

## An Examination of Data I/O Speed on a Parallel Data Storage System

MURATA, Ken T.<sup>1\*</sup>; WATANABE, Hidenobu<sup>1</sup>; UKAWA, Kentaro<sup>2</sup>; MURANAGA, Kazuya<sup>2</sup>; YUTAKA, Suzuki<sup>2</sup>; TATEBE, Osamu<sup>3</sup>; TANAKA, Masahiro<sup>3</sup>; KIMURA, Eizen<sup>4</sup>

<sup>1</sup>National Institute of Information and Communications Technology, <sup>2</sup>Systems Engineering Consultants Co., LTD., <sup>3</sup>University of Tsukuba, <sup>4</sup>Ehime University

This paper is to propose a cloud system for science, which has been developed at NICT (National Institute of Information and Communications Technology), Japan. The NICT science cloud is an open cloud system for scientists who are going to carry out their informatics studies for their own science.

The NICT science cloud is not for simple uses. Many functions are expected to the science cloud; such as data standardization, data collection and crawling, large and distributed data storage system, security and reliability, database and meta-database, data stewardship, long-term data preservation, data rescue and preservation, data mining, parallel processing, data publication and provision, semantic web, 3D and 4D visualization, out-reach and in-reach, and capacity buildings.

In the present study, we examine performance of parallelization of I/O on the NICT Science Cloud system. We examine two types of data file system; parallel file system (GPFS) and distributed file system (Gfarm). The later file system shows a tremendous fast I/O, as fast as 23 GB/sec using only 30 servers. We should pay attention to this I/O speed (23GB/sec is 184 Gbps) from the viewpoint of network speed. Since general network speed so far is 10 Gbps or 40 Gbps in a cloud system, this 184 Gbps is fast enough that the I/O cannot be a bottle-neck for big data processing.

We also discuss that the distributed file system shows better scalability compared with the GPSF system. Parallelization efficiency in the present examination is higher than 90% in case of parallel file system. This suggests that, in the near future, we can expect higher I/O speed using more file servers. For instance, if the I/O speed is as fast as 100 GB/sec, it takes only 1,000 sec. (17 min.) to read 100 TB data files.

