

Study on Data Staging Technique for Efficient Post-Processing in Large-Scale CFD Computation

Report Number: R18EACA42

Subject Category: JSS2 Inter-University Research

URL: <https://www.jss.jaxa.jp/en/ar/e2018/9110/>

● Responsible Representative

Shinji Shimojo, Director, Cybermedia Center, Osaka University

● Contact Information

Shinji Shimojo (shimojo@cmc.osaka-u.ac.jp)

● Members

Shinji Shimojo, Keichi Takahashi

● Abstract

Conventional post-processing of CFD simulations was achieved by saving the entire simulation output on a parallel file system and then processing the output. However, this approach is becoming increasingly challenging due to the limitations in storage size and IO bandwidth. Therefore, data staging, where the simulator transfers its output to a post-processing application during runtime, is attracting attention. In this research, we evaluate the feasibility of leveraging data staging technologies on HPC environments exemplified JSS2 and analyze the requirements for data staging middleware and HPC environment.

● Reasons for using JSS2

We used JSS2 because it allows communication between its main compute system (SORA-MA) and pre/post-processing system (SORA-PP).

● Achievements of the Year

We assumed that a CFD simulation and a post-processing application are executed on the SORA-MA and SORA-PP system, respectively. As for the data staging middleware, we used Adaptive IO System (ADIOS2), a data movement middleware developed at the Oak Ridge National Laboratory (ORNL). First, we confirmed that staging communication using ADIOS2 is possible within MA and PP, respectively. Next, we tried to directly communicate from a compute node in MA to a compute node in PP, but this attempt failed. This problem was due to the configuration of the interconnect that did not route packets from one system to another.

To solve this problem and establish staging communication between the MA and PP systems, we ran a proxy (adios-reorganize) that bridges the two systems on a compute node in MA that connects to PP (IO node). As a result, we successfully demonstrated that staging communication from MA to PP is feasible.

Throughout our evaluation on JSS2, we found some issues in ADIOS2 with its SPARC64 XIfx processor support and communication across heterogeneous systems. We addressed these issues by collaborating with the ADIOS2

developers at ORNL. In addition, we experienced difficulties in obtaining an IO node because the job scheduler deployed on JSS2 does not allow users to request IO nodes explicitly.

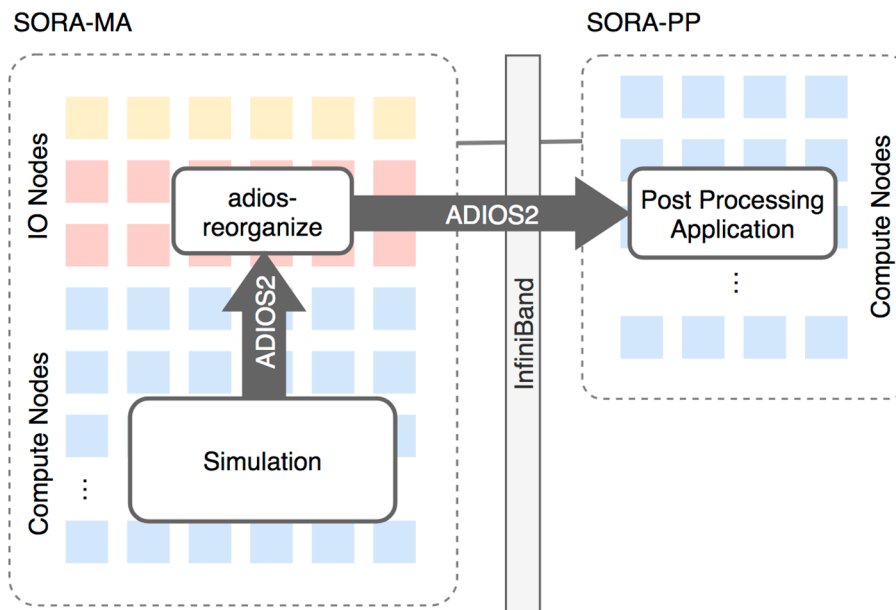


Fig. 1: Data staging between SORA-MA and SORA-PP using ADIOS2

● **Publications**

N/A

● **Usage of JSS2**

● **Computational Information**

Process Parallelization Methods	MPI
Thread Parallelization Methods	N/A
Number of Processes	1 - 128
Elapsed Time per Case	5 Minute (s)

● **Resources Used**

Fraction of Usage in Total Resources*1 (%): 0.00

Details

Computational Resources		
System Name	Amount of Core Time (core x hours)	Fraction of Usage*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

File System Resources		
File System Name	Storage Assigned (GiB)	Fraction of Usage*2 (%)
/home	9.54	0.01
/data	95.37	0.00
/tmp	1,953.13	0.17

Archiver Resources		
Archiver Name	Storage Used (TiB)	Fraction of Usage*2 (%)
J-SPACE	0.00	0.00

*1: Fraction of Usage in Total Resources: Weighted average of three resource types (Computing, File System, and Archiver).

*2: Fraction of Usage: Percentage of usage relative to each resource used in one year.