Real-time Data Processing via Gfarm/Pwrake and Applications on the NICT Science Cloud

*Kazuya Muranaga[1], Ken T. Murata[2], Kazunori Yamamoto[2], Yoshiaki Nagaya[2], Kentaro Ukawa[1], Junichi Murayama[1], Yutaka Suzuki[1], Osamu Tatebe[3], Masahiro Tanaka[3], Eizen Kimura[4]

1.Systems Engineering Consultants Co., LTD., 2.National Institute of Information and Communications Technology, 3.University of Tsukuba, 4.Department of Medical Informatics Ehime Univ.

This paper is to present a technology to perform parallel data processing on parallel storage system, 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, Gfarm/Pwrake is used for the big-data processing. The Gfarm is one of the parallel storages system developed in Tsukuba University. The Pwrake is a task scheduler designed for the Gfarm. With combination of two technologies, real-time data processing is easily constructed on the NICT Science Cloud. Several examples to use this Gfarm/Pwrake real-time data processing are discussed for meteorological satellite data processing, weather radar data processing, and other real-time data processing's.