Japan Geoscience Union Meeting 2014
(28 April - 02 May 2014 at Pacifico YOKOHAMA, Kanagawa, Japan)

STT60-09 Room:211 Time:April 30 16:30-16:45

Techniques of Big-Data Processing on the NICT Science Cloud

MURATA, Ken T.[1*] ; WATANABE, Hidenobu[1] ; UKAWA, Kentaro[2] ; MURANAGA, Kazuya[2] ; YUTAKA, Suzuki[2] ; TATEBE, Osamu[3] ; TANAKA, Masahiro[3] ; KIMURA, Eizen[4]

[1]NICT, [2]Systems Engineering Consultants Co., LTD., [3]University of Tsukuba, [4]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 an I/O performance of data file system; distributed file system (Gfarm). The Gfarm 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. We also discuss that the distributed file system shows high scalability: Parallelization efficiency in the present examination is higher than 90% in case of parallel file system. We finally discuss high-performance data processing on the NICT Science Cloud. We have already archived several examples using our technique for both Earth and Space observation data and simulation data. The speed up of the data processing is more than 60 times for scientific big-data.