

博士論文審査報告書

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	高德

情報理工学専攻 並列・分散アーキテクチャ研究

2013年3月

ICT (Information and Communication Technology) の発展に伴い, リアルタイムの情報ソースから提供されるデータが爆発的に増大すると共に, その解析が急務となっている. センサ情報, 金融情報, Web 情報といった, 永続的かつ大量に生成されるデータストリームは即時性が高く, 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大させることができる. データストリームを利用するアプリケーションには, Twitter のような Web データのリアルタイム解析を行うものから, 金融取引などの低レイテンシ処理が必要なものまで様々なものがある. 特に金融市場におけるアルゴリズム取引でのレイテンシ競争は今やマイクロ秒単位であり, 処理レイテンシを可能な限り削減することが求められている.

さらに, 今後のスマートシティ化に伴い, データストリーム量は飛躍的に増大し, その活用機会は一段と高まるものと予想される. このような大規模なデータストリームに対応するためには, 並列分散処理が必要になる. マルチコアプロセッサやコモディティクラスターの普及により, 並列分散処理の環境は安価に構築できるようになったが, 開発者誰もが並列分散処理を行うアプリケーションを簡単に実装できるわけではない. より簡単に通信処理や排他制御を実装できるフレームワークの必要性が高まっている.

本論文では, このような背景の下, データストリームを並列分散処理するためのフレームワーク **QueueLinker** を提案し, これを実装してその有効性を評価している. 本フレームワークの利用により, 通信処理や排他制御などの並列分散処理を意識することなく, プログラム開発を行うことが可能となる. さらに, **QueueLinker** が持つコア技術として, 低レイテンシ処理を実現するオペレータ並列実行手法と分散バックアップ手法を提案している. そして, **QueueLinker** の応用例として, 並列分散型の Web クローラを実現している.

以下, 各章の概要とそこでの成果について説明する.

第 1 章は序論であり, 本研究の意義と得られた結果の概要を示している.

第 2 章ではデータストリームの特徴とデータストリームを処理するデータストリーム管理システム (**DSMS**, **Data Stream Management Systems**) について述べている. **DSMS** は連続問い合わせ (**CQ**, **Continuous Query**) をデータストリームに対してリアルタイムに処理する. **SQL-Like** な言語で表現される **CQ** は, 関係代数オペレータで構成されるプラン木に変換され実行される. **CQ** の処理においては, 関係代数オペレータの実行戦略が **CQ** のレイテンシを左右することになる. しかし, データストリームは **DSMS** の外部から到着するデータであり, データの性質や到着頻度が事前には分からない点や, 外部イベントなどにより到着頻度が変化する点が実行戦略を困難なものとしている. **CQ** に対して低レイテンシ処理を実現するためには, これらデータストリームの流量変化に対応した処理を行う必要がある.

既存研究では, 低レイテンシを実現するために, **CQ** を構成する関係代数

オペレータをどのような優先度で実行するかという実行戦略について提案がなされている。しかし、既存の実行戦略は、関係代数オペレータと CPU コアの配置が考慮されていない点、データストリームの流量変化に頑健な実行戦略となっていない点について問題提起している。

第 3 章では、提案フレームワークである **QueueLinker** について述べている。**QueueLinker** は、**Producer-Consumer** モデルによる API を提供している。開発者はデータ並列処理の基本単位となる「モジュール」の Java 実装と、モジュール間のデータ転送経路を示す論理グラフを与えることで、並列分散処理を容易に実現できる。各モジュールに対してデータ並列処理方法を指定することで、データ並列による並列分散実行が **QueueLinker** により自動的に行われる。具体的には、データ内容のハッシュ値に基づいて、**QueueLinker** がデータ転送先のモジュールインスタンスを決定し、そのインスタンスが動作している計算機及びスレッドを特定しデータを転送する。開発者は通信や並行制御に関わる記述を行う必要がないが、各モジュールの並列数や動作する計算機を開発者側が制御することも可能である。このように、通信処理や排他制御といった並列分散処理を意識することなく、プログラム開発を行うことができるフレームワークとなっている。

第 4 章では、**CQ** の低レイテンシ並列処理技法として、データストリーム流量の変化に応じてオペレータの CPU コアへの再割り当てを効率よく行う手法を提案している。本手法は、**QueueLinker** に実装されており、**CQ** の関係代数オペレータ（複数の命令から構成）をモジュールとし、プラン木を論理グラフと考えることで、**QueueLinker** 上で実行することができる。

具体的には、**QueueLinker** 上での処理単位であるモジュールに対して、CPU コアレベルでの割り当て制御を行うと共に、処理のオーバーロードが発生しない範囲で、CPU コア間のデータ通信量が少なくなるように CPU コアへの配置を行っている。その上で、データストリーム流量が変化した際に、適切な再割り当てを動的に計算すると共に、既に実行中のモジュールとタプル（データ）を移動させる方法を提案している。本提案の貢献は、モジュールとタプル（データ）の移動を矛盾なく行う点にあり、再配置計算アルゴリズムは、既存手法を含めて様々なものを用いることができる。評価実験では、10 マイクロ秒レベルで処理が可能な **CQ** を用い、データストリームの流量を最大 4 倍変化させた時、レイテンシ変動を 1 マイクロ秒以下に抑えることができることを確認している。これにより、金融市場におけるアルゴリズム取引のように数マイクロ秒単位の処理レイテンシが問題となるアプリケーションにおいても、提案手法が有効であることを示している。

第 5 章では、**CQ** の分散処理において低レイテンシ処理を可能とするバックアップ技法を提案している。耐故障性の向上は、ストリーム処理において重要な課題であり、一般的にはプライマリの他にセカンダリを設けることで対応がなされている。これに対して、本章では、セカンダリを耐故障性向上

のためにだけ用いるのではなく，セカンダリをプロアクティブに利用することにより，低レイテンシを実現する **Chase Execution** を提案している．

Chase Execution は，バックアップであるセカンダリをプライマリとは異なるモジュール配置で動作させる．評価実験を通じ，プライマリとセカンダリの内，早く結果を出力した方の結果（データ）を採用することで，プライマリのみ，あるいはセカンダリのみを用いる場合に比較して，レイテンシが削減できることを確認している．本手法はバックアップが必要な場合に，バックアップの出力をも利用することでレイテンシを削減することができるという新しい考え方を提案している点が大きく評価できる．

第6章では，**QueueLinker** を用いた並列分散型の **Web** クローラを提案している．**Web** クローラを **QueueLinker** のモジュールにより実装することで，並列分散クロールが容易に実現される．本クローラは全てのモジュールがデータ並列で動作し，**QueueLinker** の機能により，任意の並列数と分散数で動作できる点が大きな特色である．既存の分散 **Web** クローラは，収集対象となる **Web** サーバ単位での負荷分散戦略をとっている．しかし，この戦略は容易に分散クロールを実現できる一方で，負荷の偏りが問題となる．これに対して，提案する **Web** クローラは，クローラの機能ごとに並列処理を適用できるため，上述した負荷の偏りの問題を解決できる．実証実験により，毎秒 4,000 ページのダウンロードに成功しており，既存研究での最高性能とされる毎秒 1,789 ページを大幅に上回る性能を達成しており，実用面でも大きく評価できる．

第7章では，まとめと今後の課題について言及している．

以上を要するに，本論文では，データストリーム処理のための新しい並列分散処理フレームワーク **QueueLinker** を提案し，データストリームを利用するアプリケーションにおいて重要となる低レイテンシ処理を，特に **CQ** を対象として実現している．さらに，提案フレームワークにより **Web** クローラのような実用的なアプリケーションを簡単に構築できると共に，大きな性能向上を達成できることを示しており高く評価できる．よって本論文は，データ工学の発展に大いに貢献するものであり，博士（工学）早稲田大学の学位論文として価値あるものと認める．

2013年3月

審査員

主査	早稲田大学教授	博士（工学）（早稲田大学）	山名 早人
	早稲田大学教授	Ph.D.（イリノイ大学）	村岡 洋一
	早稲田大学教授	工学博士（慶應義塾大学）	中島 達夫
	ニュージャージー	Ph.D.	
	工科大学 准教授	（南カリフォルニア大学）	Andrew Sohn