Data-intensive Computing in the GeoBrain Web Service Environment

Liping Di, Peisheng Zhao, Meixia Deng
Center for Spatial Information Science and Systems, George Mason University
4400 University Drive, MSN 6E1
Fairfax, VA 22030
ldi@gmu.edu

Earth observation, mainly through satellite remote sensing, has collected huge amount of data about Earth and its environment. The data have great scientific and socio-economic values. However, those data have to be processed and analyzed before the value is realized in applications. The processing and analysis of the data are both data and computationally intensive and large amount of computing resources are needed, which are only affordable by a few well-funded scientists or institutions. As such, many of the collected data are severely underutilized and the potentials of the societal benefits of the Earth observation are not fully realized. In the past decade, the Center for Spatial Information Science and Systems has spent significant amount efforts to develop the Web service technology and systems, such as GeoBrain, to solve this problem. GeoBrain is a geospatial service, modeling, and knowledge building system based on Service Oriented Architecture (SOA). It has adopted open standards from the Open Geospatial Consortium (OGC), the World Wide Web Consortium (W3C), the Organization for the Advancement of Structured Information Standards (OASIS), and the International Organization for Standardization (ISO). GeoBrain allows world-wide users to discover, access, and analyze peta-bytes of Earth observation data in the distributed repositories. It also provides an open geospatial knowledge environment through its portal system for capturing, discovering, preserving, sharing, and disseminating geospatial knowledge. GeoBrain has been operational since 2005 although the development has been continued. This presentation will discuss the issues on conducting data-intensive computing in the Web service environment, the implementation of data-intensive computing capabilities in GeoBrain, and the efforts to migrate such capabilities from the cluster computing environment to the Grid and later on to Cloud ones.