Supporting Event-based Unified Data Access/query over Integrated Data-views for Decision Making in Geographic Information Systems

A. Sayar [a,b.] *, G. C. Fox [a,b,c], M. Pierce [a]

[a] Community Grids Laboratory, Indiana University
501 N. Morton Suite 224, Bloomington, IN 47404
{asayar, gcf, mpierce}@cs.indiana.edu
[b] Computer Science Department, School of Informatics, Indiana University
[c] Physics Department, College of Arts and Sciences, Indiana University

KEY WORDS: Decision making, GIS, federation, Web Services

Abstract

Geographic information is critical to effective and collaborative decision making in earth-related disaster planning, crisis management and early-warning systems. Decision making in Geographic Information Systems (GIS) increasingly relies on analyses of spatial data in map-based formats. Maps are complex structures composed of layers created from distributed heterogeneous data and computation resources belonging to the separate virtual organizations from various expert skill levels.

We describe a distributed computing infrastructure based on service oriented architecture principles for understanding and managing the production of knowledge from distributed observation, simulation and analysis through integrated data-views in the form of multi-layered map images. Our infrastructure is based on common data model, standard GIS Web-Service components with capability metadata and a federator. The federator aggregates GIS services and enables unified data access/query and display over integrated data views. Integrated data views are defined in federator's capability metadata as composition of layers provided by standard GIS Web-Services. Our Grid [grid, grid2] approach is based on the WS-I+ [wsi]. Key issues that we investigate include (1) a service-oriented GIS framework upon Web Services and Open Geographic Standards; (2) federator architecture for aggregating GIS Web Service components and enabling uniform data access/query and display over integrated data-views; (3) layer-structured integrated data-views as map images; and (4) high-performance design features for decision support in GIS requiring quick response times. From our investigations, we demonstrate that our framework enables seamless access to heterogeneous resources by hiding system complexities from the decision makers through the application specific hierarchical data definition as layer structured map images.

* Corresponding author.

Background and Motivation

Geographic Information Systems (GIS) [giscgl] are systems for creating, storing, sharing, analyzing, manipulating and displaying spatial data and associated attributes.

The general purpose of GIS is extracting information/knowledge from the raw geo-data. The raw data is collected from sensors, satellites or other sources and stored in databases or file systems. The data goes through the filtering and rendering services and is ultimately presented to the end-users in human recognizable formats such as images, graphs, charts etc. GIS is used in a wide variety of tasks such as urban planning, resource management, emergency response planning in case of disasters, crisis management and rapid response etc.

Over the past decade, GIS has evolved from traditional centralized mainframe and desktop systems to collaborative distributed systems. Centralized systems provide an environment for stand-alone applications in which data sources, rendering and processing services are all tightly coupled and application specific. Therefore, they are not capable of allowing seamless interaction with the other data or processing/rendering services. On the other hand, the distributed systems are composed of autonomous hosts (or geographically distributed virtual organizations) that are connected through a computer network. They aim to share data and computation resources collaborating on large scale applications.

Modern collaborative GIS requires data and computation resources from distributed virtual organizations to be composed based on application requirements, and accessed and queried from a single uniform access point over the refined data with interactive display tools. This requires seamless integration and interaction of data and computation resources. The resources span over organizational disciplinary and technical boundaries and use different client-server models, data archiving systems and heterogeneous message transfer protocols

These issues motivated us to develop a federated service-oriented data-grid framework providing unified data access and query over integrated data-views for the decision support in Geographic Information Systems.

A Federated, Service-Oriented GIS Framework

Our framework supports collaborative decision making over integrated data views, described in layer-structured hierarchical data provided by a federator. The users access the system as though all the data and functions come from one site. The data distribution and connection paths stay hidden and formulated as hierarchical data defined in federator’s capability metadata. The users access the system through integrated data-views (maps) with the event-based interactive mapping display tools. Tools create abstract queries from the users’ actions through action listeners and communicate with the system through federator.

Federation is based on federating service-oriented standard GIS Web Services capabilities metadata and their standard service interfaces for accessing, querying, and rendering data.

Creating such a federated design has some advantages in data sharing, reliability, and system growth (interoperability and extensibility).

Capability is a metadata about the data and services together. It includes information about the data and corresponding operations with the attribute-based constraints and acceptable request/response formats. It supplements the Web Service Description Language (WSDL) [wsdl], which specifies key low-level message formats but do not define an information or data architecture. These are left to domain specific capabilities metadata and data description language (GML) [gml]. Capabilities also provide machine and human readable information that enables integration and federation of data/information. It also aids the development of interactive, re-usable client tools for data access/query and display.

Our federated, service-oriented GIS framework is composed of two parts [see *Figure 1*]. The first part consists of interoperable service-oriented GIS components compliant with Open Geographic Standards, and the second part is federator for composing the GIS service components according to the application requirements and integrated data-views defined in its capability metadata.

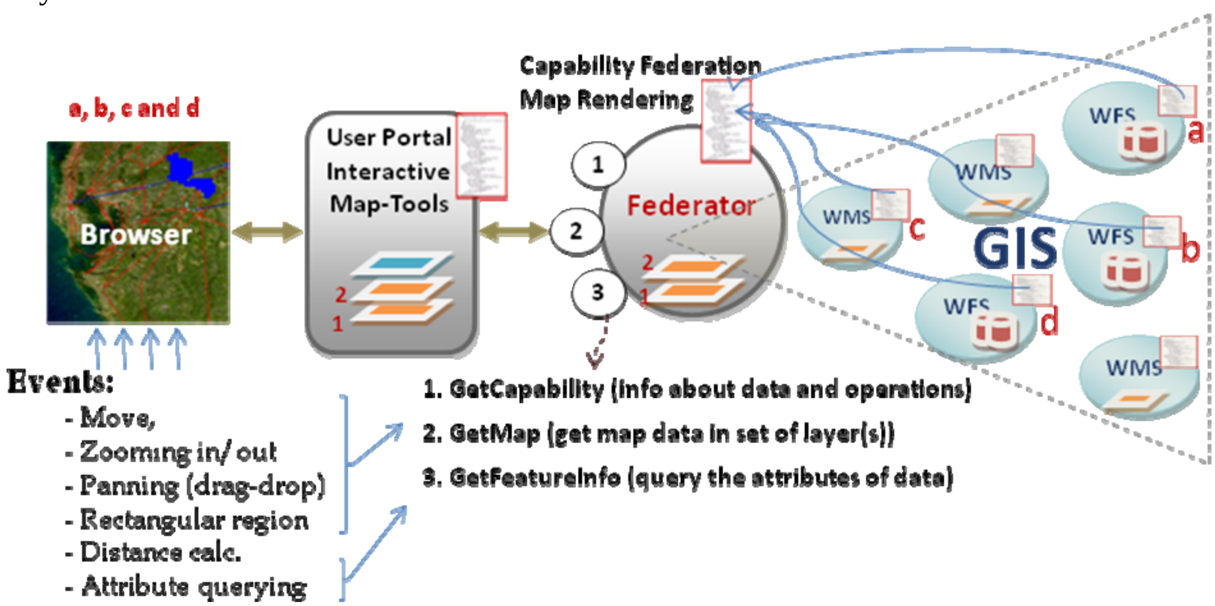

Figure 1: Federated, service-oriented Geographic Information System framework supports decision making over integrated data-views.

There are two types of GIS services that we consider: data rendering services (Web Map Services (WMS) [wms]) and data providing services (Web Feature Services (WFS) [wfs]). There are also two types of data, vector data provided by WFS in the form of XML-encoded common data model (GML) and binary layer-images provided by WMS. GML is a data description language which is XML encoding of the heterogeneous data. It consists of two parts, content (core data) and presentation (attribute and geometry elements). GML's properties enable data to be displayed, queried and easily integrated.

GIS services developed along web service standards have limited performance. To improve performance, we extend services to support streaming data. The streaming data transfer is done through the topic-based publish/subscribe messaging middleware [narada].

The federator aggregates the standard GIS Web Services' capabilities metadata and presents a single capability image to the user through application-based hierarchical data defined in its federated capability metadata. This enables unified data access/query/display from a single access point through abstract queries from event-based interactive map tools. These generic tools enable seamless interaction with the system through federator or any other compatible WMS.

Performance: Reliability and performance are crucial issues for federated services and are one of our main areas of investigation. As we will describe, federation in many circumstances increases performance through concurrent and distributed/parallel image rendering. Performance features of the proposed federated framework [*Figure 1*] mostly focus on applications of caching and parallel processing approaches based on the geo-data characteristic. Geo-data is described based on their locations, and based on that they are unevenly distributed and variable sized. In order to address these issues, we develop novel federator-oriented caching and parallel processing approach such as pre-fetching and client-oriented caching taking the locality issue into considerations to support high-performance in GIS requiring quick response times.

Applications: Application-based hierarchical data is defined as an integrated data-view in the federator's capability metadata. It actually defines a static workflow starting from the federator and ending at the original data sources (WFS serving GML or WMS serving map layers). The services are linked through the references defined in their capability metadata. Decision makers' interactions with the system are carried over the integrated data views through event-based interactive map tools. Integrated data-views are defined in the hierarchical data format as explained below:

Application ->Map -> Layer -> Data {GML and/or binary images} ->Raw data (any type).

A map is application-based, human-recognizable, integrated data display and is composed of layers. A layer is a data rendering of a single homogeneous data source. Layers are created from the structured XML-encoded common data model (GML) or binary map images (raster data). Heterogeneous data source are integrated to the system through the mediators in the form of GML (WFS-based mediation) or binary map images (WMS-based mediation). The mediators have resource specific adaptors for request and response conversions and appropriate capability metadata describing the data and resources.

Our experiences with GIS have shown that federated, service-oriented, GIS-style information model can be generalized to many application areas such Chemistry and Astronomy. We call this generalized framework Application Specific Information System (ASIS) and give blueprint architecture in terms of principles and requirements. Developing such a framework requires first defining a data description language expressing the primitives of the domain, second, key service components, service interfaces and message formats defining services interactions, and third, the capability file requirements enabling inter-service communications to link the services for the federation.

The chapter will be organized as follows. We first survey GIS and decision making in early warning systems such as disaster planning and crisis management. We also describe Pattern Informatics (PI) [pi02] earthquake forecasting application as a motivating use case. Second, we address heterogeneity and interoperability issues and propose a service-oriented GIS architecture by integrating Open Geographic Standards with Web Service principles. Third, we present the federation architecture federating service-oriented GIS components to provide application-based integrated data-views for the decision support. Fourth, we validate the architecture with performance tests. Finally, we summarize the principle and requirements to create proposed GIS-style information model for general science domains.

REFERENCES

[giscgl] GIS Research at Indiana University Community Grids Lab, Project Web Site: http://www.crisisgrid.org.

[gml] Cox, S., Daisey, P., Lake, R., Portele, C., and Whiteside, A. (eds) (2003), OpenGIS Geography Markup Language (GML) Implementation Specification. OpenGIS project document reference number OGC 02-023r4, Version 3.0.

[grid] Grid Computing: Making the Global Infrastructure a Reality edited by Fran Berman, Geoffrey Fox and Tony Hey, John Wiley & Sons, Chichester, England, ISBN 0-470-85319-0, March 2003. http://www.grid2002.org.

[grid2] The Grid 2: Blueprint for a new Computing Infrastructure, edited by Ian Foster and Carl Kesselman, Morgan Kaufmann 2004.

[narada] Pallickara S. and Fox G., “NaradaBrokering: A Distributed Middleware Framework and Architecture for Enabling Durable Peer-to-Peer Grids” ACM/IFIP/USENIX, Rio Janeiro, Brazil June 2003.

[pi02] Tiampo, K. F., Rundle, J. B., McGinnis, S. A., & Klein, W.: “Pattern dynamics and forecast methods in seismically active regions”, Pure Applied Geophysics 159, 2429-2467 (2002).

[wfs] Vretanos, P. (2002) Web Feature Service Implementation Specification, OpenGIS project document: OGC 02-058, version 1.0.0. Volume,

[wms] de La Beaujardiere, J., Web Map Service Specifications, OGC project document reference number OGC 04-024. 2004.

[wsdl] Christensen, E., et al., Web Services Description Language (WSDL) 1.1. 2001, March.

[wsi] Malcolm Atkinson, David DeRoure, Alistair Dunlop, Geoffrey Fox, Peter Henderson, Tony Hey, Norman Paton, Steven Newhouse, Savas Parastatidis, Anne Trefethen and Paul Watson.Web Service Grids: An Evolutionary Approach UK e-Science Technical Report July 13 2004 published in a special issue of Concurrency&Computation: Practice&Experience 17, 377-389, 2005.