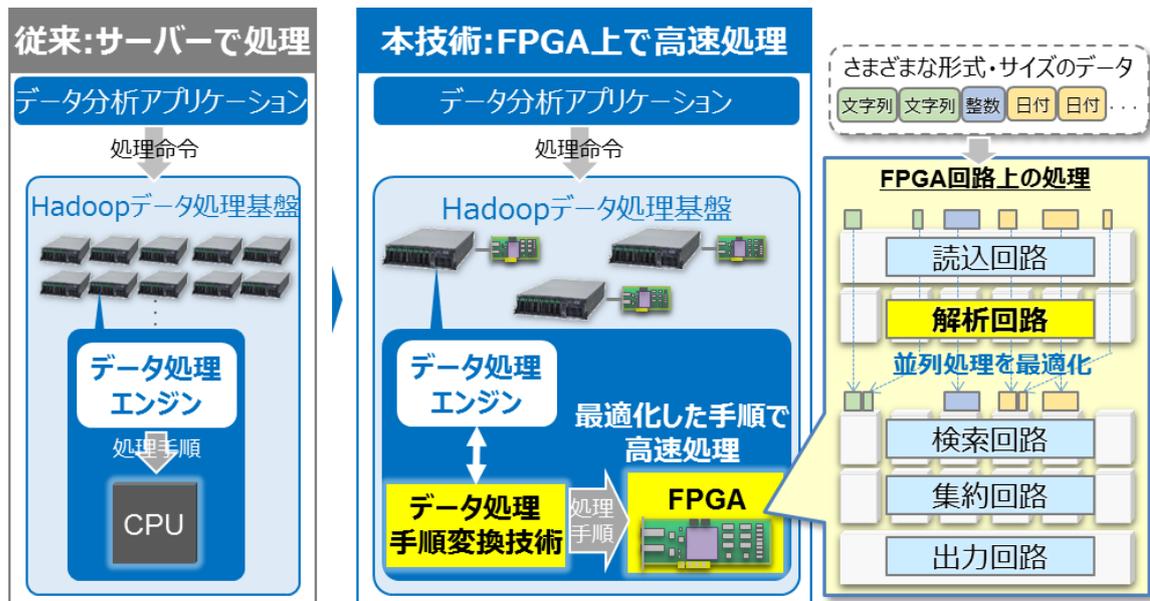


News Release

2017年11月14日
株式会社日立製作所

OSS ベースでのビッグデータ分析を最大 100 倍に高速化する技術を開発 高速分析システム構築の低コスト化を実現



本技術概要

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、オープンソースソフトウェア(OSS)の Hadoop データ処理基盤^{*1}で行うビッグデータ分析を、最大 100 倍に高速化する技術を開発しました。本技術は、本来はソフトウェア処理に合わせて作られているデータ処理手順を、ハードウェア上の並列処理に適した手順に変換した上で、さまざまな形式のデータを FPGA^{*2}上で高速に処理することを可能にします。これにより、高速分析を行う場合でも多数のサーバーが不要となるため、データサイエンティストが行うインタラクティブ分析や、ビジネス現場での迅速な意思決定、タイムリーな情報提供サービスなどを、低コストで実現します。日立は、今後、本技術を金融や通信などの分野に適用し、実証実験を進めることで、データアナリティクスサービスの基盤を支えていきます。

近年、IoT に用いられるセンサーをはじめ、金融機関の口座取引記録や、ソーシャルメディアの情報などの大量かつ多様なデータを、さまざまな視点や条件でインタラクティブに分析して業務やサービスに反映させる、ビッグデータ分析の重要性が高まっています。分析には、オープンソースの Hadoop データ処理基盤が広く使用されていますが、高速化のためには多数のサーバーが必要であり、機器や管理のコスト増大が課題となっています。

日立は、2016年にFPGAを用いた高速データ処理技術を開発^{*3}しています。しかし、この技術は、Hadoop データ処理基盤とはデータ管理方式の異なる、自社のデータベースを対象とした技術であり、データベース管理ソフトウェア自体にも個別のカスタマイズを行っているため、Hadoop データ処理基盤への適用は困難でした。

そこで、今回日立は、Hadoop データ処理基盤において FPGA を用いた高速データ処理を実現する技術を開発しました*4。開発した技術の特長は以下の通りです。

1. FPGA 向けに処理効率を最適化する、データ処理手順変換技術

Hadoop データ処理基盤のデータ処理エンジンは、CPU を用いてソフトウェアで順次処理を実行することを前提に、データの取得・検索・演算といったデータ処理の手順を最適化しています。しかし、この処理手順をそのまま実行しても、ハードウェアが得意とする並列処理による高速化の効果を十分に生かすことができません。そこで、Hadoop の処理手順を解析し、分散処理効率を考慮した上で、FPGA での並列処理に適した順序に処理命令を変換する、データ処理手順変換技術を開発しました。これにより、FPGA の回路を無駄なく効率的に利用可能となります。

2. さまざまなデータの形式を解析して FPGA で高速に処理する回路設計

従来は、FPGA 上で処理を行う場合は、日付や数値、文字などのデータ形式を、ハードウェアで処理しやすい形式に限定し、データの種類ごとに専用の処理回路を用意する必要がありました。しかし、Hadoop データ処理基盤では、例えば同じ日付のデータでも UNIX 通日型、ユリウス通日型など複数の形式を許容するため、多くの専用処理回路が必要となり、搭載規模に限界のある FPGA の回路領域を有効に活用できません。そこで、さまざまなデータ*5の形式やサイズを判別する解析回路を設け、データ形式やサイズに応じて複数のデータを一つの回路でまとめて処理するなど、並列処理の最適化調整を行う回路設計を考案しました。これにより、さまざまなデータに対応できるだけでなく、検索処理や集約処理の回路を無駄なく利用した並列処理が可能となり、高速データ処理を実現します。

本技術を Hadoop データ処理基盤に適用し、サンプルデータの分析を行った結果、データ処理性能を最大 100 倍に高速化できることを確認しました。これにより、高速分析に必要なサーバー数を大幅に削減することができるため、現在広く行われている Hadoop ベースのビッグデータ分析の低コスト化が可能です。日立は、今後、お客さまとの実証実験を通じて、本技術の実用化をめざします。

なお、本成果は、2017 年 11 月 13 日(月)～16 日(木)に米国・デンバーで開催される「SC17(The International Conference for High Performance Computing, Networking, Storage and Analysis)」において、技術展示を行う予定です。

*1 Hadoop データ処理基盤：大規模データの配置や分析を、複数のサーバーに分散させて処理するオープンソースのソフトウェアフレームワーク「Hadoop」を用いた計算基盤。サーバーの単純な増加で処理性能を向上させることが可能。

*2 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ハードウェア。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。

*3 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

<http://www.hitachi.co.jp/New/cnews/month/2016/08/0803a.html>

*4 関連国際特許出願：10 件。

*5 Hadoop など、オープンソースデータ処理基盤で一般的に用いられる標準データフォーマット「Parquet」をサポート。

■照会先

株式会社日立製作所 研究開発グループ 研究管理部 [担当:小平、安井]
〒185-8601 東京都国分寺市東恋ヶ窪一丁目 280 番地
電話:042-323-1111(代表)

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。
