

データ集約科学の基盤システムに関する研究

A Study on Infrastructure System for Data Intensive Sciences

川島英之

慶應義塾大学環境情報学部

1. 研究目的

近年の計算科学ではシミュレーション結果を分析するために並列分散データ分析が必要になりつつある。並列分散データ分析を実現するには、分析処理系が必要である。本プロジェクトの目的は、スーパーコンピュータを用いてこの分析システムを高速化する技法を探求することだった。

2. 研究成果の内容

本研究の成果は並列分散処理スキームである **MapReduce** をスクラッチから開発し、その **shuffling phase** を独自で高性能化したことである。この研究では、偏ったデータによって引き起こされる **MapReduce** シャッフルリングの問題に対処するために設計された3つのメモリ内シャッフル方法を提案し、検証をおこなった。結合シャッフルアーキテクチャ (**CSA**) は、対応するブロックのメタデータを含む両方のブロック、シャッフル転送の単位、およびメタブロックをシャッフルするために、単一の対全対全交換を使用する。分離シャッフルアーキテクチャ (**DSA**) は、メタブロックとブロックのシャッフルを分離し、それぞれのシャッフルプロセスに異なる全対全交換アルゴリズムを適用して、偏った分布におけるストラグラの影響を軽減する。

Skew-Aware Meta-Shuffle (DSA w / SMS) を使用した分離型シャッフルアーキテクチャは、各ワーカプロセスのメモリ消費量に基づいてブロックの適切な配置を自律的に決定する。このアプローチは、一部のワーカプロセスがノードメモリの制限を超える可能性がある、非常に偏った状況を対象とした。この研究では、**In niBand** や **Intel Omni-Path** などの高性能インターコネクトを採用した、プロトタイプのインメモリ **MapReduce** エンジンでの3つのシャッフルメソッドの実装を評価した。本研究の結果は、**DSA w / SMS** が極端に歪んだデータ分布に対して唯一の実行可能な解決策であることを示唆した。

3. 学際共同利用として実施した意義

本研究を学際共同利用として実施したことにより、筑波大学計算科学研究センターに在籍する多数の応用研究者（気象等）と議論を行うことができ、システム研究者としてはシステム設計に関して多くの気づきを得ることができた。

4. 今後の展望

提案アルゴリズムを更に発展させ、様々なアプリケーションに適用可能にしたい。

5. 成果発表

(1) 学術論文

(2) 学会発表

Harunobu Daikoku, Hideyuki Kawashima, Osamu Tatebe: Skew-Aware
Collective Communication for MapReduce Shuffling. BigData 2018: 3331-3340

(3) その他

使用計算機	使用計算機 に○	配分リソース※	
		当初配分	追加配分
COMA			
Oakforest-PACS	○	16384	
※配分リソースについてはノード時間積をご記入ください。			