

データ分析速度を最大 100 倍に高速化する処理技術を開発 セルフサービス分析を実現するリアルタイムデータ分析システムのプロトタイプを構築

株式会社日立製作所(執行役社長兼 CEO:東原 敏昭/以下、日立)は、このたび、ハードウェア(FPGA)*1 の内部メモリを活用してデータを高速に処理するためのデータベース管理システムと、FPGA 内でデータを高速に並列処理する技術を開発しました。開発した技術を使用しない場合と比較して、データ分析速度を最大 100 倍に高速化することに成功しました。また、開発した 2 つの技術を、日立のグループ会社である Pentaho Corporation(ペンタホ社)が開発した、分析結果を出力するビジネス分析ソフトウェア*2Pentaho Business Analytics、およびデータを格納するフラッシュストレージ*3 と組み合わせることで、リアルタイムデータ分析システムのプロトタイプを構築しました。今回開発したプロトタイプを用いることで、データ分析の専門家に依頼することなく、大量の業務データをその場で分析することができるセルフサービス分析の高速化を実現し、業務効率の向上に貢献していきます。

近年、データサイエンティストなど専門家が行っていた大量のデータを分析する業務を自分自身で行うセルフサービス分析が注目されています。具体的には、銀行の窓口担当者が来店した顧客から聞き取った情報を分析システムにその場で入力し、分析結果を踏まえて顧客のニーズにマッチした商品を提示するといった事例があげられます。

このようなセルフサービス分析を実現するためには、大量のデータを迅速に処理する高い処理性能が必要です。セルフサービス分析においてシステムが実行する処理には、データ読み出し処理と、データ分析処理があり、データ読み出し処理の性能は、データを格納する記憶媒体をハードディスクドライブに代えてフラッシュストレージを用いることで、既に数 10 倍から数 100 倍に向上しています。一方で、データ分析処理がデータの読み出し処理の速度に追いつかず、その性能向上がセルフサービス分析の課題になっていました。

そこで日立は、FPGA の内部メモリを活用してデータを高速に処理するためのデータベース管理システムと、FPGA 内でデータを高速に並列処理する技術を開発し、データ分析速度を最大 100 倍に高速化することに成功しました。また、今回開発した 2 つの技術を、分析結果を出力するビジネス分析ソフトウェア Pentaho Business Analytics、および、データを格納するフラッシュストレージと組み合わせることで、リアルタイムデータ分析システムのプロトタイプを構築しました(図 1)。

開発した 2 つの技術の概要は以下の通りです。

1. FPGA の内部メモリを活用してデータを高速に処理するためのデータベース管理システム

FPGA は高速な内部メモリ(数 MB)を備えています。データ分析処理に適したカラム型データベース*4 で使用されるデータ格納形式では、データの圧縮ルールや格納位置などを示すデータ管

理情報のサイズが FPGA の内部メモリより大きいため、低速な外部メモリに格納する必要がありました。データ管理情報は、カラム型データを外部メモリに読み込む際に参照が必要となりますが、低速な外部メモリに格納すると処理速度が低下してしまう原因となります。そこで、FPGA の内部メモリのサイズにデータ管理情報が収まるように、データベースを細分化してフラッシュストレージ内に保持し、データ管理情報を FPGA の内部メモリで処理するデータベース管理システムを開発しました(図 2) (国際特許出願済み*5)。

2. FPGA でデータを高速に並列処理する技術

高速にデータ処理を行う場合、データを並列化して処理する方法が一般的に用いられています。しかし、カラム型データベースで処理する場合、一つの項目の処理が完了するまで次の処理に進むことができず、並列化が困難でした。そこで、一定の個数ずつ順番に処理することができるカラム処理方式を開発するとともに、データを選択するデータフィルタ回路、および、データをグループ分けして合計値や平均値などを算出するデータ集約演算回路を FPGA に実装することで、並列化処理を実現しました。

今後、超高速データベースエンジンを搭載した、日立の高速データアクセス基盤「Hitachi Advanced Data Binder プラットフォーム」*6に本技術を導入し、データ分析処理の一層の高速化をめざすとともに、ビジネスの現場においてリアルタイムデータ分析システムの検証を行い、業務効率向上の実現に貢献します。

本成果は、2016年8月9日(火)～11日(木)に米国・サンタクララで開催される Flash Memory Summit 2016 において、技術発表、並びにデモ展示を行う予定です。

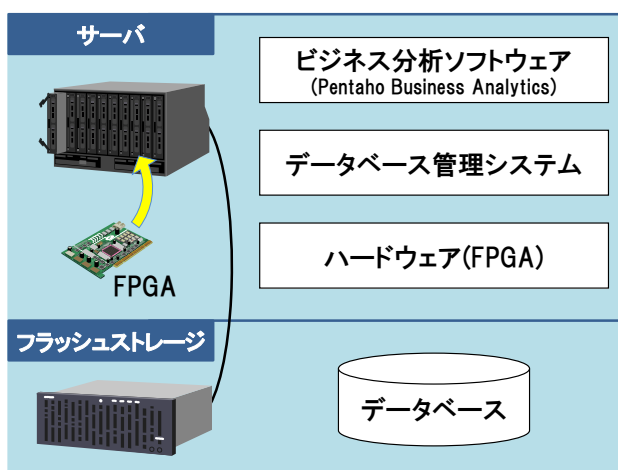


図1 リアルタイムデータ分析システムの構成図

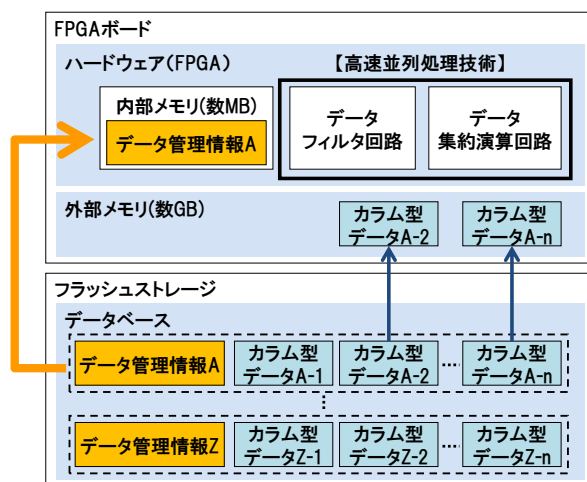


図2 開発技術の詳細説明図

- *1 ハードウェア(FPGA)(Field Programmable Gate Array) : 製造後に購入者や設計者が構成を設定できる集積回路。特定用途向けの集積回路と比較して、低コストで開発が可能。
- *2 ビジネス分析ソフトウェア: データ分析方法の指定やデータ分析結果の閲覧を行うためのソフトウェア。
- *3 フラッシュストレージ: データ記憶媒体としてフラッシュメモリを用いたストレージ装置。
- *4 カラム型データベース: 取引額などデータを構成する項目ごとの処理を効率的に実行するように設計されたデータベース。
- *5 関連特許出願 13 件。
- *6 内閣府の最先端研究開発支援プログラム「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的サービスの実証・評価」(中心研究者: 喜連川 東大教授/国立情報学研究所所長)の成果を利用。

■照会先

株式会社日立製作所研究開発グループ 技術統括センタ 研究管理部[担当: 井川、鷹栖(たかのす)]
〒244-0817 神奈川県横浜市戸塚区吉田町 292 番地
電話: 050-3135-3409 (直通)

以上

このニュースリリース記載の情報(製品価格、製品仕様、サービスの内容、発売日、お問い合わせ先、URL 等)は、発表日現在の情報です。予告なしに変更され、検索日と情報が異なる可能性もありますので、あらかじめご了承ください。
