Japan Geoscience Union Meeting 2012
(May 20-25 2012 at Makuhari, Chiba, Japan)

U01-14 Room:105 Time:May 23 15:30-15:50

Toward a Big Data Science: A challenge of Science Cloud

MURATA, Ken T.[1*], WATARI, Shinichi[1], NAGATSUMA, Tsutomu[1], KUNITAKE, Manabu[1], WATANABE, Hidenobu[1], YAMAMOTO, Kazunori[1], KUBOTA, Yasubumi[1], MURAYAMA, Yasuhiro[1], KATO, Hisao[1], TSUGAWA, Takuya[1], SHINAGAWA, Hiroyuki[1], JIN, Hidekatsu[1], TANAKA, Takashi[2], SAKAGUCHI, Kaori[1], SAITO, Shinji[1], NISHIOKA, Michi[1], ISHIBASHI, Hiromitsu[1]

[1]NICT, [2]Kyushu University

During these 50 years, along with appearance and development of high-performance computers (and super-computers), numerical simulation is considered to be a third methodology for science, following theoretical (first) and experimental and/or observational (second) approaches. The variety of data yielded by the second approaches has been getting more and more. It is due to the progress of technologies of experiments and observations. The amount of the data generated by the third methodologies has been getting larger and larger. It is because of tremendous development and programming techniques of super computers.

Most of the data files created by both experiments/observations and numerical simulations are saved in digital formats and analyzed on computers. The researchers (domain experts) are interested in not only how to make experiments and/or observations or perform numerical simulations, but what information (new findings) to extract from the data. However, data does not usually tell anything about the science; sciences are implicitly hidden in the data. Researchers have to extract information to find new sciences from the data files. This is a basic concept of data intensive (data oriented) science for Big Data.

As the scales of experiments and/or observations and numerical simulations get larger, new techniques and facilities are required to extract information from a large amount of data files. The technique is called as informatics as a fourth methodology for new sciences.

Any methodologies must work on their facilities: for example, space environment are observed via spacecraft and numerical simulations are performed on super-computers, respectively in space science. The facility of the informatics, which deals with large-scale data, is a computational cloud system for science.

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

The 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.

Figure is a schematic picture of the NICT science cloud. Both types of data from observation and simulation are stored in the storage system in the science cloud. It should be noted that there are two types of data in observation. One is from archive site out of the cloud: this is a data to be downloaded through the Internet to the cloud. The other one is data from the equipment directly connected to the science cloud. They are often called as sensor clouds.

In the present talk, we first introduce the NICT science cloud. We next demonstrate the efficiency of the science cloud, showing several scientific results which we achieved with this cloud system. Through the discussions and demonstrations, the potential performance of sciences cloud will be revealed for any research fields.

Keywords: Big Data, Science Cloud, OneSpaceNet