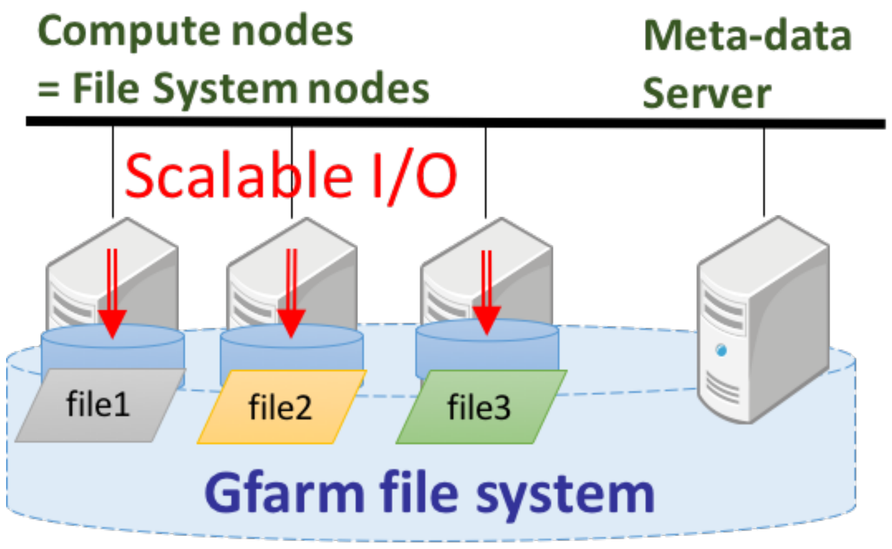
Software Researches for Big Data and Extreme-Scale Computing

Gfarm: a High Performance Distributed File System for Supercomputing [1,2]

http://oss-tsukuba.org/en/software/gfarm

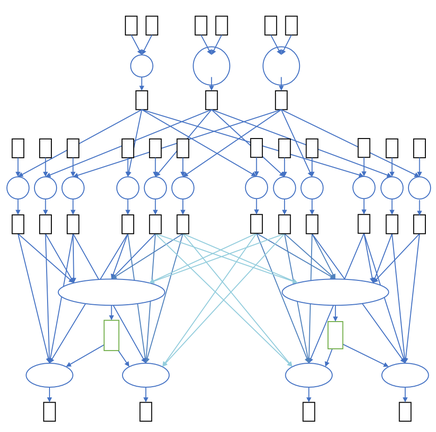


It is a Node-local burst buffer and a Grid file system.
Features include

- Open source
- Exploit local storage, and data locality for scalable I/O performance
- No single point of failure
- MapReduce, MPI-IO, Pwrake workflow system, Batch queuing system with data locality enhancement
- InfiniBand support
- Data integrity is supported for silent data corruption
- 20,000 downloads since March 2007
- Production systems: 8PB JLDG, 100PB HPCI Storage, etc.

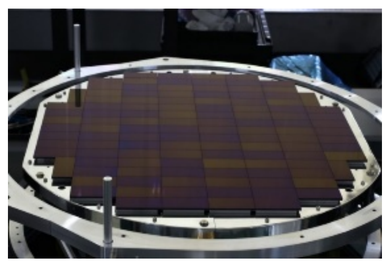
Applying Pwrake Workflow System and Gfarm File System to Telescope Data Processing [3]

DAG of HSC pipeline



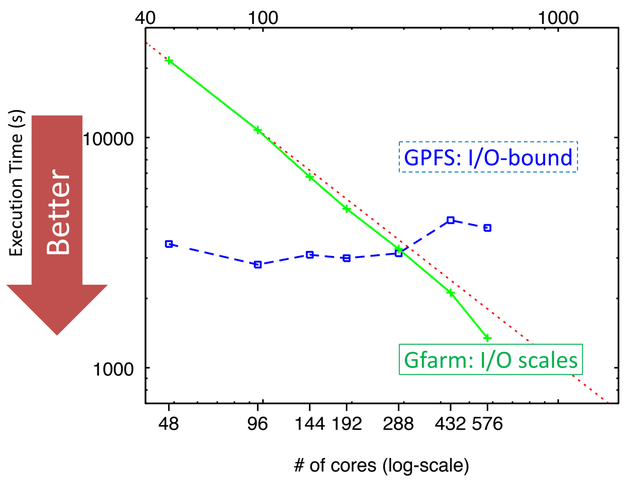
Subaru Telescope
Image Credit: NAOJ

Hyper Suprime-Cam
Image Credit: NAOJ



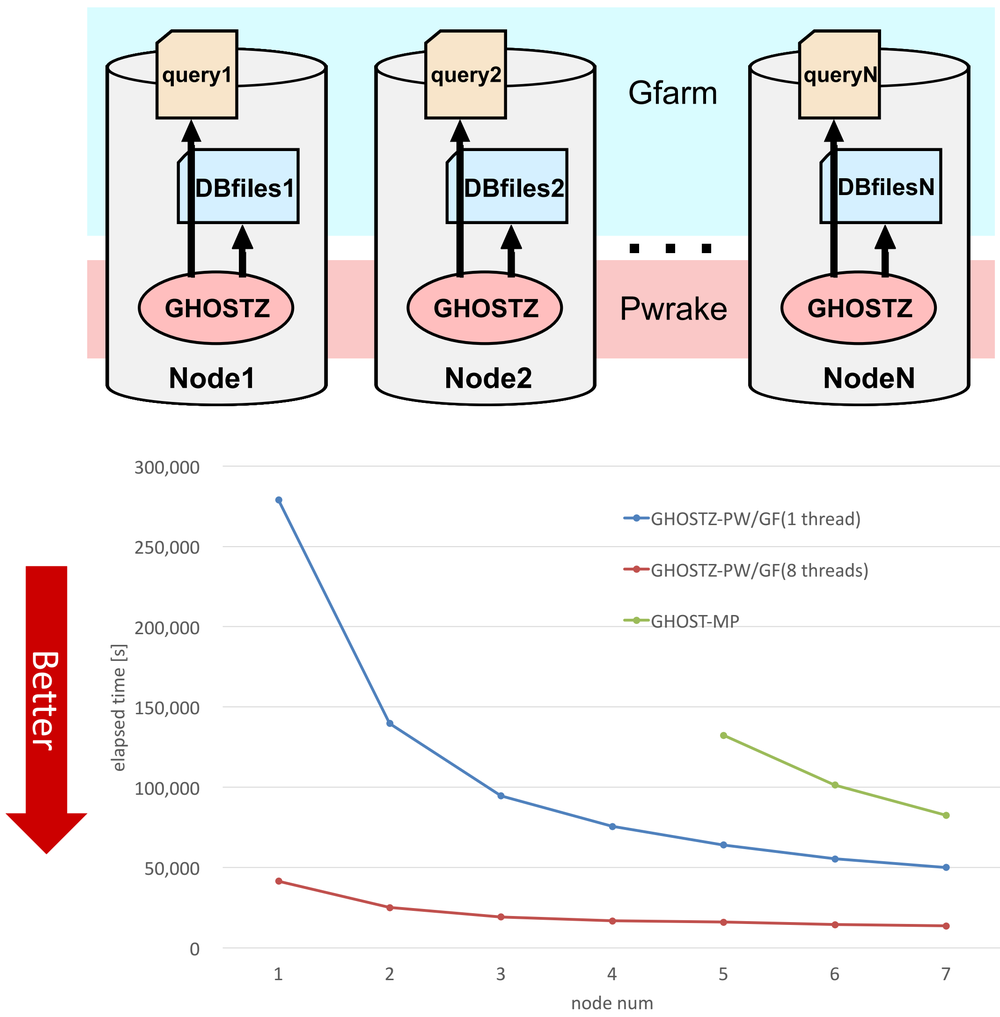
Hyper Suprime-Cam (HSC) mounted on the Subaru telescope generates 300GB/night data. To speed up the pipeline processing of HSC data, we introduced **Gfarm file system** for scalable I/O and **Pwrake workflow system** for efficient core utilization. Measurement using 576 cores shows that our method improved the performance of the pipeline by a factor of 2.2 from the original method.

Scalability of I/O-bound task



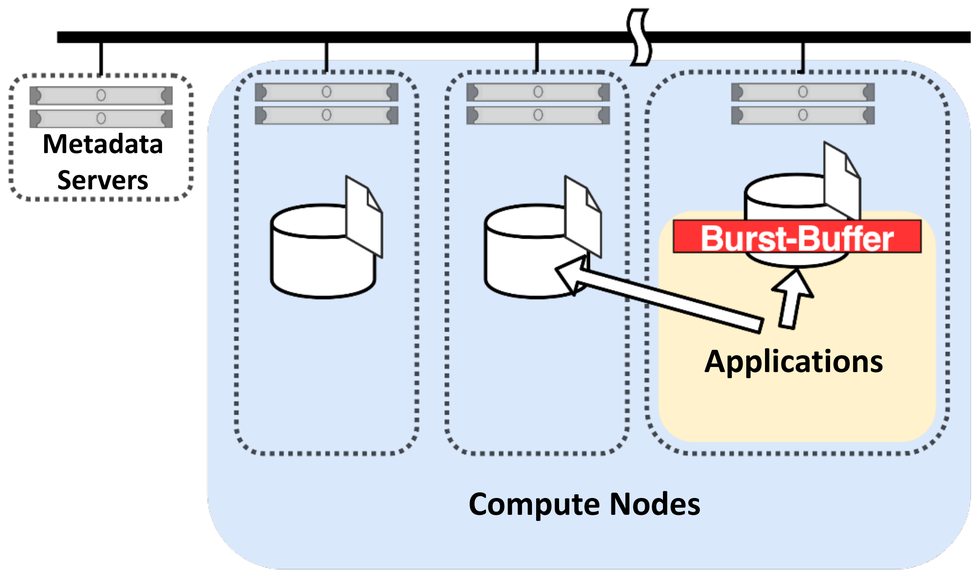
Pwrake: https://github.com/masa16/pwrake

Distributed and parallel homology search system with Gfarm and Pwrake [4]



In recent years, acquired genomic data size is rapidly increasing, which results in memory shortage in a cluster node and long execution time. GHOSTZ with Pwrake/Gfarm solves these issues using node-local storage and locality aware workflow execution. It outperforms GHOST-MP 6x times in an alignment step when using 7 nodes.

Accelerating Distributed File System with client node-local Burst Buffer



Node-local burst buffer is expected to be a promising caching layer for a distributed file system. To design a node-local burst buffer, we evaluate current solutions including Gfarm/BB, BeeOND, and Octopus using NVMe SSD and NVRAM.

Reference
[1] Osamu Tatebe, Kohei Hiraga, Noriyuki Soda, "Gfarm Grid File System," New Generation Computing, Ohmsha, Ltd. and Springer, Vol. 28, No. 3, pp.257-275, 2010.
[2] Gfarm File System, http://oss-tsukuba.org/en/software/gfarm
[3] Masahiro Tanaka, Osamu Tatebe, Hideyuki Kawashima: "Applying Pwrake Workflow System and Gfarm File System to Telescope Data Processing", Proceedings of 2018 IEEE International Conference on Cluster Computing (CLUSTER), pp. 113-122, 2018
[4] Kenta Machida, Osamu Tatebe: "Distributed and Parallel Homology Search System with Gfarm/Pwrake," IPSJ SIG Technical Reports, Vol. 2017-HPC-162, No. 10, 9 pages, 2017

Acknowledgment
This work is partially supported by the JST CREST Grant Numbers JPMJCR1303 and JPMJCR1414, and JSPS KAKENHI Grant Number JP17H01748.

contact address: pr@ccs.tsukuba.ac.jp