MPI communication optimization for CFD application on supercomputer with complex node connection topology

Report Number: R24EFHC0313
Subject Category: Common Business
URL: https://www.jss.jaxa.jp/en/ar/e2024/26610/

Responsible Representative

Megumi Kamata, Security and Information Systems Department Supercomputer Division

Contact Information

Megumi Kamata, Yoimi Kojima(kamata.megumi@jaxa.jp,kojima.yoimi@jaxa.jp)

Members

Megumi Kamata, Yoimi Kojima

Abstract

Computational Fluid Dynamics (CFD) is an essential technology for aerospace engineering and major application of the high-performance computing. Most of CFD software employs MPI for parallelization. In large-scale supercomputers, the computational nodes which are allocated for CFD can be scattered over the interconnect and the communication between faraway nodes suffers parallel efficiency of the CFD software.Some modern supercomputer systems such as Fugaku and JSS3 (JAXA supercomputer system generation 3) employ advanced interconnect (Tofu interconnect) to provide efficient and high-throughput communication between computation nodes. However, due to the high complexity of the network topology, it is difficult to achieve optimal MPI communication if we rely only on the automatic rank allocation which has no connection with the data structure of CFD application.

In this study, we aim to achieve optimal MPI communication for CFD application on supercomputer system with complex node connection topology, such as Tofu interconnect..We implement MPI rank reordering function to a CFD solver developed by JAXA and evaluate its effect.

Reasons and benefits of using JAXA Supercomputer System

This study aims to improve the performance of CFD applications on supercomputers. Since it is necessary to verify the performance by running the program on an actual supercomputer, a supercomputer was utilized for this purpose.

Achievements of the Year

We implemented MPI rank optimization method for CFD application and evaluated its effect on supercomputer with complex node connection topology. The results indicates that the optimization could improve strong scaling in certain situation.

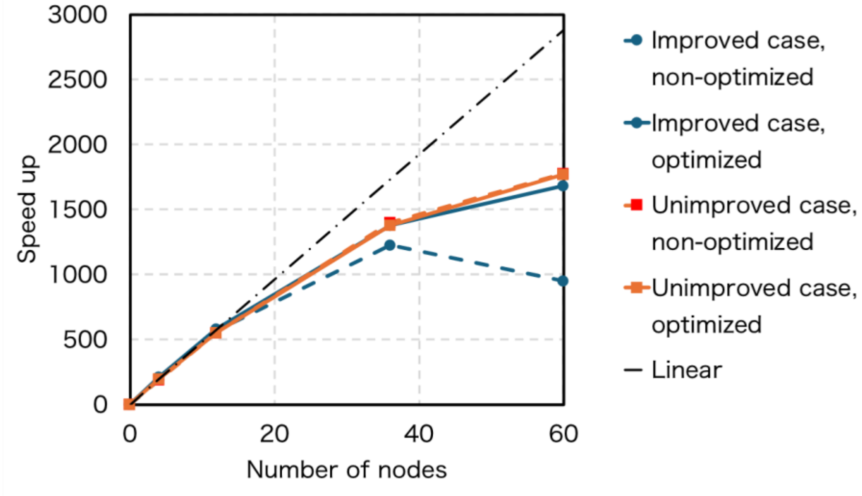

Fig. 1: The figure shows the improved case (blue curves) which shows significant improvement of the strong scaling when the MPI rank optimization is applied. On the other hand, in the unimproved case, the strong scaling is not improved regardless if it is optimized or not

Publications

- Other

Presented at JAXA booth in SC24

Usage of JSS

- **Computational Information**

Process Parallelization Methods	MPI
Thread Parallelization Methods	N/A
Number of Processes	48 - 2880
Elapsed Time per Case	30 Minute(s)

- **JSS3 Resources Used**

Fraction of Usage in Total Resources*1(%): 0.11

Details

Computational Resources		
System Name	CPU Resources Used (core x hours)	Fraction of Usage*2(%)
TOKI-SORA	3,028,682.76	0.14
TOKI-ST	3.99	0.00
TOKI-GP	0.00	0.00
TOKI-XM	0.00	0.00
TOKI-LM	0.00	0.00
TOKI-TST	0.00	0.00
TOKI-TGP	0.00	0.00
TOKI-TLM	0.00	0.00

File System Resources		
File System Name	Storage Assigned (GiB)	Fraction of Usage[*2] (%)
/home	84.58	0.06
/data and /data2	8,604.29	0.04
/ssd	726.55	0.04

Archiver Resources		
Archiver Name	Storage Used (TiB)	Fraction of Usage[*2] (%)
J-SPACE	17.59	0.06

[*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.

- **ISV Software Licenses Used**

ISV Software Licenses Resources		
	ISV Software Licenses Used (Hours)	Fraction of Usage[*2] (%)
ISV Software Licenses (Total)	0.00	0.00

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