

大規模 CFD 解析におけるポスト処理効率化のためのデータステージング技術に関する研究

報告書番号：R18JACA42

利用分野：JSS2 大学共同利用

URL：<https://www.jss.jaxa.jp/ar/j2018/8982/>

● 責任者

下條真司, 大阪大学 サイバーメディアセンター 応用情報システム研究部門

● 問い合わせ先

下條真司(shimojo@cmc.osaka-u.ac.jp)

● メンバ

下條 真司, 高橋 慧智

● 事業概要

CFD シミュレーションの大規模化にともない、従来のように全てのシミュレーション結果を並列ファイルシステムに保存し、シミュレーション終了後にポスト処理を実行することはストレージ容量や IO 帯域幅の制約により困難になると予想される。そのため、シミュレーション実行中にポスト処理プログラムへ計算結果をリアルタイムに転送する、データステージング技術が注目されている。本研究では、JSS2 に代表される HPC 環境におけるデータステージング技術の実現可能性を検討し、データステージングミドルウェア、および、HPC 環境に求められる要件を分析する。

● JSS2 利用の理由

JSS2 は計算用主システム (SORA-MA) とプリポスト処理用副システム (SORA-PP) 間で通信が可能であることから利用した。

● 今年度の成果

SORA-MA システム上で CFD シミュレーション、SORA-PP システム上でポスト処理をそれぞれ実行すると想定し、これらのプログラム間でデータステージングの実現可能性を検討した。データステージングのためのミドルウェアとしては、米国オークリッジ国立研究所において Exa-scale Computing Project の一環として開発されている Adaptive IO System (ADIOS2) を採用した。まず、ADIOS2 を用い SORA-MA システム上の計算ノード間、および、SORA-PP システム上の計算ノード間でステージング通信が可能であることを確認した。次に、MA 上の計算ノードから PP 上の計算ノードへ直接通信を試みた。しかし、システム間を越えるパケットをルーティングしないよう相互結合網が設定されていたため、実現できなかった。

MA-PP 間の通信を実現するため、PP と接続されている MA 上の一部の計算ノード (IO ノード) 上で MA-PP 間の通信を中継するプログラム (adios-reorganize) を動作させる手法を考案した。その結果、MA 上の計算ノードから PP 上の計算ノードへのステージング通信が可能であることを確認できた。

以上の実験により、JSS2 においてデータステージングが可能であることを実証した。その中で、ADIOS2 の SPARC64 XIfx プロセッサ対応、および、異種アーキテクチャ間のステージング通信に課題があることが明らかになったので、ADIOS2 の開発チームと連携し、ADIOS2 に対して機能拡張・バグ修正を実施した。また、HPC 環境側の課題として、MA の IO ノードを明示的に要求する機能がジョブスケジューラに存在すれば、より効率的な資源利用が可能であることを明らかになった。

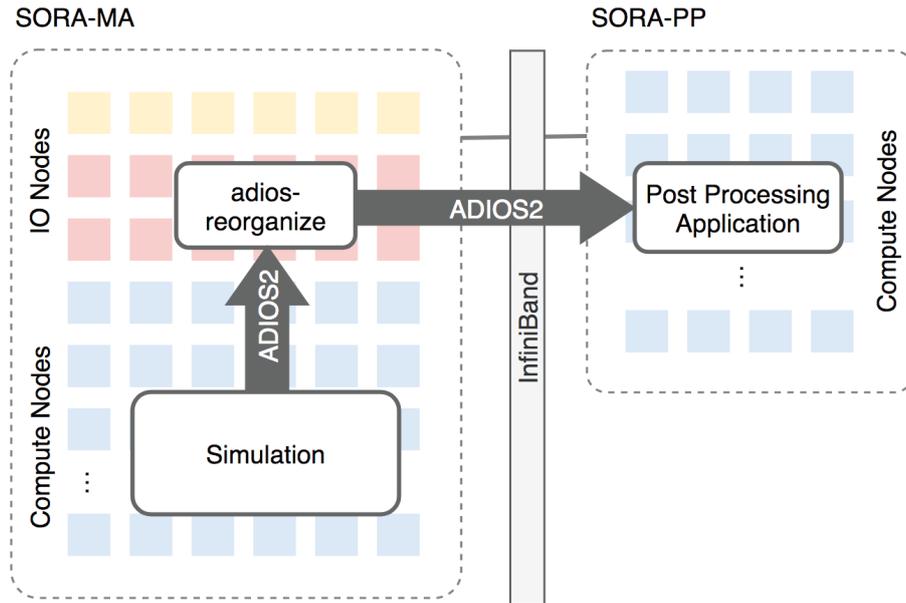


図 1: ADIOS2 による SORA-MA/SORA-PP 間のデータステージング通信

● 成果の公表

なし

● JSS2 利用状況

● 計算情報

プロセス並列手法	MPI
スレッド並列手法	非該当
プロセス並列数	1 - 128
1 ケースあたりの経過時間	5 分

● 利用量

総資源に占める利用割合※1 (%) : 0.00

内訳

計算資源		
計算システム名	コア時間(コア・h)	資源の利用割合※2 (%)
SORA-MA	261.56	0.00
SORA-PP	366.46	0.00
SORA-LM	0.00	0.00
SORA-TPP	0.00	0.00

ファイルシステム資源		
ファイルシステム名	ストレージ割当量(GiB)	資源の利用割合※2 (%)
/home	9.54	0.01
/data	95.37	0.00
/ltmp	1,953.13	0.17

アーカイバ資源		
アーカイバシステム名	利用量(TiB)	資源の利用割合※2 (%)
J-SPACE	0.00	0.00

※1 総資源に占める利用割合 : 3つの資源(計算,ファイルシステム,アーカイバ)の利用割合の加重平均

※2 資源の利用割合 : 対象資源一年間の総利用量に対する利用割合