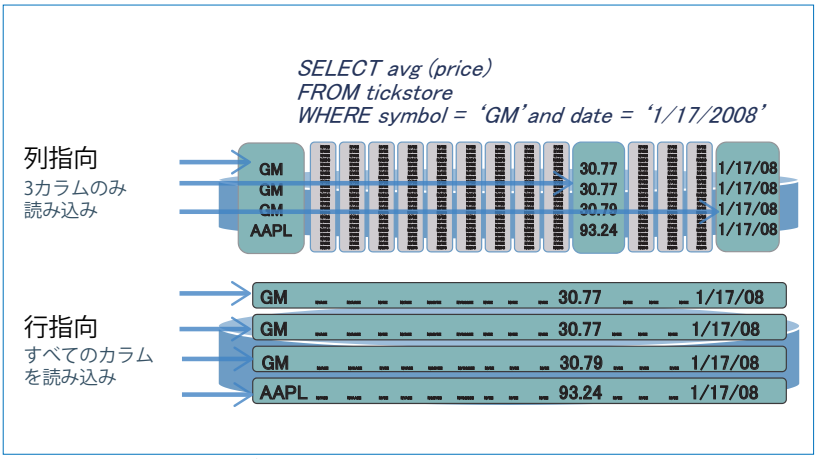


図 -5　列指向と行指向のデータアクセス方式

向データベースでは，インデックス化されていない列を検索する場合，必要ない列も含めてすべてのデータにアクセスする必要があった．これに対して列指向データベースの場合，処理に必要な列のみに限定してアクセスができるため，対象データ容量が大きければ大きいほど，アクセス効率が高くなる（図 -5 参照）．

GPU とリスク管理例

■ GPU とは

GPU とは Graphics Processing Unit の略称で，本来はゲームなどのグラフィック描画用途で利用される．その GPU の優れた描画処理性能が，描画を伴わない高速演算に活用できることが見出され始めたのは 2000 年頃だが，当初はグラフィック処理専用の言語を利用しての演算処理応用が大変困難であったため，適用範囲は学術研究用途が中心だった．通常の CPU で同等の環境を実装する場合，非常に高額となるため，モンテカルロ法やブラックショールズ法など，非常に大規模な演算を必要とするオプション取引のプライシングでは，コスト面や省電力，省スペースなどの点で GPU が大きな優位性を持つといえる（表 -3 参照）．

2012 年に登場した最新機種では，1 個の GPU に 3,000 個以上の計算コアを搭載できるようになった．これにより，単精度浮動小数点演算の性能は 4 倍以上の 6TFlops（1 秒間あたり 6 兆回の演算能力）となっている．

■ CUDA の登場

GPU の主要ベンダである NVIDIA 社は 2006 年に GPU での高速演算プログラミングを容易にすることを可能とする CUDA（Compute Unified Device Architecture）と呼ばれる言語を発表した．

CUDA 登場前の課題

- 演算処理記述に向かないグラフィクス API と GPU のハードウェアアーキテクチャを詳しく理解する必要があった
- グラフィック処理専用の言語で，複雑なプログラミングが必要だった
- アドレス指定が可能な読み込みや書き込みなどの基本的機能もなかった

CUDA 登場により解決された課題

- C/C++ での実装が可能となり，グラフィック処理の詳しい理解が必須ではなくなった
- 倍精度演算がサポートされ，性能改善の対象範囲が拡張された

CUDA は C 言語の拡張セットとなり，C もしくは C++ の知識を持っていれば，習得が非常に容易であることが特長となる．CUDA の登場により，GPU の利用範囲や利用頻度は飛躍的に拡張することになる．2012 年 5 月に開催された，NVIDIA 社の GTC（GPU Technology Conference）では，2008 年と 2012 年の CUDA の利用状況が劇的に変

化していることを説明している（表-4参照）．

GPUを活用したスーパーコンピュータは，2008年時点では，東京工業大学のTSUBAME1.2のみであったが，現在は上位5位のスーパーコンピュータのうち，3システムがGPU環境となっており，あわせて35システムまで増加している．論文件数では，ビッグデータ対応技術として話題に上がることの多いHadoop技術（大規模分散処理用Javaフレームワーク）に関する論文に対して4倍以上のCUDA論文が公開されている．CUDAの登場によってGPUの適用範囲は拡大され，数年前から金融工学の分野でも活用されるケースが増えている．ヒューレット・パッカードでは，既存のリスク管理プログラムをCUDAに移行してどの程度の性能改善が期待できるかをアセスメントするサービスを提供している．また，既存プログラムのサンプルをCUDAに移行した上での検証も行っている．

■ GPU導入の課題

GPU導入時には下記のような課題が考えられる．

- GPUシステムの実装ノウハウの不足
- CUDA言語開発ノウハウの不足
- パッケージ製品のGPU対応が少ない
- GPU利用を前提としたサーバ製品が少ない
- CPU-GPU間の通信帯域が不足

GPUシステムにおいては，GPUに特化した運用や性能監視，稼働状況，温度監視など，通常のサーバ製品とは異なるノウハウが必要となる．CUDA言語は，CやC++の知識があれば習得しやすいが，GPUの性能を最大限に引き出すためにはGPUの特長を踏まえた多くのチューニングポイントが存在するので，豊富な経験に基づいた適切で効率の良い開発やチューニングが重要となる．GPUに対応したパッケージ製品はまだまだ少ないが，欧米の製品を中心に徐々に増えている．過去に導入したシステムをカスタマイズ可能なパッケージとして導入可能としているベンダもある．GPU利用を前提としたサーバ製品としては，GPUベースのスーパーコンピュータとして有名な東京工業大学のTSUBAME2.0で利用されているヒューレット・パッカードのSL390モデルが存在する．このモデルは2Uハーフワイドのサーバに3個のGPUが搭載可能で，2個のGPUが専用のCPU-GPU間の通信バスを持つことが可能となっている．また，マルチGPU環境でのノード間の高速な通信を可能とするために，オンボードで高速なInfiniBandポートが搭載可能である．SL390は2012年に後継モデルとなるSL250が登場し，メモリやPCIバスのスピードが向上するなど大幅な改善が図られている．このようにGPU導入に伴う課題解決にはGPU環境の導入実績や経験豊富なベンダやメーカの選定が非常に重要となる．

モデル名	Tesla M2090
計算コア数	512
メモリサイズ	6GB
単精度浮動小数点演算	1,331GFlops
倍精度浮動小数点演算	665GFlops

表-3　Tesla GPUの性能

	2008年	2012年
CUDAダウンロード数	15万件	150万件
GPUによるスーパーコンピュータ	1	35
CUDAを教えている大学	60	560
CUDA論文	4,000	22,500

表-4　2008年／2012年のCUDA利用状況比較

■ GPU導入事例

金融業界でのGPU導入事例としては，ブルームバーグが債券のプライシング処理で大幅な性能やコスト改善を果たしている．CMO（不動産抵当証書担保債券）のような資産担保証券のプライシング計算を毎晩130万件実行していた．彼らは，モンテカルロ法に基づいたシングルファクタモデルの計算を2ファクタモデルとし，より精度の高いプライシングをすることを検討していた．システムのスケールアップを検討した結果，当時の800コアから合計8,000コアで，1,000台のサーバに増やす必要があることが判明した．そこで非現実的で，柔軟性に

データ型	言語	処理時間
倍精度浮動小数点演算	C#	570.6 秒
	CUDA	3.952 秒
単精度浮動小数点演算	C#	1531.8 秒
	CUDA	1.659 秒

表 -5　カブドットコム証券の GPU 導入効果

欠ける構成を避け，GPU によるコンパクトで計算能力の高い構成が検討された．最終的には 1,000 台のサーバ構成を避け，たった 48 個の GPU で，同等の処理を 16 時間から 2 時間と，1/8 に短縮することに成功した．サーバの利用スペースも 1/42 まで縮小し，導入コストは 1/28，消費電力は 1/3 と，大幅な性能とコストメリットを享受することとなった．

BNP パリバは，コンピュータのグリーン化への取り組みの中で，省電力で計算能力の高い GPU に注目し，500CPU を超える既存システムを，たった 2 個の GPU システムに置き換えることで，以下のようなメリットを得ることができた．

- 全体的な消費電力が 1/190
- 削減応答時間が 1/5 に短縮
- トレーダーの市況認識度が向上
- 運用コストと環境への影響の削減

BNP パリバは世界中で取り引きされる株式と，商品先物取引の金融モデルのパラメータ推測や，数値最適化アルゴリズムにおいて GPU を活用している．

日本国内では，カブドットコム証券が GPU ベースの高速演算システムを活用し，先物・オプション取引リスク管理ツール「先物 OP ナビ TM」の無償提供を 2012 年春より開始している．このツールは，日経 225 オプション取引における全口座のリスク計算を秒単位の間隔で実現している．オプション取引の指標として用いられているデルタ，ガンマ，ベガ，セータといった Greeks 指標をリアルタイムに表示する．損益チャートにてリアルタイムのリスク状況を可視化し，通知機能も備えている．導入に先立って日本ヒューレット・パッカードが検証を実施し，表 -5 のような導入効果を事前に確認することができた．

高速化の応用例

これまでの技術要素を組み合わせた高速化構成の例としては，InfiniBand と IO アクセラレータを組み合わせた高速ストレージ構成などが考えられる．さらに Fusion-io 社の DirectCache を利用することで，IO 処理サーバ側にも IO アクセラレータを実装し，ストレージキャッシュのように扱うことも可能だ．これは従来型の FC ストレージの IO 処理遅延に悩む環境では，既存のアプリケーションやストレージの構成を変更せずに IO 性能を改善できるため，緊急で問題を回避したい場合などに有効である．ほかにも GPU と InfiniBand を組み合わせ，複数 GPU サーバの通信速度を改善するために，Mellanox 社の Direct GPU を利用する方法もある．多数のサーバのメモリ上に分散してデータを保持することで，データアクセス処理を高速化するインメモリデータグリッド製品である GemFire のノード間レプリケーションを，InfiniBand で高速化する手法も考えられる．このように，高速な技術を単純に導入するだけでなく，アセスメントや実際の機器を利用した事前検証などを活用し，システムの特性を深く理解することで，どの技術を，どのように組み合わせて最適なコストと効果のバランスを実現するかということが最も重要である．

(2012 年 5 月 31 日受付)

古橋 勇作 ‖ Yusaku.furuhashi@hp.com
日本ヒューレット・パッカード テクノロジー・コンサルティング事業部コンサルタント．欧米の最新動向を踏まえた，金融業界を始めとしたハイパフォーマンス技術の導入を推進している．