

Designing a monitoring system for and collection statistics for HPCI shared storage

Hidetomo Kaneyama
Riken Center for Computational Science
Japan
hidetomo.kaneyama@riken.jp

Hiroshi Harada
Riken Center for Computational Science
Japan
hiroshi.harada@riken.jp

Asuka Ohta
PARARS
Japan
ask@parars.riken.jp

Hiroyuki Nakatsuka
Metahack Japan
Japan
hiroyu-n@metahack.jp

1. INTRODUCTION

The HPCI Shared Storage is a large-scale network storage system for the High Performance Computing Infrastructure (HPCI), which enables researchers from Japanese institutions to access and utilize supercomputers. HPCI Shared Storage offers 50 PB of logical storage space for the storage and management of research data generated by HPCI supercomputers in Japan, as of January 2023.

To ensure stable operation of this storage system, the monitoring environment has been updated since 2020 to enhance information collection and update the automatic alerting system. We have also developed a mechanism to provide user usage information for each user and group. This monitoring environment has enhanced the collection of statistical information and has been able to collect stored data information and other information at a high frequency.

The statistical information shows that there are many files that have not been accessed for more than 3years (dark-files). These dark-files may have already been forgotten due to members leaving the research group. Our goals is to encourage users to delete or use such data. Some users may also be storing large amounts of small data. The data transfer of a large number of small files may cause transfer delays due to metadata access load and other factors. Also, for huge data, replication, etc., may take a long time to read, and the data may remain unused after writing. To address this issue for users, we are currently implementing compression and split transfers. This enhanced collection of statistical information promotes active use and improvement.

2. Service Monitoring System

The service monitoring is shown in Fig.1.HPCI Shared Storage utilizing a network file system known as Gfarm[1]. Gfarm's metadata is stored in a PostgreSQL database, from which detailed user information is directly retrieved and made available to users through the using of Grafana. For added security, certain user data is segregated into separate tables within TimescaleDB. The service monitoring authentication is used via Shibboleth in accordance with HPCI. Additionally, an exporter utilizing Gfarm's API is utilized to monitor operational information, access status, and service availability. alive monitoring and port checks are conducted using node_exporter and blackbox_exporter respectively. In the event of any failures, alerts are promptly dispatched through slack, email and automatic phone notifications via Twilio.

3. Data statistics for HPCI shared storage

The access status of files stored within the HPCI shared storage is presented in Table 1. As of 2022, user usage is robust, with a more 50% of all number of files being access between 2021 and 2022. HPCI Shared storage is currently experiencing a lot of access from Fugaku, so it is that active use of Fugaku is driving shared storage usage. However, a significant number of files, about 20%, have not been accessed in over 3 years. Given that this information is obtained on a per-group or per-user, our goals is to provide similar

information to users and groups to encourage delete dark-files and optimize the utilization of storage space

Table 2 displays information on the sizes of files stored within the HPCI shared storage. The prevalence of under 10MB files is more 67%, and these files often impede the transfer speed due to the increased metadata load during I/O. Given that these files are infrequently accessed, it is crucial to implement efficient transfer methods, such as compression. To address this issue, we are developing a compression transfer function on the client side.

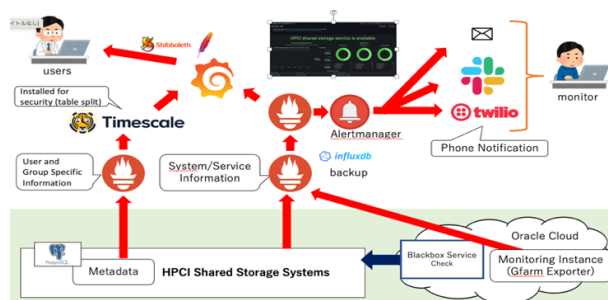


Fig 1. Service Monitoring System

Access Period	File Space		File Number	
	Value	Rate	Value	Rate
~ 3month	4.08PB	19.74%	29M	15.68%
3m ~ 1year	8.08PB	39.09%	76M	41.72%
1y ~ 3y	4.27PB	20.66%	44M	24.23%
3y ~ 5y	3.19PB	15.43%	18M	9.82%
5y ~	1.05PB	5.08%	16M	8.56%

Table 1. Capacity and number of files for storage by access period

FileSize	File Space		File Number	
	Value	Rate	Value	Rate
~ 10MB	0.21PB	1.00%	119 M	67.55%
10MB~100MB	0.87PB	4.16%	26 M	14.94%
100MB~1GB	6.25PB	29.89%	26 M	14.91%
1GB ~ 10GB	7.25PB	34.68%	4 M	2.49%
10GB~100GB	3.31PB	15.83%	177 K	0.10 %
100GB~	3.02PB	14.43%	8579	0.00 %

Table 2. Capacity for storage and number of files by file size

REFERENCES

[1] Osamu T. Gfarm Grid File System. 2010.7 NEW GENERATION COMPUTING SPRINGER