

令和元年度 情報工学コース卒業研究報告要旨

石川 研究室	氏 名	牧田 直樹
卒業研究題目	メニーコアシステムにおける分散ストリーム処理 システムの性能評価 － 遅延に関する評価 －	

Web サービスのトラフィック解析や e コマースでの取引など、ビッグデータに対するリアルタイムな解析・処理の要求が増加している。そういった場面でストリームデータ処理が重要視されている。ストリームデータ処理とは、無制限に発生する大量のデータをリアルタイムに処理することである。

そのユースケースから、ストリーム処理システムは高い処理性能を求められ、様々な分散並列ストリーム処理システムの研究開発が学術・産業両面で活発に行われている。

主な分散並列ストリーム処理 OSS (open source software) はシェアードナッシングアーキテクチャを採用しており、複数のマシンを使用することで性能のスケールアウトを可能としている。その一方で、近年ではメニーコア CPU による各マシン単位での性能のスケールアップの潮流が見られる。

しかし、メニーコア CPU を搭載したマシンにおける既存の分散並列ストリーム処理 OSS の基本的な性能評価に関する知見は十分とは言えない。

そこで、本研究では既存の分散並列ストリーム処理 OSS のメニーコア CPU における処理性能を調査し、特にレイテンシの面からその評価を行った。本研究では時間窓を連続する問い合わせとみなし、「時間窓が生成された時間」と「時間窓から結果が出力された時間」の差をレイテンシとして定義した。イベント時間窓を用い、主に時間窓の幅、データセットのキーの偏り、並列数の3つをパラメータとして変化させ、評価を行った。

図1はOSSストリーム処理フレームワーク Apache Flink において時間窓の幅を 100 ms、キーの偏りをなしとした場合のベンチマーク結果の一例である。Apache Flink に対してベンチマークを行った結果、並列数の増加によるスループットの向上は見られたが、レイテンシは並列数と比例して増加している傾向が見られた。

今後の課題として、Apache Samza, Apache Storm, Spark Streaming などの OSS に対しても同様のベンチマークを行い、その性能を比較することが挙げられる。

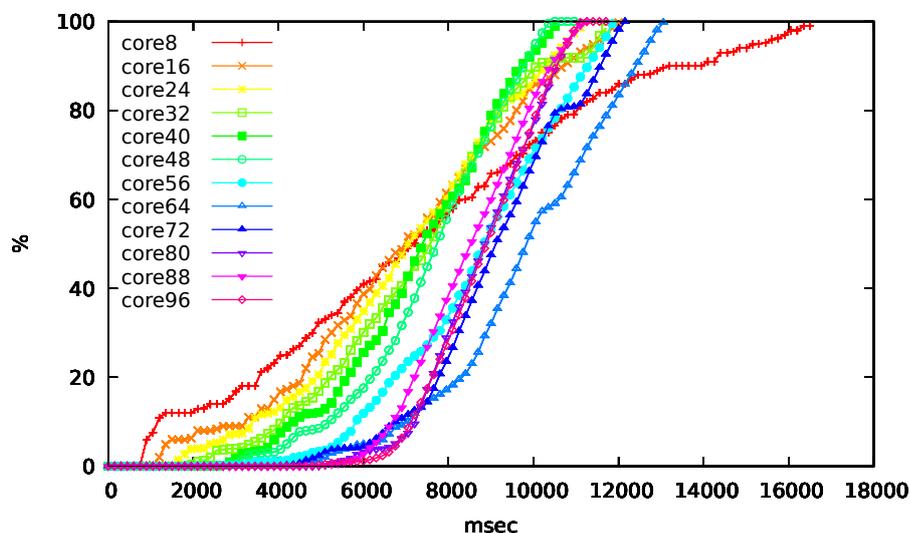


図1: Apache Flink におけるベンチマークの一例