CONCEPT AND REALIZATION OF FEDERATED SPATIAL DATABASE

Jianya Gong Hanjiang Xiong Yandong Wang Lite Shi

State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS)
Wuhan University, 129 Luoyu Road, Wuhan, China, 430079
(jgong, wyd, xhj)@rcgis.wtusm.edu.cn, lilyshi@hp01.wtusm.edu.cn

Commission IV, WG VI/2

KEYWORDS: Spatial Database, Federated Database System, Geographic Information System

ABSTRACT:

Geographic information systems manage a large volume of spatial data. We often use it to solve some problems such as urban planning and land use management. Interoperability of GIS is the ability to access spatial data and associated services in a distributed and heterogeneous processing environment. There are many existing spatial databases and GIS applications that are built on those databases today. In order to solve problems, we often need the cooperation of many spatial databases, even if there are more complex spatial relationship and diversity between the existing spatial data models. In this paper, we will discuss the concept and methodology of a Federated Spatial Database System to implement the interoperability of GIS. At the end of paper, we introduced a project named VirtualWorld 2000. The goal of this project is to build a federated spatial database system to access the existing component spatial databases distributed in the LAN, which include graphic databases, DEM databases and image databases built with Oracle8i spatial cartridge, SQL Server 7.0 and GeoStar 3.0 file system.

1. INTRODUCTION

Geographic information systems manage a large volume of spatial data. We often use it to solve some problems such as urban planning and land use management. Interoperability of GIS is the ability to access spatial data and associated services in a distributed and heterogeneous processing environment [5]. There are many existing spatial databases and GIS applications that are built on those databases today. In order to solve problems, we often need the cooperation of many spatial databases, even if there are more complex spatial relationship and diversity between the existing spatial data models.

Interoperability among traditional geographical information systems requires solving two major problems:

- How to access spatial data distributed on the network.
- How to allow cooperation between existing heterogeneous spatial databases.

The first problem is dealt with by using networking techniques based on the client/server paradigm. For example a distributed object environment, such as CORBA can be used to provide interconnectivity among several geographical information systems.

The second problem has recently been the focus of extensive researches in the field of databases. Several solutions have been identified:

- Schema integration based solutions aim to combine all information into a single global schema.
- Canonical data-model based solutions hide heterogeneity and provide a good framework for multi-model translations.
- Multi-database language based solutions allow users to query several information sources at the same time.

In addition to the above problems, interoperability of GIS must address the following issues:

- Support for spatial data which exhibit more complex relationship than traditional applications;
- Support for spatial operators as spatial analysis and spatial query;
- Support for new types of heterogeneity specific to spatial information systems (generic, spatial contextual, models, and operators heterogeneity).

In section 2, we will discuss the distributed spatial data model of the federated spatial database system. We will build a distributed spatial data model that extends the Open GIS Simple Features Specification in order to manage the DEM and image data. In section 3, we will discuss the architecture of the federated spatial database system. We will build architecture with the middleware components and spatial agents. In section 4, we will discuss the spatial data model integration between the component spatial databases. In section 5, the spatial query and operators of the federated spatial database system will be discussed. We need to define the visual spatial query language for the federated spatial database system. In section 6, the spatial object migration in the federated spatial database system will be discussed as well.

At the end of this paper, we will introduce a project named VirtualWorld 2000, the goal of which is to build a federated spatial database system to access the existing component spatial databases distributed in the LAN, which includes graphic databases built with Oracle8i spatial cartridge, SQL Server 7.0 and GeoStar 3.0 file system, digital terrain databases and digital ground image databases obtained with photogrammetry or remote sensing technology. We can build the 2D or 3D applications on this system and keep those spatial databases autonomous, that is to say, all the existing applications built on them are not necessary to be rebuilt. Also, it is easy for us to add or delete a component spatial database.

2. THE CONCEPT OF THE FEDERATED SPATIAL DATABASE SYSTEM

In 1990, Kamel represented the concept of the Federated Database Management System [2]. This approach allows users and applications to access data across several heterogeneous databases (relational, object-relational and object –oriented DBMS) while maintaining their autonomy. Soutu. C discussed the methodology for developing a federated database system [3]. Many experts had discussed the query, schema transformation and other issues about the Federated Database System.

The integration of spatial databases has encountered the same problem. The cooperation of several spatial databases demands that the application should access those spatial databases transparently. With the theory, we can build a federated spatial database system to implement the interoperability of GIS. A Federated Spatial Database System (FSDBS) consists of autonomous component spatial database systems that participate in the federation to allow controlled sharing of their spatial data. There are several major characteristics:

- The members of the FSDBS may be heterogeneous, distributed and autonomous spatial databases. They can either join in or disengage the FSDBS freely and will not affect the whole system.
- The Federated Spatial Database System has a common spatial data model to hide the heterogeneity among different local spatial databases.
- The Federated Spatial Database System provides a set of global spatial query languages to query and access the local spatial databases.

3. THE COMMON SPATIAL DATA MODEL OF FSDBS

A Federated Spatial Database System has many local spatial databases. These local spatial databases will have different spatial data models. So we should define a common spatial data model to support a set of spatial concepts to hide heterogeneity between spatial models. The Open GIS Consortium (OGC) has presented us a good spatial model to integrate the different spatial models. Also, OGC has released an implementation specification in 1999 - the OpenGIS Simple Feature For OLE/COM Revision1.1. As more and more software packages have been adopting this specification, we can adopt it as the common data model of FSDBS as well. But the model has a deficiency, that is, the specification cannot define the image objects and DEM objects. We hope that the local spatial databases of the FSDBS can be DEM, image or 3D vector databases. The common model should be extended in the following way (Figure 1):

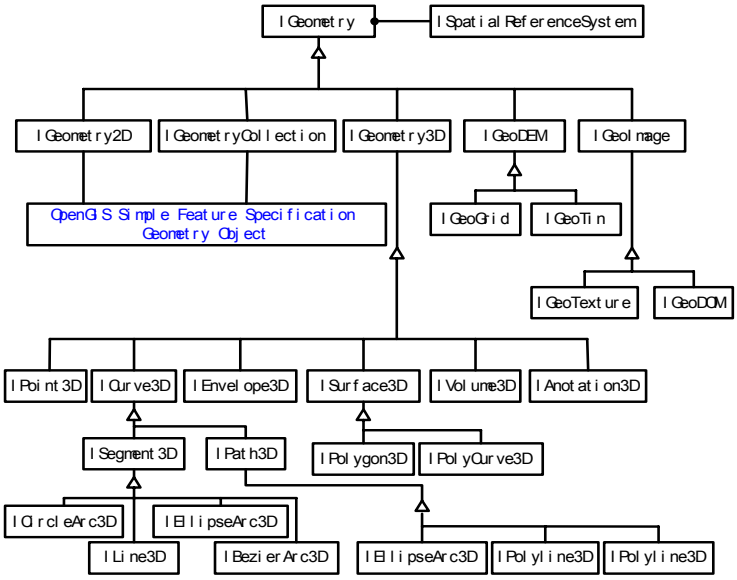

Figure 1 The Extended Spatial Object Architecture for FSDBS

4. THE SCHEMA DEFINITION AND INTEGRATION OF FSDBS

According to the methodology of developing a federated database system [3], the architecture of a federated database system can be described by a five-level schema. There is five-level schema: local schema, component schema, exported schema, the federated schema and the external schema.

Spatial Data Coverage is the core concept of the five-level schema. We can see the FSDBS as:

1) All the local spatial databases are consisted of several exported spatial data coverages.
2) FSDBS has several federated spatial data coverages. Each federated spatial data coverage is coming from one or more exported spatial data coverages.
3) From the users' point of view, the FSDBS is consisted of several users' spatial data coverages. Each user's spatial data coverage is also coming from the federated spatial data coverage.

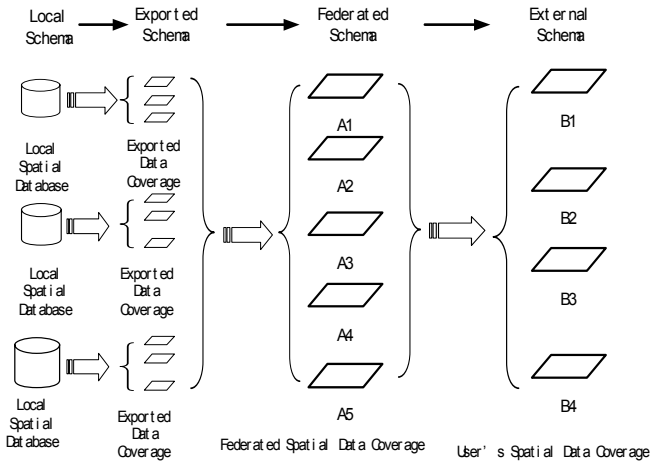

Figure 2 The Schema and Transformation of FSDBS

5. QUERYING THE FSDBS

In order to access the heterogeneous, distributed and autonomous spatial databases, we should have a spatial query language for the FSDBS. All the user's spatial data requests are translated into the global query language. The FSDBS discomposes the query into several sub-query commands and send them to each local spatial database. The spatial databases execute the command with the local mode, and retrieve the result to FSDBS. The FSDBS then integrates the result and retrieve the result to the user. In order to interpret the procedure of the execution of user's query, an example is given as follows:

Suppose that we have a Chinese digital map with a scale of 1:1000000. The digital map has three data coverages: the railway coverage stored in SQL Server 6.5, the road coverage stored in Oracle8.1.6, the Province coverage and County Coverage stored in files which are managed by GeoStar3.0 (a GIS software package developed by Wuhan University).

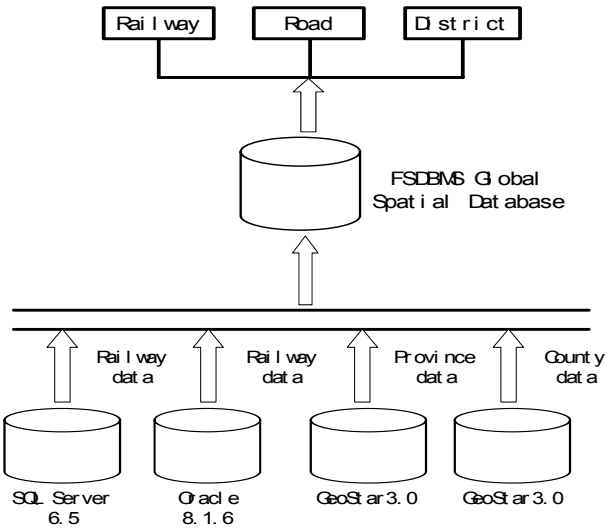

Figure 3 A FSDBS for One Part of the 1:1000000 Chinese Digital Map

The four spatial databases export four coverages. But in a user's view, the FSDBS only have three coverages in the exported schema: railway, road and distinct coverage.

A windows query maybe describe as following:
SELECT * FROM FSDBS WHERE Rect(minX,minY,maxX,maxY).

6. THE IMPLEMENTATION OF FSDBS

We have built a project named as VirtualWorld 2000, the motivation of which is to build a FSDBS to manage the existed DEM, image and vector graphic databases and provide the services for urban planning, while keeping the old applications built on these databases not to be modified. The architecture of VirtualWorld 2000 is as follows (Figure 4):

It notes that "FSDBS" means the query is execute in all the spatial databases that are members of the federal.

The procedure of the FSDBS spatial query will be as follows:

1) To query the global metadata database and decompose the query command into three absolute sub-query commands:
SELECT * FROM Coverage Railway WHERE Rect(minX,minY,maxX,maxY)
SELECT * FROM Coverage Road WHERE Rect(minX,minY,maxX,maxY)
SELECT * FROM Coverage Distinct WHERE Rect(minX,minY,maxX,maxY)

2) With the meta database, each of the three sub-query commands will be decomposed to several sub-queries. This decomposition should be insured that each sub-query command only relates to one local spatial database.

(1) The command, "***SELECT * FROM Coverage Railway WHERE Rect (minX,minY,maxX,maxY)***", will be sent to the wrapper of SQL Server 6.5.

(2) The command, "***SELECT * FROM Coverage Road WHERE Rect (minX,minY,maxX,maxY)***", will be sent to the wrapper of the Oracle 8.1.6.

(3) The command, "***SELECT * FROM Coverage Distinct WHERE Rect (minX,minY,maxX,maxY)***", will be decomposed into two sub-query commands: ***SELECT * FROM Coverage Province WHERE Rect(minX,minY,maxX,maxY),*** which will be sent to the wrapper of spatial database managed by GeoStar3.0, and ***SELECT * FROM Coverage County WHERE Rect(minX,minY,maxX,maxY),*** which will be sent to the wrapper of another spatial database managed by GeoStar3.0

3) Each wrapper converts the query from global spatial query language to local spatial query language or procedure. For example, the query: ***SELECT * FROM Coverage Road WHERE Rect(minX,minY,maxX,maxY)***, which has been sent to the wrapper of Oracle 8.1.6 should be converted as following:
SELECT * FROM Road WHERE sdo_filter(Road.GeoObject, mdsys.sdo_geometry{2003,NULL,NULL,mdsys.sdo_elem_info_array(1, 1003,3),mdsys.sdo_ordinate_array(minX,minY,maxX,maxY)}, 'querytype=window')='TRUE', and sent to Oracle 8.1.6 DBMS.

4) Each local spatial database executes the sub-query and retrieves the result to FSDBS with the wrapper's schema transformation. The FSDBS also execute the schema transformation, integrate the result and retrieve it to the user.

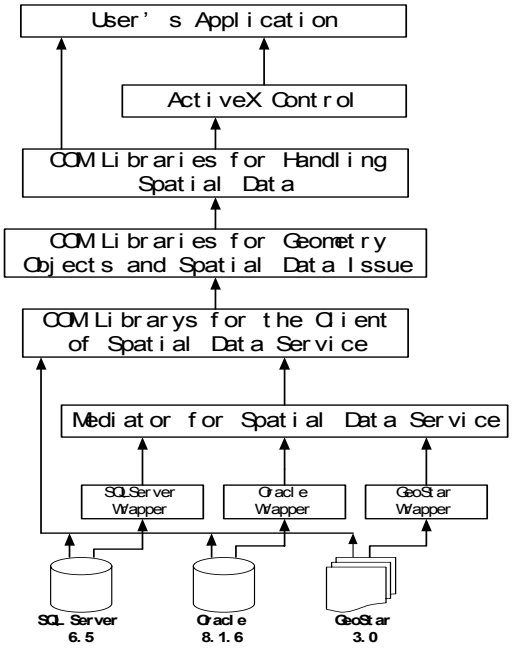

Figure 4 The Architecture of VirtualWorld 2000

7. CONCLUSIONS

The interoperable GIS is becoming more and more important today. The FSDBS provides us a new idea to build an interoperable GIS platform. This paper shows what is a FSDBS and how to begin the research of FSDBS. There are many problems need to be solved: the Global spatial Query Language, The Transaction and Security management and spatial object operator and so on.

REFERENCES

Holger, G., 2001. Object-oriented modeling of data sources as a tool for the integration of heterogeneous geoscientific information. *Computers & Geosciences*, 27, pp. 975-985.

Kamel, M. N. and Kamel, N. N., 1990. The federated database management system: an architecture of distributed systems for the 90's. In: *Proceedings of the 2nd IEEE Workshop on Future Trends*, pp. 346 –352.

Sheth, A. P. and Larson, J. A., 1990. Federated database systems for managing distributed heterogeneous and autonomous databases. *ACM Computing Surveys*.

Soutou, C. 1993. Towards a methodology for developing a federated database system. In: *Proceedings of ICCI '93: Computing and Information*, pp. 560 –564.

Tryfona, N. and Sharma, J., 1996. On information modeling to support interoperable spatial databases. In: *Proceedings of the 8th International Conference on Advances Information Systems Engineering, CaiSE'96*, Vol. 1080, Springer, pp. 210-221.

ABBREVIATIONS

FSDBS	Federated Spatial Database System
GIS	Geographic Information System
OGC	OpenGIS Consortium