Scientific Satellite Data Processing Using JSS3

● **Responsible Representative**

Iku SHONOHARA, Unit Director, Science Satellite Operation and Data Archive Unit, ISAS

● **Contact Information**

NAKAHIRA, Satoshi(nakahira.satoshi@jaxa.jp)

● **Members**

Satoshi Nakahira, Ken Ebisawa, Eiji Honjoh

● **Abstract**

The data size of space science data has been growing due to the increasing performance of instruments and longer observation time periods, and it now takes several months to recreate data due to e.g. calibration changes. Therefore, we attempted to implement fast data processing using JSS.

● **Reasons and benefits of using JAXA Supercomputer System**

The environment is ready to execute parallel processing without building a cluster environment on its own, and is expected to have IO performance that does not easily become a bottleneck even if many data read processes are run for a single input data set.

● **Achievements of the Year**

Accumlated data of Monitor of All-sky X-ray Image (MAXI) from August 2009 to February 2022, 12 years and 7 months, were reprocessed in a batch. Parallelization was performed by converting the daily input data into a daily data set, and then sequentially submitting the data set to the system so that the number of concurrent executions would be less than 200. As a result, the process of converting 11 TB of input data into 20 TB of output data took about 17 hours, and the time required to process one day's worth of data was not significantly different from the time required by a regular data processing server, indicating that the speed can be linearly improved.

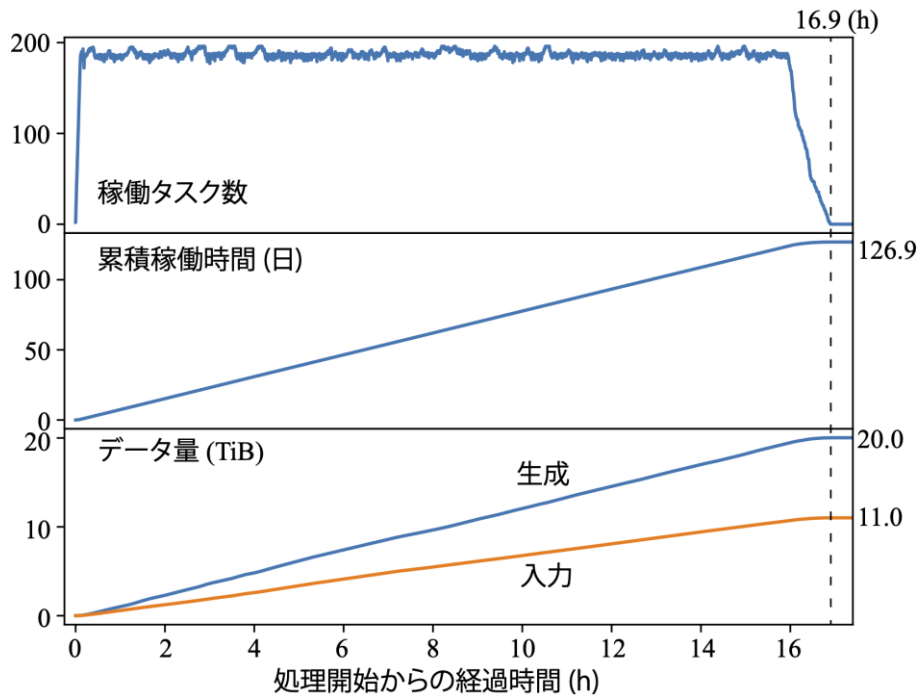Translated with www.DeepL.com/Translator (free version)

Fig. 1: The horizontal axis is the elapsed time (hour) from the start of processing, and the vertical axis is the number of concurrent jobs, cumulative sum of processing time, and total generated data volume, respectively.

● **Publications**

- Oral Presentations

Space Science Data Analysis Simposium, Feb. 18, 2022, "Scientific Satellite Data Processing and Data Analysis Using JSS3"

● **Usage of JSS**

● **Computational Information**

Process Parallelization Methods	Since the data conversion corresponds to one day's input data to one day's output data, the number of jobs is checked and submitted with the jxsub command while adjusting the interval.
Thread Parallelization Methods	N/A
Number of Processes	1 - 195
Elapsed Time per Case	16.9 Hour(s)

● **JSS3 Resources Used**

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

Details

Computational Resources		
System Name	CPU Resources Used (core x hours)	Fraction of Usage[*2](%)
TOKI-SORA	43.10	0.00
TOKI-ST	34,276.33	0.04
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	26.67	0.03
/data and /data2	17,433.33	0.19
/ssd	816.00	0.21

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.

- **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.

- **ISV Software Licenses Used**