Geophysical Research Abstracts
Vol. 15, EGU2013-13425, 2013
EGU General Assembly 2013

A Virtual Remote Storage and Data Visualization on the Science Cloud

Hidenobu Watanabe and Ken T. Murata
National Institute of Information and Communications Technology (NICT)

Super-computer is now one of the most important tools for scientific research works. Computational power of super-computers is getting higher and higher. Along with the development of its computational power, data size yielded by super-computers is also getting larger. Typical data size by one job is already in TB order. For near future works, we need to prepare PB scale data analysis.
Recently, big-data science is paid attention because of such large-scale data is popular. However, since data analysis environment to process such large scale data is not established in any science cloud or academic cloud, big-data science is not successful yet. One of the reasons of this difficulty is techniques to support researchers working on cloud systems, including middle-wares.
For example, we save many data files with large-scale on cloud storages. However, reading these files take too long time. Assume that we have 10TB data files on a normal NAS storage. It takes more than 10 days to read all of the data files. This is because of the low speed disk I/O between hard-disk drive and controlling server.
We have been developing a technique of a virtual storage system with high speed I/O using parallel storage system named Gfarm (Grid farmware). With combining parallel disk servers and making a virtual storage, we achieved as high as 7Gbps storage. What should be noted is that the 7Gbps is not network speed, but disk I/O speed. Since disk I/O speed of a local SSD storage is about 4Gbps, this suggests that our virtual storage provide us with higher I/O speed than local disks.
After saving or reading data files, we next analyze data files. The virtual disk system should be synchronized these data analysis. In the present talk, we are going to present the sequential data processing (visualization) using the virtual high speed storage system.