WEB ENABLED ACCESS, RETRIEVAL AND DISSEMINATION OF GEOGRAPHICALLY DISTRIBUTED SPATIAL DATA

Ashok Kaushal*, V.K.M. Raja^, Sandeep K Srivastava^

* ASL Advanced Systems Pvt Ltd, Bangalore, India
^ Center for Development of Advanced Computing, Pune, India

KEY WORDS: Spatial Data Clearing House, Grid Computing, Metadata, Directory Server, Catalog Server, Web Technologies, Open standards, Interoperability

ABSTRACT:

Decision making in the fast moving global village needs an access to data/ information which is not only spatial in nature, but is distributed at geographically distinct locations. Here data has a reference to varied nature of survey, maps, positioning (air-photo, satellite images, GNSS), and reports that relate to specific place or region. With the exception of select Web based search engines (Google Earth, Virtual Earth, Yahoo Maps), limited effort are made to serve the decision makers (including a common man) in finding data/ information related to a particular location. While volume and availability of data with spatial and temporal context from multiple sensors/ sources is increasing with every day, there is a little effort to access and disseminate this data in a seamless manner. It is expected that the recently adopted Map Policy by Government of India will facilitate publication of existence of such data/ information in electronic form and sharing the same from multiple locations, so as to avoid duplication of effort in generation of same by multiple institutions/ organizations. Herein, there is an effort for providing a single window solution to update, maintain and access spatial data/ information (distributed at geographically distinct locations) by the user using Grid computing, Web technologies and Open standards.

1. INTRODUCTION

India is fast emerging as a knowledge society with emphasis on "transparent" e-governance using Information and Communication Technology (ICT). In the recent past, there is a substantial investment in communications technology (wireless, broadband and satellite) to exploit the potential of the information age. The recent initiatives by the Government of India, including “Right to Information” (RTI) and announcements for development of information infrastructure through Public Private Partnership (PPP), leads to believe that within a few years an unprecedented capability will exist for sharing of data/ information along "electronic superhighways."

Amongst the variety of datasets that would be involved, spatial data/ information will be of major content. These spatial data/ information sets are vital to make decisions at local, regional, state and central level for implementation of action plans, infrastructure development, natural and man-made disaster management, natural resource management and business development. Crop monitoring, land use assessment, environment concern and flood mitigation in addition to fleet management, marine navigation and safety of life are few examples.

The Web based Spatial Data Clearing House (SDCH) (Figure 1) will encompass technologies, standards, policies and people necessary to promote and share spatial data/ information at all levels of research & development (R&D), industries and academia.

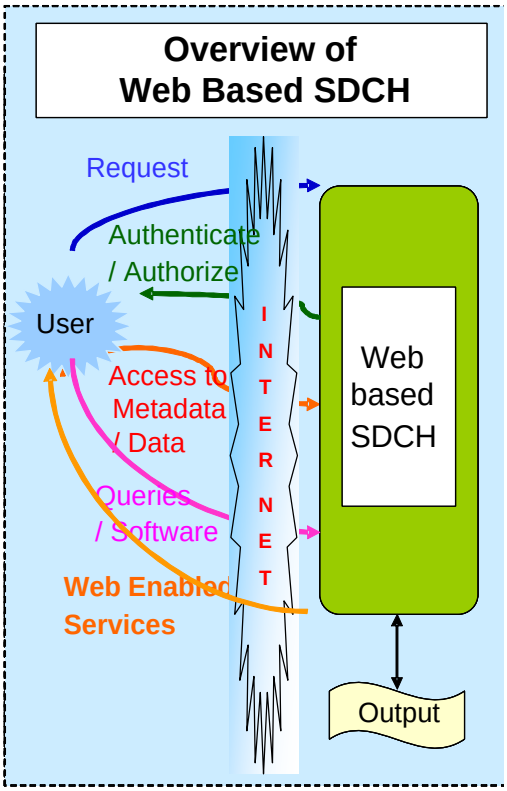

Figure 1 Overview of Web Based SDCH

Although, the need of the hour is to share data/ information for the benefit of the society, but the volume of data, their geographical distribution and low communication bandwidth along with the configuration of the receiving devices have been the major constraints to achieve the much-desired mission. The Web based SDCH implemented using Grid computing will provide the solution.

The Grid enabled and Web based implementation is a science of today and is going to be technology of tomorrow. It may be defined as an intelligent, high-performance and cooperative problem-solving or decision support system. It will provide a platform to access, share and manage worldwide resource of data, information, knowledge, and services.

2. GRID COMPUTING

Grids enable the sharing, selection, and aggregation of a wide variety of resources including supercomputers, storage, data/ information, sensors, networks and specialized devices that are geographically distributed and owned by multiple organizations for solving large-scale compute intensive and data intensive problems in science, engineering, and commerce. Grid computing can be defined as possibility of using distributed computers as a single, unified computing and storage resource. With Grid computing, the users can share data/ information, methods and models, software components, computing capabilities, services, etc. for a wide range of applications.

Grids were originally conceived for solving large-scale scientific applications using resources located at geographically distinct locations. The same are being experimented for many other applications including providing information services among multiple departments and domains using services on the World Wide Web (WWW). The common requirement across such opportunities is to share the compute and storage resources placed at geographically distinct locations in controlled and secured manner. In comparison to a conventional distributed system, the Grid enabled SDCH has following advantages:

- The spatial data of multiple scales, time, domains located on multiple nodes/ servers across distinct locations becomes seamlessly available and more effective for decision making as issues of system heterogeneity, security and rescheduling of resource get wrapped at Grid level.
- The system can be extended and scaled with the incremental growth of spatial data as there is no need for special software and analysis/ modeling tool a prior installed on the system. Adding a new node/ server to the Grid allows becoming part of the bigger system.
- The system provides access to data created and maintained by multiple organizations on multiple servers through shared infrastructure without being duplicated by the users independently.
- The system provides abundant resource and potential for managing/ archiving/ retrieving/ mining/ maintaining large-scale spatial data, its processing, analysis and modeling.
- While different applications will have varying spatial data usage requirements, the Grid based system will allow experimenting various models remotely on common data infrastructure.

C-DAC is collaborating with multiple organizations for development of select applications on Grid during 'Proof of Concept' of GARUDA project. The vision for the GARUDA is to have a unified and integrated approach to building distributed environments that incorporate computation, data management, scientific instruments and human collaboration. The proposed integration will facilitate to use large-scale, multi institutional, and dynamic, distributed application environments for doing science and commerce that is not possible, or is difficult and expensive, today.

3. COMPONENTS OF THE PROPOSED SYSTEM

The architecture of the proposed system is described in Figure 2 and will consist of:

3.1. Standards

The Grid enabled and Web based SDCH will require defining and using the Open standards, which are accepted universally. These standards will include but not limited to defining content and schemas, design and process, network protocols, exchange and transfer of spatial data. They will be of relevance to database standardization (formats, exchange, interoperability), networks-gateways and protocols, communication equipment, software implementation, etc. They will enable applications and technology to work in seamless manner. The software tools, applications and data affect each other. Hence, the processes for developing standards must consider interoperability among tools, applications and data (See also 3.7).

3.2. Metadata

The quality of the service will depend upon the quality of the data which is represented through its definition within Metadata. Hence, the development and use of Metadata standard for spatial data is one among the critical components of the proposed system. The Grid enabled and Web based SDCH will locate these Metadata in the Catalog server to share the qualifier for data before deciding to access and use the associated data by the end users.

Such Metadata should be defined by the overall needs and usage requirements of the proposed system. It is important that the stakeholders in SDCH contribute in the development of Metadata through definition of their business requirements, the content specifications

of datasets and the technology and implementation methods for servicing their users.

The major elements of Metadata about the dataset/ information includes but not limited to parameters on:

a) Identification Information
b) Data Quality Information
c) Spatial Data Organization Information
d) Spatial Reference Specification
e) Entity and Attribute Information
f) Distribution Information
g) Metadata Reference Information

Metadata will be exchanged using eXtensible Markup Language (XML) conforming to a Document Type Declaration (DTD). The support for XML in parsing and presentation solutions is widespread on the Web and is presumed in the current standards of the ISO TC 211 and OGC specifications.

3.3. Data Uploading Tool

The data servers in the system will hold spatial and non-spatial data in enterprise RDBMS (Oracle10g Spatial). These database servers will be populated using Standard Commercially off the Shelf (SCOTS) tool (GeoRasterETL from PCI Geomatics). A customized GUI will be created to upload data (both vector and raster data) in the servers using Oracle10g Spatial along with GeoRasterETL. The broad specifications of GeoRasterETL are as per under:

- Support for over 100 raster and vector formats
- Command-line and scripting support to enable loader to be integrated into production workflows
- Option to schedule within a batch-process
- Provides environment for flexible and customized transformation workflows

3.4. Spatial Data Clearing House Node

The SDCH will aid in decision-making in the context of planning and managing with various developmental activities. It will consist of several nodes which in turn will be servers where the spatial data/ information will reside as per the conformity of the system standards. The nodes will be primarily databases connected through high bandwidth and created and maintained by multiple organizations placed at spatially distinct locations. They may have different operating system and use different tools for generation of spatial data.

The data/ information (both spatial and non-spatial) will be uploaded into RDBMS (Oracle10g) using Data Uploading Tool referred above. Oracle 10g is a robust enterprise RDBMS, suitable for Grid architecture. It supports storage of spatial data using Spatial Cartridge for vector data and Geo Raster for raster imagery.

3.5. Directory Server

The support of discovery and access service for spatial data/ information in SDCH will depend upon the development of a “Search service”.

The Directory Server will contain the information regarding the available nodes, which are the repositories of data/ information residing at single or multiple organizations. The existing organizations will register with the Directory Server, which will help the SDCH to identify the potential resources.

3.6. Catalog Server

The Catalog Server will contain information regarding various resources located in multiple nodes. This information will be maintained using Metadata.

The Catalog Server will allow a user to query distributed collections of spatial data/ information through their Metadata descriptions. A user interested in locating spatial data/ information will use a “Search Service” and fill a search form, specifying queries for the data with certain properties. The search request will be passed to one or more registered Catalog Servers.

Each Catalog Server will manage a collection of Metadata entries. Within the Metadata entries, there will be instructions on method to access the data/ information being described.

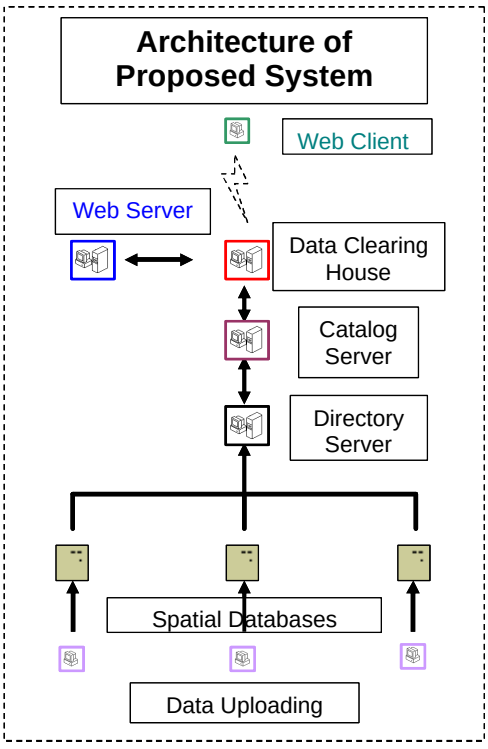

Figure 2 Architecture of Proposed System

3.7. Interoperability

Data interoperability will be incorporated into the system to aid in accessing multi format datasets (vector and raster). The databases placed at multiple locations/ organizations may contain datasets in their respective proprietary data formats, which will be standardized and made interoperable using standards (GML, WFS, WMS, WCS) defined by the Open Geospatial Consortium (OGC) for seamless access and transfer to other servers on the system.

3.8. Web Server

The Web server will be responsible for accessing the spatial data/ information from the backend and publishing it on the Web using standard Web Browsing tools (IE, Netscape, Mozilla and others). This will work along with the electronic clearing house to provide the spatial data and its attribute information to the end user.

The Web server will be developed using OGC Complaint SCOTS tool and will support Web Feature Service (WFS), Web Map Service (WMS), and Web Coverage Service (WCS). WMS generally serves the spatial data/ information in the form of images to the clients while WFS serves data as features and WCS as grid (raster). The spatial data served by these services will be customized to include different styles using Style Layer Descriptor (SLD) and displayed using Scalable Vector Graphics (SVG).

3.9. Web Client

The Web Client will be developed as Spatial Decision Support System (SDSS) by customizing SCOTS tools for providing an access and query the spatial data/ information for analysis.

The system will be implemented using J2EE along with XML (eXtensible Markup Language/ GML (Geographic Markup Language) for transfer of data.

3.10. Data Clearinghouse

The Data Clearinghouse will be the mechanism to provide access to the Metadata and finally to the actual data/ information sets through WWW. The clearinghouse will have systems to authenticate and authorize data requests.

Often spatial data volumes may be large and download through networks may not be feasible. In such scenario, the system will generate media bearing the requested data for transmission by mail.

The Data Clearinghouse will use the search and access protocols engines to look for and discover data and information. The “Search Service” will query the Catalog Server for information using Metadata from the underlying data servers in collaboration with the Directory Server.

The Data Clearinghouse will also be responsible to give the client necessary order details for purchase of the spatial data/ information. A user friendly GUI will be developed for end-user to access and display spatial data/ information.

3.11. Data Compression

The data compression tools will help in compressing/ uncompressing of maps and images. This will help in easy transfer of the data over the network.

3.12. Data Security

The entire system will be overlapped with a strong security overlay that will provide system level, network level and storage level security. Security of information will be a prime requirement, as transfer of spatial data/ information over the network will have potential risk of being misused.

For this purpose, a robust system of access control/ authorization both for individual servers on grid will be needed. Multiple firewalls and DMZ (Demilitarized Zones) will be setup. Depending upon the requirement, the entire data will be encrypted during storage to avoid mishandling of data.

4. CONCLUSIONS

Crop monitoring, land use assessment and flood mitigation along with fleet management, marine navigation, and others need timely access to spatial data/ information created and maintained at multiple geographically distinct locations for decision making.

The proposed Spatial Data Clearing House, using Grid computing, Web technologies and Open standards, provides a single window solution to access, retrieve and disseminate data/ information across multiple servers/ domains/ organizations placed at geographically distinct locations in a seamless manner using WWW.

5. REFERENCES FROM OTHER LITERATURE

CARL REED, 2005. Interoperability, standards, and the geospatial industry. GIS Development

KAUSHAL A., SRIVASTAVA S. AND KHARE M. 2005. Grid computing for disaster management. National seminar on GIS for rural development (G4RD) with focus on disaster management, NIRD, MORD, Hyderabad (March 09-11, 2005)

WANG, Y., GE, L., RIZOS, C., & BABU, R., 2004. Spatial data sharing on grid. Geomatics Research, Australasia, 81, 3-18.