様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２９　年　　６　月　　７　日現在

機関番号：　１２６１２
研究種目：　基盤研究(C)（一般）
研究期間：　2014～2016
課題番号：　２６３３００６１

研究課題名（和文）専用ハードウェアを用いたデータストリーム管理システムの開発

研究課題名（英文）Development of a data stream management　system using special hardware assistance

研究代表者

吉永　努（Yoshinaga, Tsutomu）

電気通信大学・大学院情報理工学研究科・教授

研究者番号：６０２１０７３８

交付決定額（研究期間全体）：（直接経費）　　3,700,000円

研究成果の概要（和文）：ストリームデータ処理の主要演算として集約演算と結合演算を対象とし，それらを実行するための専用ハードウェアを設計し，FPGAに実装した．また，ユーザがビッグデータ処理アプリケーションで使用するクエリプランをSQLで入力し，FPGAで実行するためのハードウェア構成情報を出力するクエリコンパイラと，ネットワーク結合した複数のFPGA搭載ホスト計算機を活用するためのシステムソフトウェアを試作した．
FPGAボードを搭載する4ノード計算システムを用いた実験の結果，ウィンドウサイズ4096タプル／ノードの場合，ソフトウェアに比べて約9.2倍の結合演算スループット，58.8倍の電力当たり性能を達成した．

研究成果の概要（英文）：We designed special hardware modules which accelerate aggregation and join operations for stream data process in big data analytics. We also developed a query compiler which generates configuration information for FPGAs from SQL-based query plans as well as distributed system software to utilize a PC cluster with interconnected FPGAs.
We have obtained approximately 9.2 times faster join throughput and 58.8 times performance per energy, compared to conventional software-based join operation, on an FPGA-enhanced 4-node PC cluster when a sliding window size is 4096 tuples per node.

研究分野：　リコンフィギャラブルシステム

キーワード：　データストリーム処理　リコンフィギャラブル計算　FPGA　専用ハードウェア　結合演算　集約演算　スライディングウィンドウ

１．研究開始当初の背景

（1）コンピュータ，携帯端末，センサーなどあらゆる情報機器がネットワークに接続され，それらからもたらされるビッグデータ活用が進むようになった．また，情報機器のスマート化やネットワークの高速化を背景に，データの大規模化が加速し，ストリームデータに対するリアルタイム処理の重要性が高まりつつあった．

（2）蓄積されたビッグデータの分析は実用化しつつあるものの，ストリームデータ処理はやや新しい概念であり，高速ネットワークから連続的に入力されるデータを取りこぼすことなく低遅延・高スループットで且つ柔軟にクエリ処理するためのデータストリーム管理システムはまだ実用化されていなかった．

２．研究の目的

以下の３つの研究目的を掲げ，本研究を実施した．

（1）高速・低電力なストリームデータ処理用ハードウェアを開発し，グリーンコンピューティング技術に貢献する．

（2）上記 1 で開発するストリームデータ処理用ハードウェアを用いてデータベース演算を実行するためのシステムソフトウェア開発し，ビッグデータ処理アプリケーション実行コストを削減する．

（3）上記 1 と 2 を用いたストリームデータ演の実行性能／電力を実験的に明らかにすると共に，その有効性を示す．

３．研究の方法

以下の３つの研究サブテーマを立て，それぞれのサブテーマ間で連携して研究を行った．

（1）高速・低電力なストリームデータ処理用ハードウェアの開発：　ストリームデータの結合演算と集約演算を実行するハードウェアモジュールを設計し，ハードウェア記述言語（HDL）を用いて FPGA への実装を行った．開発したハードウェアモジュールを各ノードに搭載する図１のような PC クラスタ（図１は４ノード構成の例）を構成し，複数 FPGA ボード間を専用の高速ネットワークで相互結合する．

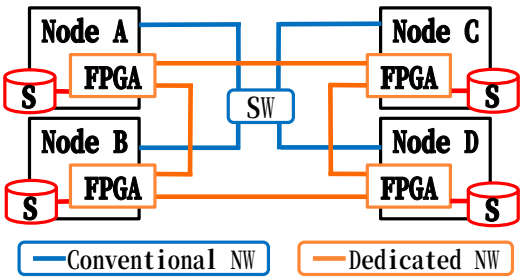

図１．FPGA ボード搭載 PC クラスタ

（2）FPGA 搭載 PC クラスタ用システムソフトウェアの開発：　図１に示した FPGA 搭載 PC クラスタの各ノード上で動作し，複数 FPGA 間のデータ転送や FPGA に実装するストリームデータ処理用ハードウェアを制御するための分散システムソフトウェア TAMAMO を開発した．

（3）性能評価：　開発した FPGA 搭載型専用ハードウェアを用いて，ストリームデータの結合演算と集約演算の実行性能／電力を測定し，ソフトウェア実行との比較を行った．

４．研究成果

（1）ストリームデータ結合演算用ハードウェア

図２に，設計したストリームデータ結合演算用ハードウェアのアーキテクチャを示す．

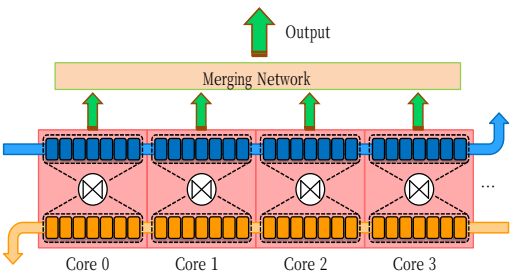

図２．Handshake Join の実装

結合演算アルゴリズム Handshake Join を我々の知る限り本研究で初めてハードウェア実装した．図２は単一 FPGA 内に 4 つの Join コアを実装する場合を示している．４つの Join コアが並列に結合演算を実行することにより，高い演算性能を実現する．さらに，Join コアを搭載する FPGA を図１のように専用ネットワークを用いて複数相互接続することにより，単一 FPGA では対応できない大きなウィンドウサイズに対応した．

（2）ストリームデータ集約演算用ハードウェア

図３に，設計したストリームデータ集約演算用ハードウェアのブロック図を示す．図中の CQPH(Configurable Query Processing Hardware)が，我々が出願中の特許（特開 2016-095606）に基づいて開発したクエリ変更に容易に対応可能な集約演算ハードウェアモジュールである．CQPH は，FPGA に直結した高速光ネットワーク GiGA CHANNEL から入力されるストリームデータを受け取って，集約演算を行った後，

LRA(Local Register Array)を介して結果をホスト PC に転送する．また，FPGA ボード上の DRAM に集約演算の途中結果を格納することで FPGA オンチップメモリだけでは扱うことができない大きなウィンドウサイズに対応することができる．

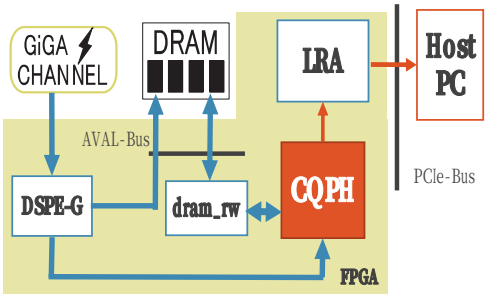

図３．データ集約演算ハードウェア

（3）FPGA 搭載 PC クラスタ用システムソフトウェア TAMAMO

図４に，開発した TAMAMO のソフトウェアスタックを示す．TAMAMO は PC クラスタのノードで動作する Linux 上で，イン・データパス計算（IDC）と FPGA 間直接通信（DDT）機能を支援する関数群を提供する．

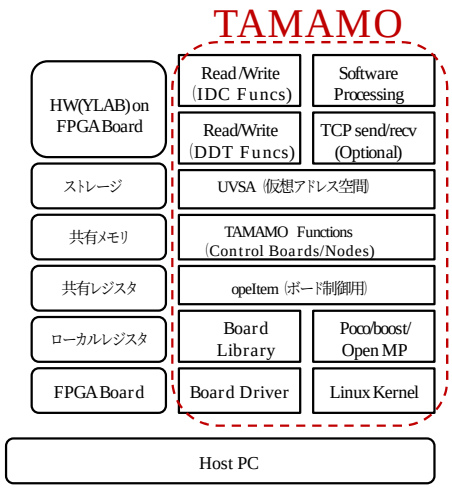

図４．TAMAMO ソフトウェアスタック

（4）結合演算性能

TAMAMO の動作する FPGA 搭載 PC クラスタのノード（Intel Core i7-6700K 4GHz, DDR4 DRAM 32GB）数を１～16 まで変化させ，FPGA ハードウェア（AVAL APX7142 改）で Handshake Join を実行した場合（HWHJ），ホスト PC 上のソフトウェアで Handshake Join を実行した場合（SWHJ），通信を TAMAMO から 10G イーサネット上の MPI_Bcast にした場合，及び TCP 通信を用いた場合の４つを比較した．なお，各ノードが処理するストリームデータのウィンドウサイズは 4096 タプル（128 ビットデータ／タプル）とした．

図５に結果を示す．TAMAMO+HWHJ は，TAMAMO +SWHJ に比べて約 5.3 倍，MPI_Bcast+SWHJ に比べて約 9.2 倍高速である．これにより，FPGA による結合処理の高速化が実験的に確認できる．また，TAMAMO は 10G イーサネット上の MPI_Bcast や TCP よりも高速な通信機能を提供することも確認できる．

電力当たりの結合演算性能は，ノード数が増えるほど FPGA が有利となり，4 ノード時に TAMAMO+HWHJ は MPI_Bcast+SWHJ に比べて約 58.9 倍である．

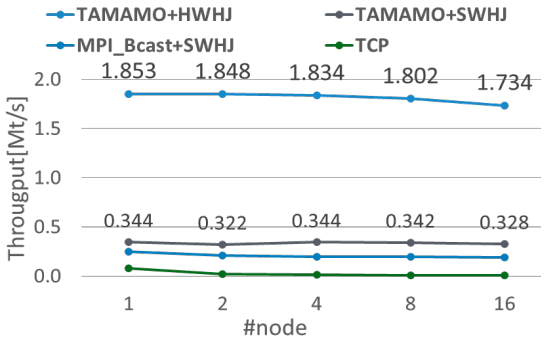

図５．結合演算性能

（5）集約演算性能

図６に集約演算性能の測定結果を示す．３つの折れ線グラフは，①GiGA CHANNEL から入力されるストリームデータを FPGA で集約演算実行した場合，②ハードディスクに格納済みのデータを PC 上のソフトウェア（C++プログラム）で集約演算実行した場合，③10G イーサネットから入力されるストリームデータを PC 上のソフトウェアで集約演算実行した場合，の比較を示す．図６の横軸は，スライディングウィンドウのウインドウ幅を表す．

FPGA 実行は，ウィンドウ幅に関わらず約

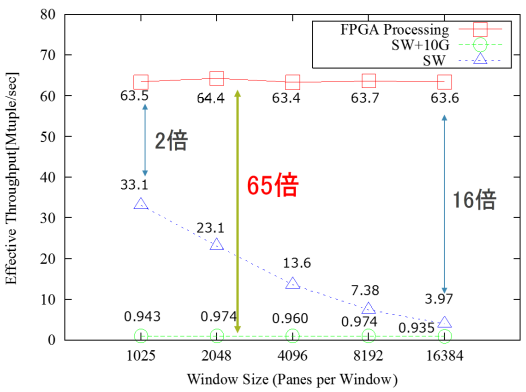

図６．集約演算性能

63.5 メガタプル／秒のスループットを維持するが，②のソフトウェア実行ではウィンドウ幅が大きくなり集約演算データが増えるに従いスループットが低下する．また，③では 10G イーサネットの通信プロトコル処理がボトルネックになり１メガタプル／秒未満の性能となった．以上より，FPGA による集約演算の有効性を確認できる．

５．主な発表論文等

〔雑誌論文〕（計　2 件）

① Masato YOSHIMI, Yasin OGE, and Tsutomu YOSHINAGA: Pipelined Parallel Join and Its FPGA-Based Acceleration, ACM Transactions on Embedded Computing System, 査読有, accepted.

② Yasin OGE, Masato YOSHIMI, Takefumi MIYOSHI, Hideyuki KAWASHIMA, Hidetsugu IRIE, and Tsutomu YOSHINAGA: Design and Evaluation of a Configurable Query Processing Hardware for Data Streams, IEICE Transactions on Information and Systems, 査読有, Vol.E98-D, No.12, 2015, pp.2207-2217.

〔学会発表〕（計　8 件）

① Masato Yoshimi, Yasin Oge, Celimuge Wu and Tsutomu Yoshinaga: Accelerating BLAST Computation on an FPGA-enhanced PC Cluster, Proc. of the Fourth International Symposium on Computing and Networking (CANDAR 2016), 査読有, pp.67-76, 2016/11/23, Higashi Hiroshima Arts and Culture Hall (Hiroshima, Higashi Hiroshima).

② Masato Yoshimi, Ryu Kudo, Yasin Oge, Yuta Terada, Hidetsugu Irie, and Tsutomu Yoshinaga: Accelerating OLAP workload on interconnected FPGAs with Flash storage, Proc. of the 2nd International Workshop on Computer Systems and Architectures (CSA'14), 査読有, pp.440-446, 2014/12/11, Shizuoka Convention & Arts Center (Shizuoka, Shizuoka City).

③ Masato Yoshimi, Ryu Kudo, Yasin Oge, Yuta Terada, Hidetsugu Irie, Tsutomu Yoshinaga: An FPGA-based Tightly Coupled Accelerator for Data-Intensive Applications, Proc. of the IEEE 8th International Symposium on Embedded Multicore/Manycore SoCs (MCSoC), 査読有, pp.289-296, 2014/09/24, Univ. Aizu (Fukushima, Aizu-Wakamatsu).

④ 多田昂介，川原尚人，吉見真聡，策力木格，吉永努： マルチノード FPGA によるストリームデータ分散結合処理，信学技法 CPSY2016-112, 査読なし, Vol.116, No.416, pp.37-42, 2017/01/23, Keio Univ. (Kanagawa, Yokohama).

⑤ 川原尚人，吉見真聡，策力木格，吉永努：ネットワーク結合型マルチ FPGA ボードを用いた集約演算クエリ処理，信学技法 CPSY2016-49，査読なし，Vol.116, No.240, pp.29-34, 2016/10/06, Makuhari-Messe (Chiba, Chiba city).

⑥ Masato YOSHIMI, Yasin OGE, Celimuge Wu, and Tsutomu YOSHINAGA: Design and Evaluation of Low-Latency Handshake Join on FPGA, IEICE, Tech. Report, CPSY2015-155, 査読なし，Vol.115, No.518, pp.253-258, 2016/03/25, Fukue Culture Center (Nagasaki, Goto).

⑦ 小川芳光，オゲヤースィン，吉見真聡，策力木格，吉永努: データストリーム集約演算 HW の並列化，信学技法 CPSY2015-119，査読なし，Vol.115, No.399, pp.79-84, 2016/01/19, Keio Univ. (Kanagawa, Yokohama).

⑧ 工藤龍，須戸里織，オゲ ヤースィン・寺田祐太・吉見真聡・入江英嗣・吉永努：複数 FPGA ボードを用いたビッグデータ分割処理の高速化，信学技法 CPSY2014-152, 査読なし，Vol.114, No.427, pp.193-198, 2015/01/29, Keio Univ. (Kanagawa, Yokohama).

〔図書〕（計　0件）

〔産業財産権〕

○出願状況（計　1 件）

名称：データ処理装置およびデータ処理方法，並びにプログラム
発明者：オゲヤースィン，吉見真聡，入江英嗣，吉永努
権利者：電気通信大学
種類：特許
番号：特開 2016-095606
出願年月日：2015 年 11 月 13 日
国内外の別： 国内

○取得状況（計　0件）

〔その他〕
TAMAMO ホームページ
https://github.com/nkawahara/tamamo

６．研究組織

(1)研究代表者

吉永　努　（YOSHINAGA, Tsutomu）
電気通信大学・大学院情報理工学研究科・
教授
研究者番号：　６０２１０７３８

(2)研究分担者

吉見　真聡　（YOSHIMI, Masato）
電気通信大学・大学院情報理工学研究科・
助教
研究者番号：　００５４８０００