

RESEARCH ON THE FRAMEWORK OF SPATIO-TEMPORAL DATA WAREHOUSE

WANG Jizhou, LI Chengming

Institute of GIS, Chinese Academy of Surveying and Mapping
No.16, Road Beitaping, District Haidian, Beijing, P.R.China, 100039
wjz@casm.ac.cn cml@casm.ac.cn

Abstract

Data Warehouse is “the topic-oriented, integrated, static datasets of various periods which are used to support decision-making in administration”. Driven by the requirement of mass spatio-temporal data management and application, Spatio-Temporal Data Warehouse was put forward, and many researchers distributed all over the world focused their energy on it. Although the research on Spatio-Temporal Data Warehouse is going deep, there are still many key problems to be solved, such as the design principle, system framework, spatio-temporal data model, spatio-temporal data process, spatial data mining and etc. In this paper, we discuss the concept of Spatio-Temporal Data Warehouse, and analyze the organization model of spatio-temporal data. Based on the above, we found the framework of Spatio-Temporal Data Warehouse, which is composed of data layer, management layer and application layer. The functions of Spatio-Temporal Data Warehouse should include data analysis besides data process and data storage. When users apply certain kind of data services, Spatio-Temporal Data Warehouse locates the right data by metadata management system, then start data processing tools to form data products or found multidimensional data cube which serves the data mining and OLAP. All kinds of distributed databases make up data sources of Spatio-Temporal Data Warehouse, including DEM, DRG, DLG, DOM, Place Name and other databases in existence. The management layer implements heterogeneous data processing, metadata management and spatio-temporal data storage. The application layer provides data products service, multidimensional data cube, data mining tools and on-line analytical process.

Keywords: Data Warehouse, Spatio-Temporal Data Warehouse, Spatio-Temporal Data Model, OLAP

1.Introduction

W.H.Inmon defined Data Warehouse as “the topic-oriented, integrated, static datasets of various periods which are used to support decision-making in administration”. Based on the Data Warehouse technology, we found Spatio-Temporal Data Warehouse by importing temporal and spatial data into Data Warehouse. Utilizing this warehouse, we abstract information according to application topics from various GIS, spatial databases and historical databases which are in different spatio-temporal scale, and supply spatio-temporal information by doing data processes to serve the scientific research, region economic decision-making, resource policy-constituting and so on.

Driven by the requirement of mass spatio-temporal data management and the progress of Data Warehouse technology, Spatio-Temporal Data Warehouse came into the world. In recent years, many researchers distributed all over the world focused their energy on it, and acquired many achievements. (Maryvonne, 2002; ZHAO,2000; LI,1999;Dimitris,2001;Giuseppe, 2001). Although the research on Spatio-Temporal

Data Warehouse is going deep, there are still many key problems to be solved, such as the design principle, system framework, spatio-temporal data model, spatio-temporal data process, spatial data mining and etc. In this paper, we discuss the concept of Spatio-Temporal Data Warehouse on the basis of an brief introduction of Data Warehouse, and analyze the organization model of spatio-temporal data. Based on the above, we found the framework of Spatio-Temporal Data Warehouse, which is composed of data sources, management system and application tools.

2. The organization of spatio-temporal data

2.1. Spatio-temporal data model

Two concepts of time are involved in research on temporal characteristic of geographic entities, which are world time and system time. The former means the time when an entity change take place in reality, and the latter means the time that records the entity change in database. In general, we only take the system time into account in GIS(Raafat, et al,1994). In Spatio-Temporal Data Warehouse, we also use the system time to mark the historical changes of entity. There are three methods to describe the changes(Ariav,1986;GONG,1997). When one or more objects change in a event, the first method records it by creating a new edition of all the tables related to these objects, the second by creating a new edition of the changed objects, and the third method records these changes by only adding a new record of the changed objects attribute field to the related table. By comparing theses methods, we can draw the conclusion that the first has the most redundancy, the second has edition controlling problems, and the third is ideal because it not only has the least redundancy, but also is easy to do querying and analyzing operations for its having historical information in the same record. Research on the spatio-temporal data model have made great progress in recent years. For examples, GONG Jian-ya put forward object-oriented spatio-temporal data model by enhancing the third method(GONG,1997), Giorgos Mountrakis founded a differential spatio-temporal model(Giorgos, et al,2002).

2.2 Storage of spatio-temporal data

The purpose of applying Spatio-Temporal Data Warehouse is to make the mass data sharing in the broadest scope. Since users have various data demands, they have various requests towards Spatio-Temporal Data Warehouse. To meet most users' demands and achieve fast responds, we use multi-layer strategy to store spatio-temporal information. Commonly, there are three layers: Data Market, Department Warehouse and Whole Warehouse(LI, et al,1999). Data Market is outcome datasets of querying in a lower level, and it mainly satisfy demands of general users. Department Warehouse which is based on the department topic can meet the demands of department leaders. And Whole Warehouse is founded for high-class decision.

3. The framework of Spatio-Temporal Data Warehouse

The functions of Spatio-Temporal Data Warehouse should include data analysis besides data process and data storage. When users apply certain kind of data services, Spatio-Temporal Data Warehouse locates the right data by metadata management system, then start data processing tools to form data products or found multidimensional data cube which serves the data mining and OLAP. Thus we think that the framework of Spatio-Temporal Data Warehouse is composed of Data Sources,

Management System and Application Tools. Figure 1 shows the framework of Spatio-Temporal Data Warehouse in detail.

3.1. Data sources

3.1.1. Distributed spatio-temporal databases

These databases are the information source of Spatio-Temporal Data Warehouse, which include all kinds of DEM, DRG, DLG, DOM, Place Name and other databases in existence. Not only the hardware and software platforms that these databases run on are diverse, but also the encoding specifications, projection systems, data formats and etc of them are in variety.

3.1.2. Special databases

Special databases store the inner data of special departments which are used to fulfill OLAP and SOLAP by associating with the spatio-temporal data, such as the populace data of police, statistical data of revenue, etc.

3.2. Management system

3.2.1. Data processing

Because the databases in existence serve various applications, there are many differences between each other in data capture methods, encoding specifications, projection systems, data classification standards and data formats, and there are even data mistakes in some databases. Hence we must do some data processing before putting them into Spatio-Temporal Data Warehouse, which includes data conversion, spatial transformation and data filter (ZHAO, et al,2000). Data conversion means unifying the data encoding and structure, adding the temporal mark, operation and semantic conversion of data sets. Spatial transformation means unifying the data coordinate and scale. Data filter means the data extraction, mainly including recomposing of data fields, deleting the useless information, translating and decoding of fields, supplement of the devoid information and verification of data integrity.

Spatio-Temporal Data Warehouse should supply data products for various users. Since there are so many users who have diversiform demands, we have to do different processing on the data towards different demands. The user-oriented data processing mainly includes data integration, data fusion and data recomposition. Data integration means the overlay of multiple data that all the data remain their own characteristics after overlaying, for an example, using DOM and digital map to produce image map. Unlike data integration, data fusion may create a new kind of data, such as the pseudocolor-composite image. Data recomposition means the organization of various geographic features, for an instance, the city frame data for the department of estate management which is composed of boundary, street and water system.

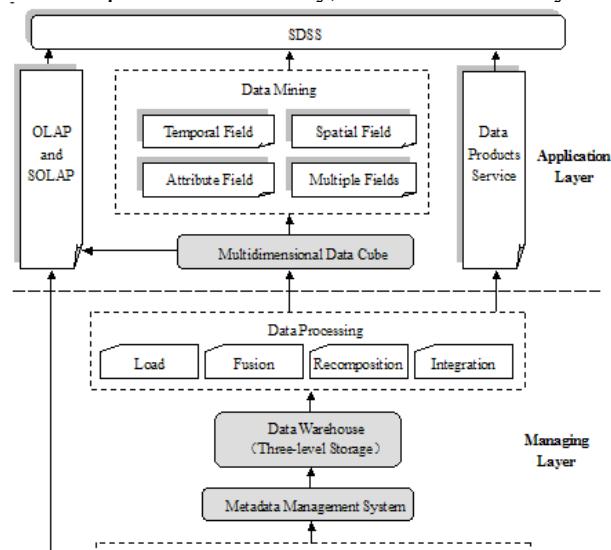


Figure 1 The Framework of Spatio-Temporal Data Warehouse

3.2.2. Metadata management system

Metadata is the background information that describes the content, quality, condition and other appropriate characteristics of the data. Metadata is a simple mechanism to inform others of the existence of datasets, their purpose and scope. Key developments in spatial metadata standards are the ISO STANDARD 15046-15 METADATA, the Federal Geographic Data Committee's content standard for Digital Geospatial Metadata FGDC, CSGDM, the European organization responsible for standards CEN/TC 287. According to these standards, spatial metadata should include eleven categories information, which are identification information, data quality information, maintenance information, spatial representation information, reference system information, entity and attribute information, distribution information, metadata reference information, citation information, time period information and contact information.

3.2.3. Spatio-Temporal Data Warehouse

The data in existence will be stored in the Spatio-Temporal Data Warehouse after processing. In the warehouse, we organize and manage the mass data using multiple dimension mechanism, which includes temporal dimension(one dimension), spatial dimension(three dimensions, X, Y and Z) and attribute dimension(multiple dimensions, name, type, address, etc.). And each dimension has various granularities, such as temporal dimension may be partitioned into year, month, day, hour, minute and second, spatial dimension may be divided into nation, province, city, county and town. So users can utilize the spatio-temporal data from diversiform points.

3.3. Application tools

3.3.1. Data products service

Spatio-Temporal Data Warehouse can supply various data products service through data integration, data fusion and data recomposition, such as the on-line query, scan and display, the on-line order and distribution of data products and so on.

3.3.2. Multidimensional data cube

Because the Spatio-Temporal data warehouse uses multiple dimension mechanism in the organization and management, it can expediently found the multidimensional data cube, or super cube data model, which serves the data mining and OLAP. The cube is composed of temporal, spatial and attribute dimension, whose actual dimension

number depends on the decided demand. For an example, when taking on the analysis of populace movement, we can build an data cube which is composed of five dimensions including temporal, X, Y, Z and populace statistical data.

3.3.3. Data mining

Data mining is a new technology that supports decision by abstracting knowledge from databases. All over the world there are many researchers who devote themselves to it. Professor Han Jiawei in Canada has implemented an prototype system based on MapInfo, which has many analysis methods of data mining including spatial comparison, spatial correlation, spatial clustering, spatial classification and so on. Professor Li Deren in China put forward that people could extract knowledge as spatial information, spatial relation, attribute relation and etc from GIS databases. We think that the spatio-temporal data comprises of three kinds of elements: temporal, spatial and attribute information, thus the spatial data mining can be partitioned into four kinds: temporal field data mining for the simulation and retrieval of change processes, the spatial field mining for geographical distribution of topic information, the attribute field mining of spatial correlation and clustering for region economic decision and the multi-fields associated mining for governmental decision.

3.3.4. OLAP and SOLAP

OLAP(On-Line Analytical Process) technology enables users to quickly analyze mass data. OLAP systems are generally based on a three-tiers architecture including a data warehouse with integrated data, an OLAP server for the dimensional view and an OLAP client, i.e. a user interface for the rapid and easy exploration of data(Han, et al,2001). SOLAP(Spatial On-Line Analytical Process) systems are built to support the spatio-temporal analysis as well as the exploration of data according to a multidimensional approach of data warehouse. Sometimes we should associate the spatio-temporal data with the special data by the unique key word in OLAP and SOLAP, which is could be the coordinate, place name, project number and so on in both the spatio-temporal data and the special data. By applying the analysis technology as rotation, drilling, nesting, slicing and visualizing on the multidimensional data cube, OLAP and SOLAP enable users to look into the spatio-temporal data from various sides and discover the potential relations of data, and at last facilitate the decision-making.

4. Conclusion

Since the spatio-temporal data is so complicated and dynamic, the existing data warehouse technology can't solve all the matters of Spatio-Temporal Data Warehouse. The research on Spatio-Temporal Data Warehouse is still in elementary stage. In this paper, we found the framework of Spatio-Temporal Data Warehouse from three tiers: data, management and application, which covers the whole process from building warehouse to supplying services. Next we will go on to take deeper study on Spatio-Temporal Data Warehouse technology which includes the data processing, metadata management, data storage, data mining and so on, and perfect the framework in practices.

References

- [1] W.H.Inmon,1992. Building The Data Warehouse. Canada: John Wiley & Sons Inc.
- [2] ZHAO Pei-sheng, YANG Chong-jun,2000,The Technology and application of spatial data warehouse. Journal of Remote Sensing, 2000,4(2) (in Chinese).