早稲田大学大学院 基幹理工学研究科

博 士 論 文 概 要

論 文 題 目

QueueLinker: A Framework for
Parallel Distributed Data-Stream Processing

QueueLinker: データストリームのための
並列分散処理フレームワーク

申 請 者

Takanori	UEDA
上田	高徳

情報理工学専攻 並列・分散アーキテクチャ研究

2013 年 1 月

モバイル機器やインターネットの普及により，センサ情報やネットワークトラフィックといった，永続的かつ大量に生成されるデータストリームが一般的になってきている．データストリームはリアルタイムに生成されるデータであり，解析処理をリアルタイムに行うことで情報の活用機会を増やすことができる．ＩＴ社会の発展と共に，情報源となる端末やセンサは急激に増えており，データストリーム量は今後さらに大きくなると予想される．大規模ストリームをリアルタイム処理するために，データストリームに対する並列分散処理を統合的にサポートするフレームワークが必要不可欠である．

データストリーム処理のアプリケーションは，処理レイテンシが重要なものから，スループットが重要なものまで多岐に渡る．たとえばアルゴリズム取引は，レイテンシが最も重要なアプリケーション例である．東京証券取引所のａｒｒｏｗｎｅｔでは片道３２マイクロ秒程度でデータを提供できるとされ，ウォールストリートのアルゴリズム取引では５マイクロ秒の差が勝敗を分けるともいわれている．いまや，アルゴリズム取引でのレイテンシ競争はマイクロ秒単位であり，処理レイテンシをベストエフォートで削減することで，競合相手より少しでも早く注文を行い，勝機を増すことができる．

ネットワーク侵入検知システムもレイテンシが重要なアプリケーション例である．侵入検知システムは，不正なパケットをリアルタイムに検知してフィルタする．この際，クライアント・サーバ間のレイテンシが延びると，ＴＣＰの性質により通信スループットが低下してしまうため，侵入検知システムにおいてはミリ秒単位の処理レイテンシが性能に影響するといえる．

一方，Ｔｗｉｔｔｅｒデータの解析結果をユーザに提示する場合を考えると，人間が閲覧する限り，秒単位のレイテンシは許容される．しかし，大量のＴｗｅｅｔを処理するためにスループット性能が要求される．さらに，今後発展が期待されるスマートシティにおいては，大量の電力センサや監視カメラなどのデータをリアルタイムに処理する必要がある．都市の発展に伴いセンサ数も増加し，ストリームとして到着するデータ量は膨大なものになる．

以上のようにデータストリーム処理のアプリケーションは，数マイクロ秒の差が問題になるアルゴリズム取引から，高スループット処理が必要なＴｗｉｔｔｅｒ解析やスマートシティのセンサデータ処理まで幅広い．大規模ストリームに対してこれらアプリケーションの要求を満たすためには並列分散処理が必要である．しかし，並列処理を実現するためのマルチスレッドプログラミングや，分散処理におけるネットワーク通信や障害対応など，並列分散処理を行うために解決すべき課題は多い．リアルタイム処理の要求から，Ｈａｄｏｏｐのようなバッチ処理型の分散処理フレームワークで処理を行うことは自然な方法ではない．データストリームは外部環境から到着するデータなため，ストレージに格納されたデータを高速に処理することを目的とした従来型データ処理の研究とは異なる研究課題

であり，データストリームに対する並列分散処理を統合的にサポートする並列分散処理フレームワークが必要となる．

申請者はデータストリームのための並列分散処理フレームワークＱｕｅｕｅＬｉｎｋｅｒを開発してきた．本論文ではまず，ＱｕｅｕｅＬｉｎｋｅｒのソフトウェアアーキテクチャについて述べる．そして，開発過程で研究を行ってきた，オペレータスケジューリング技法や実アプリケーションについて述べる．

本研究の主な貢献は以下の４点である．

（１）データストリーム処理のための並列分散処理フレームワークＱｕｅｕｅＬｉｎｋｅｒの開発：　ＱｕｅｕｅＬｉｎｋｅｒを用いると，Ｐｒｏｄｕｃｅｒ－Ｃｏｎｓｕｍｅｒモデルでモジュールを実装し，モジュールの接続関係と並列分散方法を指定することで，自動的な並列分散処理を行うことができる．プログラマは通信処理や並列分散実行に関わる実装を行う必要がない．

（２）マルチコアＣＰＵ環境における低レイテンシデータストリーム処理：　データストリームの並列処理においては，ＣＰＵコア間やスレッド間通信でレイテンシが発生するため，コア間通信を少なくすることで処理レイテンシを改善できる．本論文ではストリームの到着頻度や計算負荷に応じてオペレータへのＣＰＵコア割り当てを動的に変化させ，平均レイテンシを最小化する手法を提案した．

（３）分散処理におけるレイテンシ削減と高可用性を両立するオペレータ実行方法：　分散処理時には，計算機の障害発生に対応して高可用性を実現するため，オペレータの内部状態を複数の計算機にレプリケーションする必要がある．本論文では，プライマリとバックアップを異なるオペレータ配置で同時実行し，レイテンシ削減と高可用性を両立する手法を提案した．

（４）ＱｕｅｕｅＬｉｎｋｅｒ上で動作するＷｅｂクローラとＷｅｂデータ解析アプリケーションの開発：　ＱｕｅｕｅＬｉｎｋｅｒ上で動作するアプリケーションとして並列分散ＷｅｂクローラとＷｅｂデータ解析システムを開発し，ＱｕｅｕｅＬｉｎｋｅｒ上で動作できることを示した．

以上４つの貢献は，データストリームの並列分散処理を行う際の重要な問題であるレイテンシ短縮と高可用性の実現に貢献した研究であり，並列分散処理フレームワークＱｕｅｕｅＬｉｎｋｅｒと合わせて，今後のデータストリーム処理の分野に幅広く貢献するものと考える．

本論文は以下の構成をとる．第１章は本論文の序論として，データストリーム処理のアプリケーションについて述べ,本研究の位置付けについて明らかにする．

第２章では，データストリームの性質や，主にデータベース分野においてこれまで研究されてきた,データストリーム処理の既存研究について整理する.また，商用製品，オープンソースのデータストリーム処理系についてもまとめる．そして，データストリーム処理における研究課題と技術課題について明らかにする．

第３章では申請者が開発してきたデータストリーム処理のための並列分散処理

フレームワークであるＱｕｅｕｅＬｉｎｋｅｒのソフトウェアアーキテクチャについて述べる．ＱｕｅｕｅＬｉｎｋｅｒを用いることで，データストリームに対する処理をＰｒｏｄｕｃｅｒ－Ｃｏｎｓｕｍｅｒモデルで記述することができる．本章ではＱｕｅｕｅＬｉｎｋｅｒを例として，データストリーム処理の研究課題について改めて明確にする．

第４章では，データストリームの低レイテンシ並列処理の際に課題となる，関係代数オペレータへのＣＰＵコア割り当てについて提案手法を述べる．並列処理の際にオペレータごとにスレッドを割り当てると，ＣＰＵコア間通信やスレッド待機のオーバーヘッドによりレイテンシが増大する．逆にスレッド数が少なすぎては並列性を生かせず，処理できるデータ量に限界が生じる．提案手法ではＣＰＵアーキテクチャやスレッド待機のオーバーヘッドを考慮し，処理レイテンシを短縮するスレッド割り当て手法を提案する．マルチコア環境におけるデータストリーム処理のレイテンシ定義を与え，モデル上で最適なスレッド割り当てが動的計画法で求まることを示す．さらに，入力ストリームのデータレート変化に応じてオペレータを再配置する際，ストリーム処理を止めずにタプル適用順序を守ってオペレータを再配置する方法を提案する．

第５章ではデータストリームを分散処理する際に必用な高可用性について議論を行う．分散データストリーム処理では，各オペレータをどの計算機で実行するかが問題になる．レイテンシを重要視するアプリケーションでは，可能な限り少数の計算機で処理を実行すると計算機間通信に起因するレイテンシを削減できるが，単位時間あたりの処理対象データが増加した時，計算機の負荷が高まり逆にレイテンシが悪化する可能性がある．提案手法であるＣｈａｓｅ　Ｅｘｅｃｕｔｉｏｎは，プライマリと異なるオペレータ配置でバックアップを同時実行し，プライマリとバックアップで先に出力された方のタプルを採用することで，レイテンシの最小化と障害対応の両立を目指す．

第６章ではＱｕｅｕｅＬｉｎｋｅｒで動作するアプリケーションとして開発した並列分散Ｗｅｂクローラと Ｗｅｂデータ解析アプリケーションを示す．ＱｕｅｕｅＬｉｎｋｅｒは低レイテンシ処理に特化したフレームワークではない．本論文で提案するＷｅｂクローラは，クローリング処理をＰｒｏｄｕｃｅｒ－Ｃｏｎｓｕｍｅｒ型のモジュール群で分割実行し，全てのモジュールに任意数のスレッドと計算機を割り当てて並列分散実行することができ，Ｗｅｂサイトごとに分割処理する従来の方法よりも細かな粒度での並列分散実行により負荷分散を実現できる．また本章では，ＱｕｅｕｅＬｉｎｋｅｒ上で動作する多メディアＷｅｂデータ解析のアプリケーションについても示す．

第７章では提案手法についてまとめ,残された課題について議論する.そして，データストリーム処理およびデータ処理一般における今後の展望について述べ，本論文の結論とする．

早稲田大学　博士（工学）　学位申請　研究業績書

氏 名　　上田　高徳　　印

（2013 年　1 月　現在）

種 類 別	題名、　発表・発行掲載誌名、　発表・発行年月、　連名者（申請者含む）
論文	
○	[1] 上田高徳，佐藤亘，鈴木大地，打田研二，森本浩介，秋岡明香，山名早人，「Producer-Consumer 型モジュールで構成された並列分散 Web クローラの開発」，情報処理学会論文誌 データベース，57 号，Mar. 2013（掲載決定）.
○	[2] 上田高徳，秋岡明香，山名早人，「マルチコア CPU 環境における低レイテンシデータストリーム処理」，電子情報通信学会論文誌 D，vol. 96, no. 5, May 2013（掲載決定）.
○	[3] 上田高徳，打田研二，秋岡明香，山名早人，「データストリーム処理におけるレイテンシ削減と高可用性のためのオペレータ実行方法」，日本データベース学会論文誌，vol.10, no.3, pp.1-6, Feb. 2012.
	[4] Hiroki Asai, **Takanori Ueda** and Hayato Yamana, “Legible Thumbnail: Summarizing On-line Handwritten Documents based on Emphasized Expressions,” In *Proc. of the 13th International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI), Stockholm, Sweden, Aug. 2011 (Poster).
	[5] 久保田展行，上田高徳，山名早人，「ウェブクローラ向けの効率的な重複 URL 検出手法」，日本データベース学会論文誌，vol.8，no.1，pp.83-88，Jun. 2009.
○	[6] **Takanori Ueda**, Yu Hirate and Hayato Yamana, “The Challenge of Eliminating Storage Bottlenecks in Distributed Systems,” In *Proc. of the 1st International Workshop on Software Technologies for Future Dependable Distributed Systems* (STFSSD), Tokyo, Japan, Mar. 2009.
	[7] Sayaka Akioka, Junichi Ikeda, **Takanori Ueda**, Yuki Ohno, Midori Sugaya, Yu Hirate, Jiro Katto, Shigeki Goto, Yoichi Muraoka, Hayato Yamana and Tatsuo Nakajima, “A Scalable Monitoring System for Distributed Environment,” In *Proc. of the 1st International Workshop on Software Technologies for Future Dependable Distributed Systems* (STFSSD), Tokyo, Japan, Mar. 2009.
○	[8] **Takanori Ueda**, Yu Hirate and Hayato Yamana, “Exploiting Idle CPU Cores to Improve File Access Performance,” In *Proc. of the 3rd International Conference on Ubiquitous Information Management and Communication* (ICUIMC), Suwon, Korea, Jan. 2009.
	[9] 舟橋卓也，上田高徳，平手勇宇，山名早人，「商用検索エンジンのヒット数に対する信頼性の検証」，日本データベース学会論文誌，Vol.7，No.3，Dec. 2008.
	[10] 舟橋卓也，上田高徳，平手勇宇，山名早人，「商用検索エンジンの検索結果では取得できないランキング下位部分の収集・解析」，日本データベース学会論文誌，Vol.7，No.1，pp.37-42，Jun. 2008.

早稲田大学　博士（工学）　学位申請　研究業績書

種 類 別	題名、　発表・発行掲載誌名、　発表・発行年月、　連名者（申請者含む）
○	[11] 上田高徳，平手勇宇，山名早人，「システムコールレベルのアクセスログを用いたディスクアクセスパターンマイニング」，日本データベース学会論文誌，Vol.7，No.1，pp.145-150，Jun. 2008. [12] 片瀬弘晶，松永拓，上田高徳，田代崇，平手勇宇，山名早人，「リンク構造解析アルゴリズム高速化のための縮小 Web の構築」，日本データベース学会論文誌，Vol.7，No.1，pp.245-250，Jun. 2008. [13] Yasuaki Yoshida, **Takanori Ueda**, Takashi Tashiro, Yu Hirate and Hayato Yamana, “What's going on in search engine rankings?,” In *Proc. of the 2008 IEEE International Symposium on Mining And Web* (MAW), Okinawa, Japan, Mar. 2008.
○	[14] **Takanori Ueda**, Yu Hirate and Hayato Yamana, “EReM-DiCE: Exploiting Remote Memory for Disk Cache Extension,” In *Proc. of the 1st International Workshop on Storage and I/O Virtualization, Performance, Energy, Evaluation and Dependability* (SPEED), Salt Lake City, US-UT, Feb. 2008. [15] Takashi Tashiro, **Takanori Ueda**, Taisuke Hori, Yu Hirate and Hayato Yamana, “EPCI: Extracting Potentially Copyright Infringement Texts from the Web,” In *Proc. of the 16th International World Wide Web Conference* (WWW), Banff, Canada, pp.1151-1152, May 2007 (Poster). [16] 田代崇，上田高徳，堀 泰祐，平手勇宇，山名早人，「Web 上の文章を対象とした著作権違反自動検知システム」，日本データベース学会 Letters，vol.5，no.2，pp.25-28，Sep. 2006.
講演	[1] 上田高徳，浅井洋樹，藤木紫乃，山本祐輔，武井宏将，秋岡明香，山名早人，「ソーシャルメディアを含む多メディアビッグデータの統合的解析による情報抽出」，第 156 回データベースシステム研究発表会（DBS），Dec. 2012 (To Appear). [2] 上田高徳，佐藤亘，鈴木大地，打田研二，森本浩介，秋岡明香，山名早人，「Producer-Consumer 型モジュールで構成された並列分散 Web クローラの開発」，第 5 回 Web とデータベースに関するフォーラム（WebDB Forum），Nov. 2012. [3] 上田高徳，秋岡明香，山名早人，「マルチコア環境における低レイテンシストリーム処理のためのスレッド割り当て手法」，第 4 回データ工学と情報マネジメントに関するフォーラム(DEIM)，Mar. 2012. [4] 打田研二，上田高徳，山名早人，「カスタマイズ性とリアルタイムなデータ提供を考慮したクローラの設計と実装」，第 4 回データ工学と情報マネジメントに関するフォーラム（DEIM），Mar. 2012. [5] 鈴木大地，佐藤亘，上田高徳，山名早人，「分散 Key-Value データベースを用いた RDF ストアの構築と評価」，第 4 回データ工学と情報マネジメントに関するフォーラム（DEIM），Mar. 2012.

早稲田大学　博士（工学）　学位申請　研究業績書

種 類 別	題名、　発表・発行掲載誌名、　発表・発行年月、　連名者（申請者含む）
	[6] 田中友樹，山本祐輔，上田高徳，山名早人，「検索時間と再現率の調節可能な類似動画検索手法」，第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM)，Mar. 2012. [7] 上田高徳，打田研二，秋岡明香，山名早人，「データストリーム処理におけるレイテンシ最小化と高可用性のためのオペレータ実行方法」, 第 4 回 Web とデータベースに関するフォーラム（WebDB Forum），Nov. 2011. [8] 森本浩介，上田高徳，打田研二，山名早人，「ウェブサーバへの最短訪問間隔を保証する時間計算量が O(1)のウェブクローリングスケジューラ」，第 3 回データ工学と情報マネジメントに関するフォーラム（DEIM），Feb. 2011. [9] 上田高徳，片瀬弘晶，森本浩介，打田研二，油井誠，山名早人，「QueueLinker: パイプライン型アプリケーションのための分散処理フレームワーク」，第 2 回データ工学と情報マネジメントに関するフォーラム（DEIM），Feb. 2010. [10] 片瀬弘晶，上田高徳，山名早人，「LittleWeb: 類似ノード集約による Web グラフ圧縮手法」，第 2 回データ工学と情報マネジメントに関するフォーラム（DEIM），Feb. 2010. [11] 上田高徳, 片瀬弘晶, 森本浩介, 打田 研二, 山名早人，「QueueLinker: Distributed Producer/Consumer Queue Framework」，Web とデータベースに関するフォーラム（WebDB Forum），Nov. 2009（招待ポスター）. [12] 上田高徳，平手勇宇，山名早人，「アクセスパターンマイニングによる OS レベルでの動的な I/O 最適化」, 情報処理学会研究報告(DBS) / iDB2008, vol.2008, no.88, pp.73-78，Sep. 2008. [13] 上田高徳，「メニーコア時代における OS レベルでの I/O 最適化」，情報処理学会研究報告（jDB ワークショップ），vol.2008，no.56，p.133，Jun. 2008. [14] 上田高徳，平手勇宇，山名早人，「システムコールレベルのアクセスログに対するディスクアクセスパターンマイニング」，第 19 回データ工学ワークショップ・第 6 回日本データベース学会年次大会（DEWS），Mar. 2008. [15] 上田高徳，平手勇宇，山名早人，「リモートメモリを用いたランダムディスクアクセス高速化手法」，情報処理学会研究報告（ARC），vol.2007，no.79，pp.151-156，Aug. 2007. [16] 上田高徳，平手勇宇，山名早人，「ネットワーク上のマシンをディスクキャッシュに利用した場合の性能評価」，第 18 回データ工学ワークショップ・第 5 回日本データベース学会年次大会（DEWS），Mar. 2007. その他 10 件