



インテル® I/O アクセラレーション・テクノロジー

ネットワークおよびストレージの I/O パフォーマンス、 信頼性、効率を高めてアプリケーションの応答速度を向上

今日、データの迅速な転送と処理はビジネスの成功に欠かせないものになっています。IT マネージャは、アプリケーションのデータ転送速度を向上させるために継続的に新しいネットワーキングおよびストレージ・インフラストラクチャに投資を行いパフォーマンスの向上に努めています。しかし、投資から価値を実現するために克服しなければならないボトルネックとして、サーバの I/O に要する待ち時間に関する問題が浮上してきました。信頼性と効率性を高めてアプリケーションへのデータ供給をさらに高速にする統合プラットフォーム I/O ソリューションのインテル® I/O アクセラレーション・テクノロジーは、このパフォーマンスの問題を解決します。

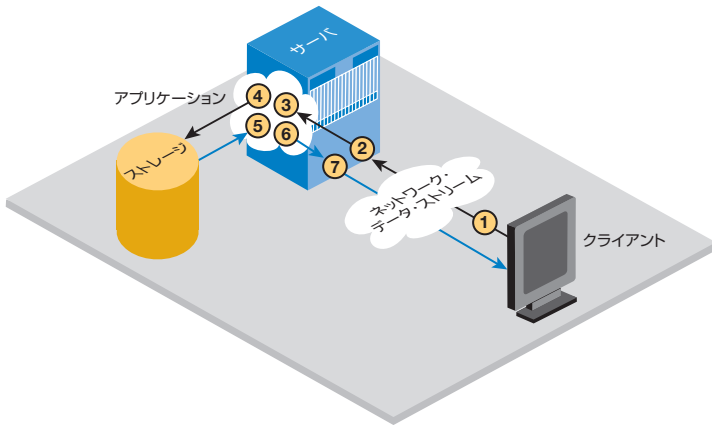
歴史的背景

これまで、ネットワーク・インフラストラクチャへの投資には、RAID (Redundant Arrays of Independent Disks) や 3 層データセンタといったアーキテクチャ面の強化やシステムの多様化に加えて、ファスト・イーサネットからギガビット・イーサネット (GbE) および 10 ギガビット・イーサネット (10GbE) への移行などがありました。最近までは、これらの進化ステージを通じて、ネットワーク・サーバのパフォーマンスをネットワーク・トラフィックの増加に対応させることは容易でした。

次第に、ネットワーク・トラフィックに対する要求がサーバの能力を凌ぐようになり、増加の一途をたどるネットワーク通信およびトランザクション処理ワークロードとのギャップが拡大を続けています。その結果、IT マネージャは、「ネットワークの帯域幅を 10 倍にする投資が行われたにも関わらず、アプリケーションの応答時間と信頼性に同等の改善が見られないのはなぜなのか」という疑問を感じています。

これは、1980 年代から I/O データの処理方法に顕著な変化が見られないことが原因です。特に、TCP/IP (Transmission Control Protocol over Internet Protocol) スタックはイーサネットやインターネットの開始後ほとんど変化していません。しかし、問題は、TCP/IP プロトコルの処理速度だけではありません。サーバの I/O のオーバーヘッド、メモリアクセス、ストレージの I/O スピードと信頼性もサーバの I/O 効率の悪さの原因となっています。

図 1. アプリケーションのデータベース



サーバのオーバーヘッドと応答の待ち時間は、応答要求のデータベース全体で発生します。これには、受信要求 TCP/IP パケットの処理、アプリケーションへのパケット・ペイロード・データのルーティング、保存情報の取り出し、ルーティングを目的とした要求元クライアントの TCP/IP パケットへの応答の再処理などが挙げられます。

これらは個々の要素の欠陥ではなく、包括的なソリューションが必要なシステム全体の問題です。インテルの研究開発チームはサーバの I/O アーキテクチャ全体に注目し、高速なギガビットおよび 10 ギガビット・イーサネットの速度とマス・ストレージの利点を活用する方法について研究してきました。その結果、より高速にアプリケーションとデータの受け渡しを行う Intel® I/O アクセラレーション・テクノロジーが開発されました。これを使用すれば、進化したネットワークとストレージの機能の利点を活用できます。

増加するサーバの I/O のトラフィック・ジャム

Intel® I/O アクセラレーション・テクノロジー (Intel® I/OAT) の重要性を理解するには、取り組んでいる問題の真の性質について理解する必要があります。理解する最適の方法は、サーバで要求の受け取り、処理、応答の処理を行うときにクライアントの要求フローを検証することです。図 1 は、このフローを図解したものです。以下の説明の番号は図中の丸で囲まれた番号に対応しています。

1. クライアントは、TCP/IP データ・パケット形式で要求を送信し、サーバは、そのネットワーク・インターフェイスを介して受信します。このデータ・パケットには、パケット識別子、ルーティング、クライアントの要求に関連する実際のデータ・ペイロードに関するヘッダ情報が格納されます。
2. サーバは、TCP/IP パケットを処理し、パケット・ペイロードを指定のアプリケーションに送信します。TCP/IP スタックを含むプロトコルの演算処理、パケット記述子とペイロードの移動に対するサーバによる複数のメモリアクセス、その他のさまざまなシステムのオーバーヘッドのアクティビティ (割り込み処理、バッファ管理など) などの処理が行われます。
3. アプリケーションは、クライアントの要求に対応し、要求に応答するためにストレージからデータを取得する必要があることを認識します。

4. アプリケーションは、ストレージにアクセスして、クライアントの要求を満たすために必要なデータを取得します。
5. ストレージは、要求されたデータをアプリケーションに戻します。
6. アプリケーションは、ストレージから受け取った追加データを使用してクライアントが要求した処理を完了します。
7. サーバは、ネットワーク接続を介して応答を TCP/IP パケットとしてクライアントに送信します。

上記のクライアントの要求およびサーバの応答サイクルは、ネットワークの開始時から存在し、引き続き今日のネットワークにおいて欠くことのできない処理として存在しています。トラフィックのボリュームが少なく、スピードが遅かった頃は、サーバはこのタスクを実行するのに十分なパフォーマンスを備えていました。現在では、パケット・トラフィックのボリュームとスピードはかなり高まりました。パケットのボリュームとスピードの高まりにより、ネットワークに対する CPU のオーバーヘッドが直接増加し、これに対応してアプリケーションに割り当てられるサイクルが減少しました。サーバ・アプリケーションのスループットは増加し、データセンタのパフォーマンスは低下しました。

Intel® I/OAT – 多角的問題に対する多角的ソリューション

図 1 では、アプリケーションの応答待ち時間は多角的でシステム全体の問題であることを示しています。アプリケーションへのペイロードの受け渡しには、パケットの受け取り、認識、処理の実行が必要です。アプリケーションは、クライアントの要求に基づき動作する際に、多くの場合、ストレージから必要なデータを取り出し、サーバに応答を返す必要があります。最後に、サーバでは、応答を TCP/IP パケットに変換し、クライアントに転送する必要があります。

Intel® I/OAT では、全体の問題を解決するためにネットワーク I/O アクセラレーションとストレージ I/O アクセラレーションの 2 種類のアプローチが使用されています。図 1 では、ネットワーク I/O アクセラレーションはパケットの変換およびパケットに関連したデータの移動に関する問題を解消し、ストレージ I/O アクセラレーションはデータ・ストレージのアクセスと信頼性に関する問題を解消しています。アプリケーションを変更せずにアプリケーションのスループットを改善してより高速にアプリケーションとデータの受け渡しを行うことを目的としています。Intel® I/OAT は、効率的なプロトコル処理、データの移動による CPU のオーバーヘッドの削減、信頼性の高いデータ転送の実現により、アプリケーションの応答時間を改善します。結果として、Intel® I/OAT は、既存のアプリケーションの健全性を維持し、同時に、将来のプロセッサの進化にも対応するプラットフォーム・アクセラレーション・テクノロジーを提供します。

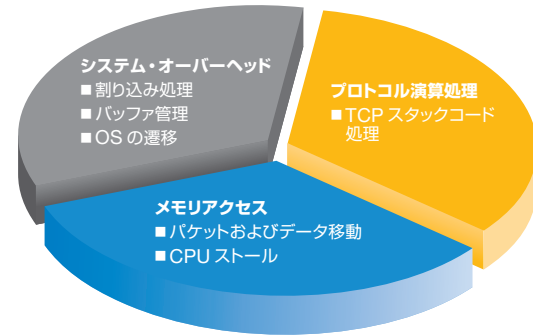
サーバのネットワーク I/O パフォーマンスの向上

図2では、インテル® I/OATのネットワーク I/O 部分が解決するシステム・オーバーヘッド・セグメントを示しています。このセグメントのサイズは、パケット・ペイロードのサイズによって相対的に異なるので注意してください。例えば、TCP プロトコルの演算処理では、電子メールのメッセージや金融取引で一般的な 1K のペイロードは 12% のセグメントを構成しますが、大規模なファイルの転送またはバックアップで一般的な 64K のペイロードは約 2 倍の 22% のセグメントを構成します。同様に、メモリのアクセスは 1K のペイロードでは 30% ですが、64K のペイロードでは 53% に増加します。

TCP 処理とメモリのアクセスについては、TOE (TCP Offload Engine) と RNIC (Remote Direct Memory Access) がすでにこの問題を解消していると考えられています。

これは正解でもあるし、不正解でもあります。TOE はインターフェイス指向で、TCP に注目したアプローチで、図2で示しているその他の処理ボトルネック・セグメントを完全には解決していません。TOE は大規模なバックエンド・ファイルの転送に効果が限定され、デュアル・プロセッサ・サーバ・マーケット・セグメントの 90% を構成するフロントエンドで動作するアプリケーションや中間層アプリケーションのパフォーマンスを実際は低下させます。図3で示しているように、TOE はバックエンド・サーバで最も効果を発揮しますが、インテル® I/OAT はデータセンタ・サーバとアプリケーション全体で一貫して高いパフォーマンスを提供します。

図2. ネットワーク I/O 処理タスク



ソース: インテル® Linux® 解析

サーバのネットワーク I/O 処理タスクは、TCP/IP パケット・サイズによって合計オーバーヘッドの割合が異なる 3 つの主要なオーバーヘッド・カテゴリに分類することができます。

RNIC については、RDMA (Remote Direct Memory Access) プロトコルで、アプリケーションのメモリ領域へのペイロード・データの直接書き込みをサポートし、データの移動オーバーヘッドを削減しています。RDMA プロトコルは、TCP/IP のような既存のネットワーク・プロトコルであり、それ自体のオーバーヘッドを示して各データ転送を編成します。この負荷を CPU に置く代わりに、RDMA では TOE エンジンを利用として使用してこの新しいワークロードを実行します。TOE と RDMA を組み合わせると、データの移動が改善され、2 つのプロトコルを処理する負荷が分散されます。ただし、RDMA を使用するには RDMA を認識するようアプリケーションを修正する必要があるため、RDMA プロトコルを実装できるのは一部のサーバ環境に限定されます。さらに、RDMA はエンド・ツー・エンド・プロトコルなので、データを転送するシステムすべてに RDMA を実装する必要があります。この実装問題とアプリケーションの選択肢が狭いために、業界への RNIC の浸透は限定的なものになっています。

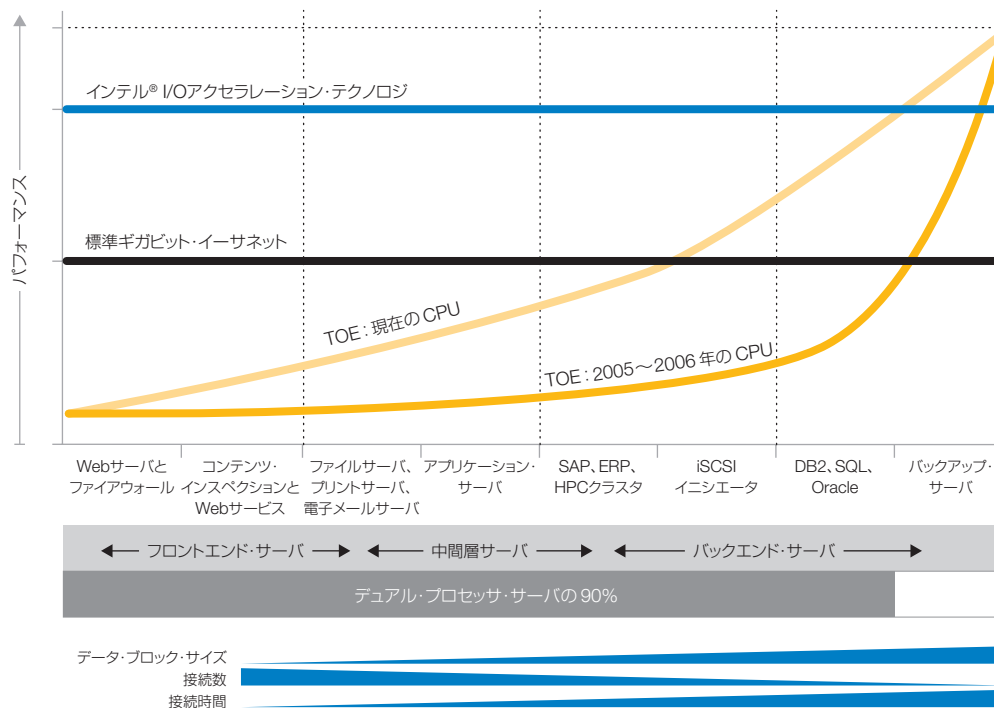
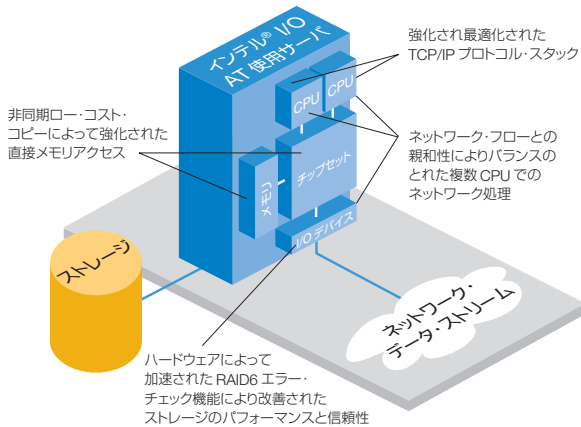


図3. 標準ギガビット・イーサネットに対する TOE およびインテル® I/OAT の相対パフォーマンス

TOE (TCP Offload Engine) ソリューションと比較しても、インテル® I/OAT はアプリケーションの I/O パケット・サイズ、接続数、または接続時間に関わらず、すべてのサーバ・アプリケーションで標準ギガビット・イーサネットより高いパフォーマンスを提供しています。

図 4. インテル® I/OAT のパフォーマンス強化



インテル® I/OAT では、すべての主要なオーバーヘッド・カテゴリに対してサーバ全体のパフォーマンスが強化され、より高速にアプリケーションとデータの受け渡しが可能になり、高い信頼性を提供しています。

対照的に、インテル® I/OAT では、サーバの I/O におけるボトルネック問題のセグメントをすべて解決していることに加えて、既存のアプリケーションや将来のアプリケーションを修正せずに使用できる TCP/IP を採用しています。インテル® I/OAT が提供するシステム全体のネットワーク I/O アクセラレーション・テクノロジーを図 4 で示します。このテクノロジーには、以下が含まれます。

ネットワーク・フローとの親和性—ネットワーク・スタック処理を複数の物理または論理 CPU に分割します。これにより、アプリケーションに CPU サイクルを割り当て、実行を高速化することができます。

非同期ロー・コスト・コピー—強化された直接メモリアクセスを提供します。はるかに少ない CPU サイクルでシステムメモリの NIC バッファからアプリケーション・バッファにペイロード・データをコピーでき、削減した CPU サイクルは可用なアプリケーション・ワークロードに戻されます。

最適化された TCP/IP スタックで改良された TCP/IP プロトコル—パケット・ペイロードからパケットヘッダの処理を最適化するために、独立したパケット・データとコントロール・パスを実現します。この強化とその他のスタック関連の強化により、プロトコルの処理サイクルが減少します。

インテル® I/OAT のネットワーク I/O アクセラレーション面の上記のテクノロジーを組み合わせれば、アプリケーションへのデータの供給が以前より 30% 高速化されます。

ストレージの I/O パフォーマンスの向上

インテル® I/OAT ソリューションの別のキー要素であるストレージ I/O アクセラレーションでは、ハードウェア・ベースのアクセラレーションを使用してアプリケーションとのデータ転送を高速化しています (図 4 を参照)。この高速化は、ストレージ・プロセッサを使用してこのワークロードを周辺機能として保持したり、RAID 6 (Redundant Array of Independent Disks) テクノロジーを追加して、RAID 5 に近いスループット・パフォーマンスを維持する一方でデータ転送時によりロバストなエラー・チェックを実行して実現しています。これにより、ディスクまたはディスク・ストレージ・サブシステムを通じたデータ移行時にデータが失われたり、内容が変わったりすることなく高速なデータ転送が行えます。

ストレージ・サブシステムを通してデータの保全性を保証するためにバイトのパリティチェックが使用されます。さらに、データのパリティチェックがディスクドライブに書き込まれれば、複数のディスクが故障したり再構築時にデータ・ブロック不良が発生したりした場合にデータが失われることを防ぎます。これにより、短いバックアップ画面、高速なディスク再構築時間、エンド・ツー・エンドなデータの保全性によりシステムの有用性と信頼性が向上します。

パフォーマンス、信頼性、効率性を高めるシステム全体のソリューション

TOE ソリューションと異なり、インテル® I/OAT は、サーバ全体のパケットおよびペイロードのすべての処理ボトルネックを解消するシステム全体のソリューションです。CPU の効率性を高め、現在のサーバで可能な速度より高速にアプリケーションとデータの受け渡しを行います。包括的なエラー・チェック機能によりストレージとのデータ転送に関連するリスクが緩和されます。最も重要なのは、インテル® I/OAT が将来のプラットフォームの改良に対応し、ハードウェアとソフトウェアの統合によりインフラストラクチャの費用をさらに削減し、ビジネスを成長させることです。

関連情報

インテル® I/O アクセラレーション・テクノロジーの詳細については、www.intel.com/go/ioat (英語) をご覧ください。

本資料に掲載されている情報は、インテル製品の概要説明を目的としたものです。本資料は、明示されているか否かにかかわらず、また禁反言によるとよらずにかかわらず、いかなる知的財産権のライセンスを許諾するためのものではありません。製品に付属の売買契約書 [Intel's Terms and conditions of Sales] に規定されている場合を除き、インテルはいかなる責を負うものではなく、またインテル製品の販売や使用に関する明示または黙示の保証 (特定目的への適合性、商品性に関する保証、第三者の特許権、著作権、その他、知的所有権を侵害していないことへの保証を含む) に関して一切責任を負わないものとします。インテル製品は、医療、救命、延命措置などの目的への使用を前提としたものではありません。インテル製品は、予告なく仕様変更される場合があります。



インテル株式会社

〒300-2635 茨城県つくば市東光台5-6
<http://www.intel.co.jp/>

インテル、Intel ロゴは、アメリカ合衆国およびその他の国における Intel Corporation またはその子会社の商標または登録商標です。
 * その他の社名、製品名などは、一般に各社の商標または登録商標です。

©2005 Intel Corporation. 無断での引用、転載を禁じます。
 2005年4月

306484-001JA
 JPN/0504/1K/SE/ICG/SK/HN