

IBM PureData System for Analytics アーキテクチャー

ハイパフォーマンスなデータウェアハウス
および高度な分析のためのアプライアンス・
プラットフォーム



目次

- 2 概要
- 2 シンプルなアプライアンスで最大のパフォーマンス
- 3 アーキテクチャー理念
- 2 システム構成要素
- 5 最大パフォーマンスの実現 - S-Blade
- 6 S-Blade の高速化 - IBM FAST エンジンの能力
- 7 クエリーのオーケストレーション
- 7 最適なクエリー・プランの作成
- 8 適切なタイミングで実行するためのスケジュール
- 10 必要とするすべての人々にすべての情報をオンデマンドで
- 11 無償体験で実際にご確認ください
- 11 IBM PureData System for Analytics について
- 11 IBM のデータウェアハウス・ソリューションおよび分析ソリューションについて
- 12 詳細情報

概要

いかなる企業であっても、成否の鍵は最適な情報を遅れずに入手し、適切な意思決定を行えるかどうかにあります。これらに適切に対処できない企業は、時間やリソースを犠牲にするだけでなく、チャンスを逃し、また組織を危険にさらしてしまうおそれさえあります。しかし、不安定な状況下において起こり得るイベントを予測する、トレンドを把握する、または最善の方法を選択するために、重要な情報から洞察を導き出すといった行為のためには、数十億件のデータ、もしくはペタバイト規模のデータを正確に分析する必要があります。このような戦略的な情報をより迅速に入手できる企業は、競合他社よりも早く反応し、優れた意思決定を行うことができます。

分析におけるイノベーションを継続することで、ビジネス全域に戦略的な情報上の利益をもたらすことができます。しかし、重要な情報を緊急に必要とする場合、その情報を提供するプラットフォームについて人々が考慮する時間はありません。このことから、プラットフォームは、照明スイッチのように簡単で確実、かつ素早いものでなければならず、無限に近いデータ・ボリュームやワークロードに対応できる必要があります。また、複雑さがユーザーに支障をきたすようなものであってはなりません。さらに、総所有コストを最小限に抑えながら、より多くのユーザーがより複雑なワークロードを実行でき、データ・ボリュームが絶え間なく増大し続けてもパフォーマンスを維持できるテクノロジー基盤を備え、長期間使用できるようなものである必要があります。

シンプルなアプライアンスで最大のパフォーマンス

Netezza テクノロジーを採用した IBM® PureData® System for Analytics データウェアハウス・アプライアンスは、シンプルなアプライアンスで業界をリードし続ける、最高の価格性能比を実現できるように開発されたプラットフォームで、従来のデータウェアハウスおよび高度な分析のあり方を大きく変革し続けています。このデータウェアハウス・アプライアンスは、一切の障害や制約がない高度な分析を行うための最先端のプラットフォームです。ユーザーやその組織にとって、それは、必要とするすべての人に最高の戦略的情報をあまねく提供する事を可能とし、あらゆる情報に対するニーズをこれまでの制約から解放する事になります。

PureData System for Analytics データウェアハウス・アプライアンス・ファミリーは、IBM が市場で最高の価格性能比を実現してきた理念に基づく、革命的な設計を採用しています。高速分析専用のアプライアンスとしての能力は、強力が高価なコンポーネントに由来するものではなく、適切なコンポーネントを統合し、これらを有効に活用することでパフォーマンスを最大化する方法に由来するものです。Massively Parallel Processing (MPP) と呼称される、マルチコア CPU と、IBM 固有の Netezza テクノロジーである FPGA Accelerated Streaming Technology (FAST™) エンジンとが結合することで、高価なシステムでも実現できない圧倒的なハイパフォーマンスを実現しています。さらに、使いやすいアプライアンスであるこのシステムは、インデックスやチューニングを一切必要とせず、導入後すぐに素晴らしい結果を実現します。シンプルなアプライアンスによる効果は、アプリケーション開発においてもその効果を発揮します。必要なイノベーションや開発を素早く実現し、多くのユーザーおよびプロセスにハイパフォーマンスな分析を提供することが可能になります。

本書では、IBM の Asymmetric Massively Parallel Processing™ (AMPP™) アーキテクチャーによりこのシステムがクエリーや高度な分析の複合的なワークロードをかつてないスピードで実現する方法について説明します。また、PureData System for Analytics データウェアハウス・アプライアンスのソフトウェアおよびハードウェアがバランスの取れた各コンポーネントを最大限に活用する方法や、数万ものユーザーによる大容量データの照会に対して最適化されたシステムが実際にどのように機能するかについて説明します。この製品は、他社には真似できない価格性能比を実現し、現在のさまざまなニーズや将来の課題に対応できる IBM 固有のデータウェアハウスおよび高度な分析専用のプラットフォームです。

アーキテクチャー理念

PureData System for Analytics は、分析処理に合わせて最適化され、柔軟に拡張できるように設計されたコンパクトなシステムに、データベース、プロセッサ、およびストレージが統合されています。このシステム・アーキテクチャーは、IBM が価格性能比で業界をリードしていることを証明してきた以下の中核となる理念に基づいています。

「データを移動させない」処理原理

Netezza テクノロジーを採用した PureData System for Analytics のアーキテクチャーは、「大規模データ・セットの処理では必要な場合以外データを移動させない」というコンピューター・サイエンスの基本原則に基づいています。

IBM アーキテクチャーではこの原理が適用されており、Field Programmable Gate Array (FPGA) と呼ばれる汎用コンポーネントを活用して、データ・ストリームのできるだけ早期、つまりデータがディスクから取り出されるとすぐ、不要なデータを除去し廃棄します。データ・ソース近くでこのデータ除去処理を行うことでデータ入出力のボトルネックを取り除き、CPU、メモリー、ネットワークといったダウンストリームのコンポーネントを不要なデータの処理から解放します。これにより、システム・パフォーマンスに大きな乗数効果をもたらすことが可能となります。

必要なデータのみ処理をする

PureData System for Analytics では、ゾーン・マップ・テクノロジーを使用しており、インデックス、パーティションといった複雑なチューニングは不要です。ゾーン・マップは、テーブル内の列の最小値および最大値をデータの保存区画毎に自動的に保管して、ユーザの照会に必要なデータを含むデータのみが読み取られるように制御しています。たいていの場合、クエリーが必要としているのは、テーブルの小さいサブセットのみです。ゾーン・マップにより、一切のチューニング作業、メンテナンスを行う事なく、ユーザは最適化された性能によるメリットを享受することができます。

バランスの取れた超並列アーキテクチャー

PureData System for Analytics のアーキテクチャーは、対称型マルチプロセッシング (SMP) および MPP の利点を併せ持ち、ペタバイト規模のデータを超高速に分析する専用のアプライアンスとして構成されています。プロセッサ、FPGA、メモリー、ネットワークを含む、アーキテクチャーの各コンポーネントは、コストや電力消費を最小限に抑えながら、ディスクの物理的特性が許容する範囲で可能な限り素早くデータを提供できるように徹底的に最適化されています。PureData System for Analytics のソフトウェアは、データ・ストリームが超並列処理を行えるようにこれらのコンポーネントが統合されており、各 MPP ノードが最大のスループットを発揮します。このバランスの取れたアーキテクチャーにより、1000 にも及ぶ処理ストリームを同時に実行するスケーラビリティを実現し、また段階的な拡張性を提供し、総所有コストを実現します。

高度分析のプラットフォーム

この MPP の原理および「データを移動させない」データ処理は、大規模なデータ・セットの高度分析にも同様に適用できます。PureData System for Analytics を使用することで、複雑な並列プログラミングやグリッド・プログラミングを考慮することなく、MPP ストリームの処理要素に、複雑な SQL だけでは実装できない アルゴリズムを容易に組み込むことができます。大容量データに対し「超並列処理上で」複雑な分析を実行できるため、遅延やコストをなくすことができます。また、パフォーマンスが大幅に改善されることにより、PureData System for Analytics アプライアンスは、データウェアハウスと高度な分析を統合するのに最適なプラットフォームとなっています。

シンプルなおプライアンス

PureData System for Analytics アプライアンスのアーキテクチャーにより、日常のオペレーションは簡素化され、ユーザーはプラットフォームの複雑さを意識せずに済みます。この製品ではシンプルさが設計上の最優先事項とされており、設計上のトレードオフが生じる場合、常にシンプルさが優先されて来ました。このソリューションは他のソリューションとは異なり、他の汎用システムで必要とされるチューニングを行わずに単に実行するだけで、厳しい照会ワークロード、複雑な分析との混合ワークロードを高速に処理します。また、導入、アップグレード、高可用性や事業継続性の確保といった一般的に時間のかかる作業であっても大幅に簡素化されるため、貴重な時間とリソースを大幅に節約できます。

加速化されるイノベーションとパフォーマンスの向上

PureData System for Analytics アーキテクチャーの主な目的の 1 つは、価格性能比の向上と革新的な機能を、競合他社のテクノロジーよりも早く、長期にわたって提供することです。オープンなブレード・ベースのコンポーネントの使用により、PureData データウェアハウス・アプライアンスは、テクノロジーの強化を迅速に取り込むことができると同時に、FPGA、バランスの取れたハードウェア構成、密結合インテリジェント・ソフトウェアの強力な影響を組み合わせることで、個別の要素によるパフォーマンスよりも大幅に優れた全体的なパフォーマンスの向上を実現できます。実際、PureData System for Analytics アプライアンス（およびそれ以前の Netezza アプライアンス）は、発表以来、2 年ごとにパフォーマンスが 4 倍以上向上しています（ムーアの法則の 2 倍¹）。これは、他の大手ベンダーを大きく上回っています。²

以下のページでは、IBM がいかにしてこれらの原理を実践するのかについて説明します。

システム構成要素

PureData System for Analytics の優れたパフォーマンスは主に、SMP フロントエンドとシェアード・ナッシング MPP バックエンドを結合して照会処理を行う Netezza テクノロジーを採用した、この製品固有の AMPP アーキテクチャーによって実現されます。

アーキテクチャーの各コンポーネントは、全体的にバランスのとれたシステムを提供するために、慎重に選別され、統合されています。各処理要素は、複数のデータ・ストリーム上で機能し、できる限り早く不要なデータを除去します。最大構成では、これらのカスタマイズされた MPP ストリームは 1000 にも及び、連動してワークロードを「超並列実行」します。

AMPP アーキテクチャー

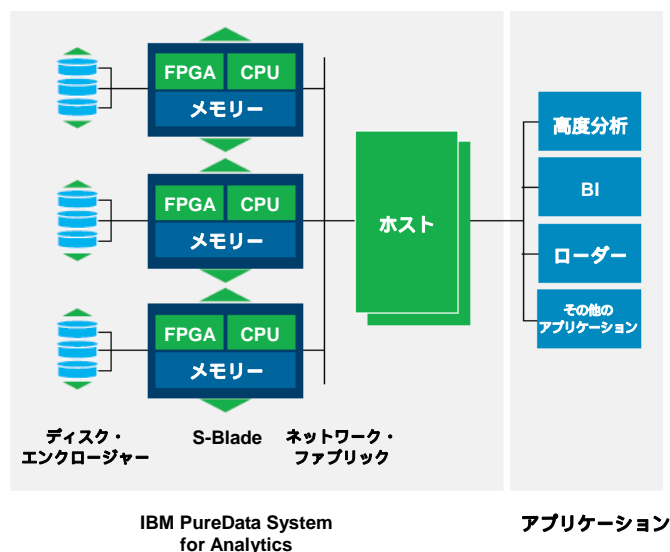


図 1: IBM PureData System for Analytics の優れたパフォーマンスは、この製品固有の AMPP アーキテクチャーによって実現されます

IBM PureData System for Analytics

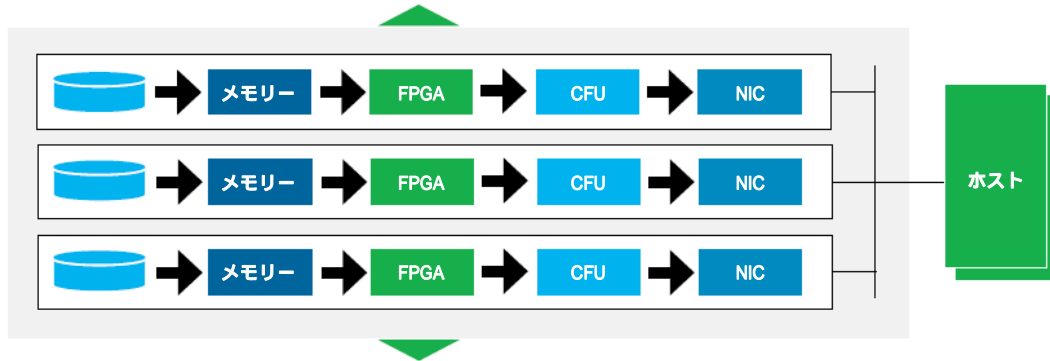


図 2: Asymmetric Massively Parallel Processing (AMPP) では、マルチコア CPU と IBM 固有の Netezza テクノロジーを採用した FPGA Accelerated Streaming Technology (FAST)TM エンジンが結合されています

ここで、アプライアンスの主要な構成要素について説明します。

- **IBM ホスト:** SMP ホストは、高可用性を実現するアクティブ/スタンバイ構成で構成された、Linux が稼働するハイパフォーマンスな IBM xSeries サーバーです。アクティブ・ホストは、外部のツールとアプリケーションに対する標準インターフェースを提供します。これは、SQL をスニペットと呼ばれる実行可能なコード・セグメントにコンパイルし、最適なクエリー・プランを作成して、実行のためにスニペットを MPP ノードに分配します。
- **スニペット・ブレード (S-Blades):** S-Blade は、アプライアンスの高速 MPP エンジン構成するインテリジェントな処理ノードです。各 S-Blade は、強力なマルチコア CPU、マルチエンジン FPGA、ギガバイト RAM を含む独立サーバーで、これらはすべてバランスが保たれ、同時に機能することで最大のパフォーマンスを実現します。CPU コアは、高度分析アプリケーションの大規模なデータ・ボリュームに対して複雑なアルゴリズムを実行できるように十分な余裕を持たせて設計されています。
- **ディスク・エンクロージャー:** ディスク・エンクロージャーには、RAID により保護された、ハイパフォーマンスな高密度の IBM ストレージ・ディスクが組み込まれています。各アクティブ・ディスクには、データベース・テーブルのデータの一部が含まれています。ディスク・エンクロージャーは、高速の相互接続によって S-Blade と接続されており、すべてのディスクが同時に最大速度でデータを S-Blade 上で処理する事が可能です。

- **ネットワーク・ファブリック:** すべてのシステム・コンポーネントは、高速のネットワーク・ファブリックにより接続されています。PureData System for Analytics は、ファブリックの帯域を十分に活用する、カスタマイズされた IP ベースのプロトコルを実行し、集中的なネットワーク・トラフィックが継続するなかでも輻輳状態を回避します。ネットワークは、1000 を超えるノードにまで拡張できるように最適化されています。また、各ノードはその他のすべてのノードに対して大規模なデータ転送を同時に開始することが可能です。

注: すべてのシステム・コンポーネントが冗長構成されています。ホストはアクティブ/パッシブですが、アプライアンス内のその他すべての保守可能コンポーネントはホット・スワップ可能です。ユーザー・データは完全にミラーリングされ、可用性の高いソリューションを実現します。

最大パフォーマンスの実現 - S-Blade

S-Blade (スニペット・ブレード: Snippet Blade) は、コモディティ・コンポーネントと PureData System for Analytics データウェアハウス・アプライアンス・ソフトウェアで構成されており、これらを結合することで、各 MPP ノードの最大のスループットを引き出すことができます。

ストレージ・アレイの専用の高速相互接続により、ディスクから送信されるとすぐにデータをメモリーに格納することができます。圧縮されたデータは、スマート・アルゴリズムを使用してメモリーにキャッシュされます。これにより、ディスクにアクセスしなくても、最も頻繁にアクセスするデータはメモリーからすぐに取り出されます。

FPGA 内で並列で実行されている FAST エンジンは、ディスクが実現可能な最大限のスピードでテーブル・データを解凍し、その 95% から 98% を除去します。その際、クエリーへの応答に関するデータのみを残し、Netezza Zone Map テクノロジーを使用して、関連するページまたは範囲のみを抽出します。ストリーム内に残ったデータは、並列で実行されている CPU コアによって並列処理されます。このプロセスは、PureData System for Analytics データウェアハウス・アプライアンスで実行されている、1000 にも及ぶこれらの並列スニペット・プロセッサで繰り返されます。その結果、高価なシステムをはるかに凌ぐ圧倒的なパフォーマンスが実現されます。

S-Blades の高速化: IBM FAST エンジンの能力

FPGA は、Netezza テクノロジーを採用した IBM データウェアハウス・アプライアンスの優れた価格性能比を実現する重要なイネーブラーです。各 FPGA には、データ・ストリームでフィルタリング機能と変換機能を実行する組み込みエンジンが含まれています。これらの FAST エンジンは、動的に再構成することができ、ソフトウェアを介して変更や拡張が行えます。これらのエンジンは、クエリー実行中に指定されているパラメーターによって各スニペットごとにカスタマイズされ、直接メモリー・アクセス (DMA) モジュールによって超高速で送信されるデータ・ストリームに働きかけます。

FAST エンジンには以下のものが含まれます。

- システム・パフォーマンスを 4 倍から 8 倍向上させる Netezza のイノベーション機能を採用した圧縮エンジン、PureData System for Analytics。³ このエンジンは、転送スピードでデータを解凍し、ディスク上の各ブロックを即座にメモリー内に展開します。その結果、ディスク (データウェアハウスで最も遅いコンポーネント) の大幅なスピードアップが実現します。
- プロジェクト・エンジンおよび制限エンジン。SQL クエリーの SELECT 節と WHERE 節のパラメーターに基づいて列や行をそれぞれ除去することにより、パフォーマンスをさらに向上させます。
- 可視性エンジン。ストリーミング・スピードで ACID (原子性、一貫性、独立性、および耐久性) に準拠するうえで重要な役割を果たします。このエンジンは、照会側に見えない方が望ましい行 (例: まだコミットされていないトランザクションに属する行) を除去します。

また、IBM FAST エンジンは、PureData System for Analytics のソフトウェアの拡張機能を介して、将来追加される革新的な新機能のための拡張可能なフレームワークを提供します。これらの新機能は、システム・パフォーマンス、セキュリティ、および信頼性をさらに向上させます。

IBM PureData System for Analytics

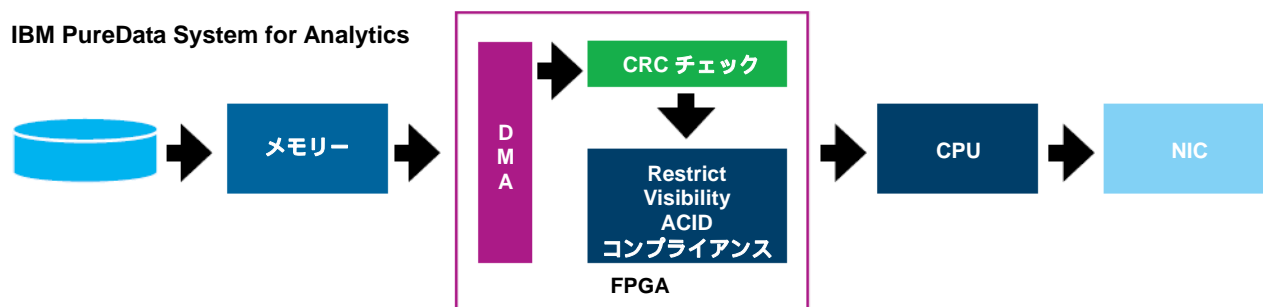


図 3: IBM PureData System for Analytics の FAST エンジンは、動的に再プログラミングできます。

クエリーの実行

PureData System for Analytics のハードウェア・コンポーネントとインテリジェントなシステム・ソフトウェアは、密接に関連し合っています。このソフトウェアは、アプライアンスのハードウェア機能を十分に活用し、多数のイノベーション機能を統合して、単純な検索か、複雑な非定型クエリーか、詳細分析かを問わず優れたパフォーマンスを実現できるように設計されています。このセクションでは、システムに組み込まれているインテリジェンスについて手順を追って説明します。

ソフトウェア・アーキテクチャー

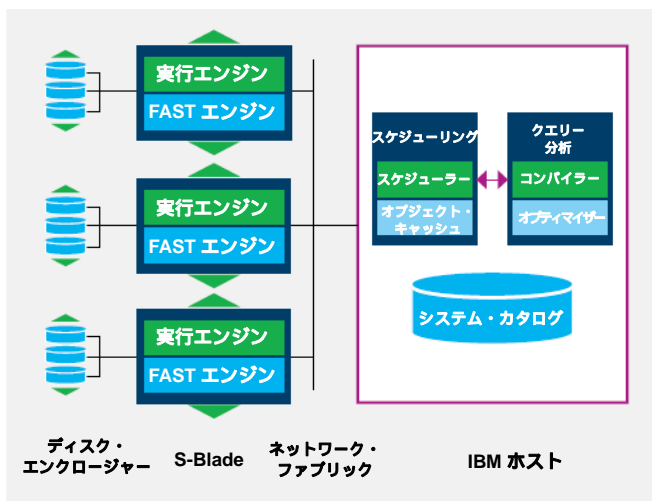


図 4: IBM PureData System for Analytics のソフトウェアは、アプライアンスのハードウェア機能を十分に活用できるように設計されています

PureData System for Analytics のソフトウェア・コンポーネントには以下のものが含まれます。

- 高度な並列最適化プログラム。より効率的に実行できるようにクエリーを変換し、処理ノードごとに各コンポーネントが最大限に活用されるようにします。
- インテリジェントなスケジューラー。ワークロードを問わず、稼働システムのスループットを最大に保ちます。
- 高速のスニペット・プロセッサ。複数のクエリーおよび複雑な分析機能を効率的に同時実行します。
- スマート・ネットワーク。Netezza テクノロジーを採用した IBM データウェアハウス・アプライアンスを介して大量のデータを簡単に移動できるようにします。

最適なクエリー・プランの作成

ユーザーがクエリーをサブミットすると、ホストはそれをコンパイルし、PureData System for Analytics の AMPP アーキテクチャーに最適なクエリー実行プランを作成します。IBM によるオプティマイザのインテリジェンスは、このシステム最大の長所の 1 つです。このオプティマイザは、システムの MPP ノードをすべて活用し、クエリーで参照される各データベース・テーブルの最新の詳細データを収集します。これらのメトリックの大半は、クエリー実行中に低オーバーヘッドで収集され、クエリーごとに個別設定されたジャストインタイムの統計データを提供します。相互に通信可能な統合されたコンポーネントと、PureData System for Analytics データウェアハウス・アプライアンスの特性により、コスト・ベースの最適化が可能になります。これにより、オペレーションに関連するディスク・コスト、処理コスト、およびネットワーク・コストをより正確に測定できます。

また正確なデータを活用することで、オプティマイザはすべてのコンポーネントを最も効率的に利用するクエリー・プランを作成できます。

オプティマイザのインテリジェンス: ジョイン順序の計算

オプティマイザのインテリジェンスの一例として、複雑なジョインで最適な順序を判別する機能が挙げられます。例えば、複数の小さなテーブルを大きなファクト・テーブルとジョインする場合、オプティマイザは、小さなテーブル全体を各 S-Blade にブロードキャストする一方で、大きなテーブルがすべてのスニペット・プロセッサ全体に配置されるようにします。この方法により、PureData System for Analytics ソフトウェア・コンポーネントを活用しながら (並列してジョインできるという AMPP アーキテクチャーの利点を活用しながら)、データの移動を最小限に抑えることができます。

オプティマイザは、これらの統計に基づき検索開始前にクエリーを変換し、データウェアハウス・システムのパフォーマンスを低下させる原因となるディスク入出力やデータ移動を最小限に抑えます。オプティマイザの判断には以下のものが含まれます。

- 適切なジョイン順序の判別
- 式の再作成
- 重複する SQL 操作の除去

スニペットへの変換

コンパイラーは、クエリー・プランをスニペットと呼ばれる実行可能なコード・セグメントに変換します。スニペットは、アプライアンスのすべてのデータ・ストリームにわたりスニペット・プロセッサが並列して実行するクエリー・セグメントです。各スニペットは 2 つの要素を備えています。1 つは個別の CPU コアによって実行されるコンパイル済みコードで、もう 1 つは、特定のスニペットのフィルタリングや処理を行う FAST エンジンのカスタマイズする一連の FPGA パラメーターです。このスニペット毎のカスタマイズにより、PureData System for Analytics では事実上、個別のクエリーに合わせ、リアルタイムにハードウェア構成が最適化されます。

コンパイラーのインテリジェンス: オブジェクト・キャッシュ

ホストは、オブジェクト・キャッシュと呼ばれる機能を使用して、クエリーのパフォーマンスをさらに加速化します。これは、パラメーターのさまざまなバリエーションをサポートする、コンパイル済みスニペット・コードの大規模キャッシュです。例えば、「where name='Bob'」という節を含むスニペットは、「where name='Jim'」という節を含むスニペットと同じコンパイル済みコードを使用しますが、設定には別の名前が反映されます。この方法により、99% を超えるスニペットのコンパイル手順をなくすることができます。

適切なタイミングで実行するためのスケジューラ

スケジューラーは、複雑なワークロード間で実行のバランスを取ることで、最大の使用効率とスループットを維持しながら、さまざまなユーザーが目的を果たせるようにします。IBM スケジューラーは、S-Blade でスニペットを実行するタイミングを決定する際に、クエリーの優先順位、サイズ、リソースの使用可能性などの要素を考慮します。アプライアンスのアーキテクチャーにより、スケジューラーはシステムの各コンポーネントからリソースの使用可能性に関する、より最新の正確なメトリックを収集できます。このスケジューラーは、ディスク処理能力のほぼ 100% を活用し、メモリーやネットワークのリソースに負荷をかけ過ぎたり、システムを酷使して効率性を低下させたりしないようにすることで、システムのスループットを最大化する高度なアルゴリズムを使用しています。これは IBM アーキテクチャーの重要な特性です。これにより、負荷が非常に高い状況下であっても、システムは最大のスループットで稼働し続けることが可能になります。

スケジューラーのインテリジェンス: リソースの過負荷の排除

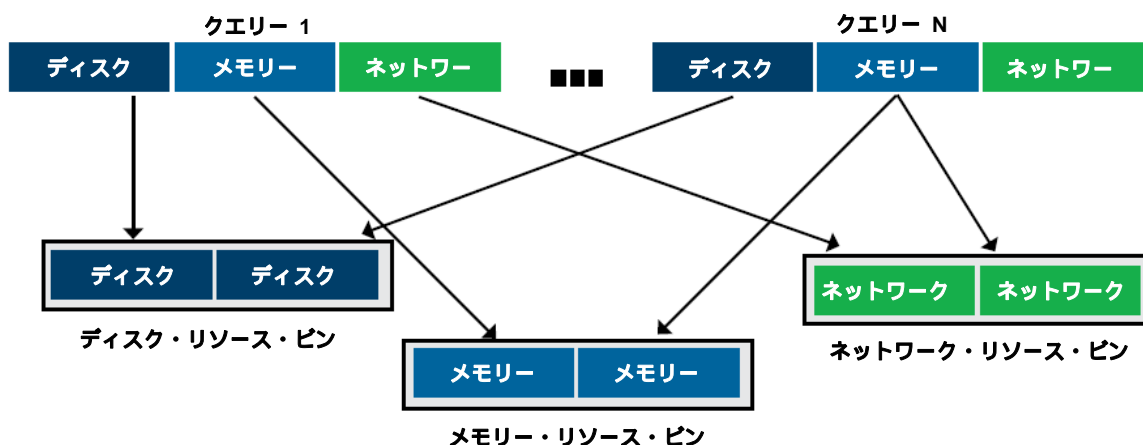


図 5: スケジューラーの指示でインテリジェント・ネットワーク・ファブリックを介してスニペットがすべてのスニペット・プロセッサにブロードキャストされます。

超並列での実行

各 S-Blade の各スニペット・プロセッサは、それぞれの実行すべき処理を受け取ります。

スニペット・プロセッサは、ホスト・スケジューラーに加え、独自のスマート・プリエンティブ (割り込み型) スケジューラーを備えています。これにより、スニペットで複数のクエリーを同時に実行できます。このスケジューラーは、クエリーの優先順位や、クエリーを実行したユーザーまたはグループ用に確保されているリソースを考慮して、特定のスニペットをいつ、どの程度の時間実行するのかについてスケジュールを決定します。実行は下記流れで行われます。

- 各スニペット・プロセッサのプロセッサ・コアは、クエリーのスニペットに含まれているパラメータによって FAST エンジン構成し、データ・ストリームをセットアップします。
- スニペット・プロセッサは、ZoneMap™ アクセラレーションと呼ばれる PureData System for Analytics のイノベーション機能を活用して、ディスク・アレイのテーブル・データをメモリーに読み込みます。この機能では、関連するページまたは範囲のみを読み取ることにより、ディスク・スキャンを削減します。また、このスニペット・プロセッサは、データ・ブロックを検索する場合にディスクにアクセスする前にキャッシュを探します。これにより、データが既にメモリー内にある場合にスキャンしなくて済むようになります。
- FPGA がデータ・ストリームに作用します。まず、データ・ストリームを転送スピードで解凍することにより、データ・ストリームの速度を最大 4 倍以上にします。⁴
- その後、FPGA FAST エンジンが、クエリーに関連しないすべてのデータを除去します。残ったデータは、メモリーへと流され、CPU コアによって並列処理されます。このデータは、通常、元のストリームと比較して大幅に小さくなっているため (平均して 2% から 5%)、プロセッサ・コアが必要とする実行時間を大幅に短縮できます。
- プロセッサ・コアは、データ・ストリームを受け取り、ソート、ジョイン、集計などの主要なデータベース操作を実行します。また、スニペット・プロセッサに組み込まれている複雑なアルゴリズムを活用して高度分析処理を行います。
- 各スニペット・プロセッサの結果がメモリー内にまとめられ、スニペット全体に関する中間結果が作成されます。このプロセスは、1000 にも及ぶスニペット・プロセッサで同時に繰り返されます。このとき、数百または数千のクエリーのスニペットが並列実行されます。

ZoneMap アクセラレーション: IBM アンチ・インデックス

ZoneMap アクセラレーションは、データウェアハウスで行の自然順序付けを活用することで、桁違いにパフォーマンスを加速化させます。この手法により、クエリーの開始と終了の範囲外にある列の値で行をスキャンせずに済みます。例えば、テーブルに週ごとのレコードが 2 年分 (~100 週間分) 含まれており、照会で 1 週目のみのデータを検索する場合、ZoneMap アクセラレーションを使用することでパフォーマンスが最大 100 倍向上します。インデックスとは異なり、ZoneMap は、管理のオーバーヘッドを一切伴うことなく、各データベース・テーブルごとに自動的に作成および更新されます。

結果の応答

現在、すべてのスニペット・プロセッサがスニペットの結果を持っている状態で、これらの結果をまとめる必要があります。スニペット・プロセッサは、インテリジェントなネットワーク・ファブリックを活用して柔軟にホストや他のスニペット・プロセッサと通信を行い、中間の計算および集計を実行します。

ホストは、スニペット・プロセッサから受け取った中間結果をまとめ、最終的な結果セットをコンパイルして、ユーザーのアプリケーションに返します。その間にも、他の複数のクエリーがシステムを流れており、完了までのさまざまな工程をえています。

ネットワークのインテリジェンス: 予測可能なパフォーマンスとスケーラビリティ

IBM のカスタム・ネットワーク・プロトコルは、大容量のデータウェアハウスに関連するデータ・ボリュームおよびトラフィック・パターン専用に設計されています。IBM プロトコルは、ネットワークに負荷をかけ過ぎることなく、ネットワークの処理能力を最大限に活用し、ほぼ一定に維持される予測可能なパフォーマンスを実現します。

トラフィックは、以下の 3 つの異なる領域でスムーズに流れます。

- ブロードキャスト・モードでホストからスニペット・プロセッサへ (1 対多)
- システム・ラック・レベルでスニペット・プロセッサからホストへ (多対 1)、S-Blade に集約
- スニペット・プロセッサからスニペット・プロセッサへ (多対多)、中間処理のために大量のデータが自由に流れる

必要とするすべての人々にすべての情報をオンデマンドで

最も規模が大きく高価なソリューションが最高のソリューションであるとは限りません。規模が小さく安価であっても最高のスマートさを備えた設計のソリューションは存在します。IBM PureData System for Analytics は、ストリーミング処理がその他の分析システムやデータウェアハウス・システムで使用される従来のコンピューティング・アーキテクチャーに勝る先天的な利点を提供することを早くから認識していました。その結果が、大規模なシステムを小さな存在へと変えてしまうパフォーマンスを備えたコンパクトなアプライアンスです。超高速で巨大なデータ・ボリュームや、数千の同時ユーザーのさまざまなワークロードに対して複雑なアルゴリズムを実行します。

処理のパフォーマンスは、お客様の成功を支援するための他社とは異なるプラットフォームを実現する他の機能によって補完されます。

- **使いやすさ:** PureData System for Analytics データウェアハウス・アプライアンスは、アプライアンスの名が示すとおり簡易な管理が可能であり、常に最大のスループットで稼働します。人によるメンテナンスの必要性を最小限に抑え、システム・ソフトウェアにより最大のスループットを確保します。
- **企業全体での優れた意思決定:** 組み込み機能により、データベースに新世代の分析機能を組み込み、開発の労力を最小限に抑えます。個別のサーバー・ハードウェアを用意する必要はなく、大量のデータ転送に時間がかかることもありません。組織のあらゆる部門で、必要とするすべての人々に一瞬で結果を返し、重要なビジネス・インテリジェンスを提供できます。
- **将来に備えた俊敏性:** このシステムは、現在の課題のみでなく将来の課題にも対応できるように構築されています。ペタバイト規模のユーザー・データに合わせた段階的な拡張や、ムーアの法則で制御された従来の進化速度をはるかに凌ぐパフォーマンスの向上を実現します。

IBM PureData System for Analytics データウェアハウス・アプライアンスを使用することで、ユーザーとその組織は、パフォーマンスの向上はもちろんのこと、極めて明確な意思決定が可能となります。しかし、本書の説明だけで納得はしないでください。PureData System for Analytics データウェアハウス・アプライアンスを評価する最も良い方法は、実際にこの製品の稼働を確認していただくことです。お客様のデータを IBM Netezza データウェアハウス・アプライアンス以上に活用できる製品は他に存在しないということがわかりいただけると幸いです。

無償体験で実際にご確認ください

お客様は、IBM PureExperience プログラム (実機検証) を通じて、PureData System for Analytics をお試しいただけます。このプログラムは無償で、お客様独自のデータを使用してシステムをテスト稼働していただけます。このプログラムでは、お客様の要件に応じた内容が柔軟にカスタマイズされ提供されます。このプログラムの詳細情報と、最寄りの地域で利用可能なプログラムの内容については、IBM 営業担当員にお問い合わせください。

IBM PureData System for Analytics について

Netezza テクノロジーを採用した IBM PureData System for Analytics は、データベース、サーバー、およびストレージを、管理が容易な単一のアプライアンスに統合したシステムです。セットアップと継続的な管理を最小限に抑え、より迅速で一貫性のある分析パフォーマンスを実現します。IBM PureData System for Analytics は、データが存在するアプライアンスにすべての分析アクティビティを統合することにより、ビジネス分析を大幅に簡素化し、業界トップ・レベルのパフォーマンスを実現します。
ibm.com/software/jp/data/puredata/analytics/ をぜひご覧ください。IBM エキスパート統合システム・ファミリーがどのようにしてあらゆる段階の複雑さを解消し、組織の真のビジネス・バリューを促進するかについて説明しています。



IBM のデータウェアハウス・ソリューション および分析ソリューションについて

IBM は、データウェアハウス、情報管理、およびビジネス分析のソフトウェア、ハードウェア、およびソリューションから成る極めて広範で包括的なポートフォリオを提供することで、お客様の情報資産の価値を最大限に高め、より素早く優れた意思決定を行うための新しい洞察を発見し、最適なビジネス結果を実現する支援をします。

詳細情報

IT をビジネス戦略の中心とし、実績のある知見を活用してリーダーシップを発揮できるようにします。PureSystems ファミリーおよび PureData System for Analytics の詳細については、日本 IBM の営業担当員または IBM ビジネス・パートナーにお問い合わせいただくか、次の Web サイトをご覧ください。

ibm.com/software/jp/data/puredata/analytics/

さらに、IBM グローバル・ファイナンスは、お客様がビジネス上必要とされているソフトウェア機能を、可能な限りコスト効果の高い、戦略的な方法で導入いただくための支援を提供します。IBM は、お客様（事前に信用審査の承認が必要）とともに、ビジネス目標や開発目標に合わせてファイナンス・ソリューションをカスタマイズし、効果的なキャッシュ管理を実現し、総所有コスト (TCO) を改善します。IBM グローバル・ファイナンスを通して、重要な IT 投資に対する資金を調達し、ビジネスのスピードアップを図ってください。詳しくは、ibm.com/financing をご覧ください。

© Copyright IBM Corporation 2013

日本アイ・ピー・エム株式会社

〒103-8510
東京都中央区日本橋箱崎町 19-21

Produced in Japan
June 2013

IBM、IBM ロゴ、ibm.com および Netezza は、世界の多くの国で登録された International Business Machines Corporation の商標です。他の製品名およびサービス名等は、それぞれ IBM または各社の商標である場合があります。現時点での IBM の商標リストについては、www.ibm.com/legal/copytrade.shtml をご覧ください。

Microsoft および SQL Server は、Microsoft Corporation の米国およびその他の国における商標です。

- 1 Intel 設立者の 1 人であるゴードン・ムーア氏は、1965 年、チップのトランジスタ数がおよそ 2 年ごとに倍増することを予見していました。主にこのプロセッサの改善により、ソフトウェア・アプリケーションのパフォーマンスが時間とともに加速化していきました。
- 2 社内の IBM Netezza and PureData System for Analytics パフォーマンス・ベンチマーク、および Netezza ならびに IBM のお客様から得られた結果に基づいています。
- 3 アプライアンスの FPGA に関する社内の IBM パフォーマンス・ベンチマークに基づいています。
- 4 4 倍の平均圧縮率に基づきます。個別の結果は異なる場合があります。



Please Recycle