Chapter XXVIII
Sharing of Distributed Geospatial Data through Grid Technology

Yaxing Wei
George Mason University, USA

Liping Di
George Mason University, USA

Guangxuan Liao
University of Science and Technology China, China

Baohua Zhao
University of Science and Technology China, China

Aijun Chen
George Mason University, USA

Yuqi Bai
George Mason University, USA

ABSTRACT

With the rapid accumulation of geospatial data and the advancement of geoscience, there is a critical requirement for an infrastructure that can integrate large-scale, heterogeneous, and distributed storage systems for the sharing of geospatial data within multiple user communities. This article probes into the feasibility to share distributed geospatial data through Grid computing technology by introducing several major issues (including system heterogeneity, uniform mechanism to publish and discover geospatial data, performance, and security) to be faced by geospatial data sharing and how Grid technology can help to solve these issues. Some recent research efforts, such as ESG and the Data Grid system in GMU CSISS, have proven that Grid technology provides a large-scale infrastructure which can seamlessly integrate dispersed geospatial data together and provide uniform and efficient ways to access the data.

INTRODUCTION

With the advancement of computing and network technologies, geospatial applications have become more and more important in both academic and commercial areas (Lo and Yeung, 2002). Geospatial applications focus on geospatial data, such as remote sensing data and survey data. Huge quantities of raw geospatial data are processed with geospatial algorithms to generate high-level data that contain information which can be further used to extract knowledge for people to better understand the earth. The extent of usefulness of geospatial data and applications has been proven across many diverse areas and disciplines, such as meteorology, global climate change, agriculture, forestry, flood monitoring, fire detection and monitoring, and geology (Lamberti and Beco, 2002). This extent is still expanding to make geospatial data and applications close to people's everyday life, such as the success of Google Earth and the popularity of mobile Global Positioning Systems (GPS).

In recent years, there were more and more satellites launched for diverse purposes. Currently there are hundreds of living earth observation satellites inspecting the earth and continuously collecting tremendous amounts of data. For example, the Earth Observation System (EOS) project of National Aeronautics and Space Administration (NASA) is producing global observation data of the land surface, biosphere, solid Earth, atmosphere, and oceans at the speed of more than 3TB/day and archiving this data into 9 data centers distributed across the United States. This exponential accumulation of geospatial data presents new challenges for effective and efficient use of the data.

Since the mid-1990s, with the exploding growth of the Internet, the focus of computing has shifted from stand-alone and locally networked environments to wide-scale, distributed, and heterogeneous computing infrastructures (Karimi and Peachavanish, 2005). Those changes enabled and compelled new ways of using geospatial data and applications. With the advancement of geoscience, more and more complex geospatial algorithms involving geospatial data from multiple sources and domains are designed. Contrary to their past monolithic design and implementation, current computing trends suggest new geospatial applications will be distributed and used in heterogeneous network environments. The capabilities to efficiently access and share the tremendous amount of distributed geospatial data are crucial to geospatial applications. Consequently, there is a need for a large-scale infrastructure which can seamlessly integrate dispersed geospatial data together and provide uniform and efficient ways to access the data. Fortunately, recent advancements of computing technologies, especially the emerging Grid technology, can make this a reality.

BACKGROUND

In the past, geospatial applications were mostly designed for a single workstation or supercomputer. The geospatial data they need to process were limited to a single storage system or locally networked storage systems. The generated high-level geospatial data and information were also difficult to be shared by other geospatial user communities due to the isolation and heterogeneity of computing platforms and storage systems. Today's complex geospatial problems need applications that can analyze large quantities of geospatial data coming from different sources which were isolated from each other in the past. For example, a statistical wildfire forecast model at 8-km spatial resolution in the conterminous USA used over 1 terabytes of data obtained from different sources, including data derived from measurements of the NASA Earth Observing System satellites and daily weather data provided by National Oceanic and Atmospheric Administration (NOAA)