



Data Grid Services for National Digital Archives Program in Taiwan

Wednesday, 1 March 2006 14:30 (15 minutes)

Digital archives/libraries are widely recognized as a crucial component of the global information infrastructure for the new century. Research and development projects in many parts of the world are concerned about using advanced information technologies for managing and manipulating digital information, ranging from data storage, preservation, indexing, searching, presentation, and dissemination capabilities to organizing and sharing of information over networks.

Digital Archive demands for reliable storage systems for persistent digital objects, well-organized information structure for effective content management, efficient and accurate information retrieval mechanism and flexible services for varying users needs. Hundreds of Petabyte of digital information has been created and dispersed all over the internet since computers had been used for information processing, and the amount still grows in the rate of tens of Petabyte per year. Grid technology offers a possible solution for aggregating and processing diversified heterogeneous Petabyte scale digital archives. Metadata-based information representation makes specific and relative information retrieval more accurately, makes information resources interoperable, and paves the way for formal knowledge discovery. Taking advantage of advancing IT, semantic level information indexing, categorizing, analyzing, tracking, retrieving and correlating could be implemented. Data Grid aims to set up a computational and data-intensive grid of resources for data analysis. It requires coordinated resource sharing, collaborative processing and analyzing on huge amounts of data produced and stored by many institutions.

In Taiwan, a National Digital Archive Project (NDAP) was initiated in 2002 with its pilot phase started in 2001. According to the record in 2005, more than 60 Terabytes digital objects was generated and archived by 9 major content holders in Taiwan. Not only delicate and gracious Chinese cultural assets can be preserved and made available via the Internet, but this approach could be proposed as a new paradigm of academic researches based on digital and integrated information resources. The design and implementation phase is ongoing and we would like to illustrate in the EGEE User Forum.

Academia SINICA Grid Computing Centre (ASGC) is in charge of building a new generation of Grid-based research infrastructure in Academia SINICA and in Taiwan based on EGEE and OSG as the Grid middleware. This infrastructure is a major component for the development and the deployment of the National Digital Archive Project (NDAP) providing long-term preservation of the digital contents and unified data access. These services will be built upon the e-Science infrastructure of Taiwan. The Storage Resource Broker (SRB) developed at SDSC, is a Middleware which enables scientists to create, manage and collaborate with flexible, unified "virtual data collections" that may be stored on heterogeneous data resources distributed across a network. The SRB system is the first and the largest (in terms of the data volume) data store in Academia SINICA right now. The system was deployed by ASGC in

early 2004, which consists of 7 sites in different institutes, linked by a dedicated fibre campus network, and provided 60 TB capacities in total. In early 2006, it will expand to 120 TB. As of January 2006, more than 30 TB and 1.4 million files have been archived in the distributed mass storage environment. All files are also preserved in two copies on different sites.

In this presentation, idea for utilizing Data Grid infrastructure for NDAP will be depicted and discussed. We will describe the use of SRB in building a collaborative environment for Data Grid Services of NDAP. In the environment, many data intensive applications are developed. We also describe our integration experience in building applications of NDAP. For each application we characterize the essential data virtualization services provided by the SRB for distributed data management.

Primary author: Mr UENG, Wei-Long (Academia SINICA Grid Computing Centre, Taiwan)

Co-authors: Mr YEN, Eric (Academia SINICA Grid Computing Centre, Taiwan); Mr LIN, Hui-Min (Academia SINICA Grid Computing Centre, Taiwan)

Presenter: Mr YEN, Eric (Academia SINICA Grid Computing Centre, Taiwan)

Session Classification: 1c: Earth Observation - Archaeology - Digital Library

Track Classification: Earth Observation - Archaeology - Digital Library