A FRAMEWORK FOR SPATIO-TEMPORAL DATA WAREHOUSE

M.LAXMAIAH[1], DR.A.GOVARDHAN[2], DR. C.SUNIL KUMAR[3]

[1]JNTUH University, Tirumala Engineering College, Hyderabad

lax.phd2006@gmail.com

[2]JNTUH University, Director of Evaluation, Hyderabad

govardhan_cse@yahoo.co.in

[3] Anna University, Dhanalakshmi College of Engineering, Chennai

sunil1977.dce@gmail.com

Abstract:

Data Warehouse (DW) is topic-oriented, integrated, static datasets which are used to support decision-making. Driven by the constraint of mass spatio-temporal data management and application, Spatio-Temporal Data Warehouse (STDW) was put forward, and many researchers scattered all over the world focused their energy on it. Although the research on STDW is going in depth, there are still many key difficulties to be solved, such as the design principle, system framework, spatio-temporal data model (STDM), spatio-temporal data process (STDP), spatial data mining (SDM) and etc. In this paper, the concept of STDW is discussed, and analyzes the organization model of spatio-temporal data. Based on the above, a framework of STDW is composed of data layer, management layer and application layer. The functions of STDW should include data analysis besides data process and data storage. When users apply certain kind of data services, STDW identifies the right data by metadata management system, then start data processing tools to form a data product which serves the data mining and OLAP. All varieties of distributed databases (DDBs) make up data sources of STDW, including Digital Elevation Model (DEM), Diagnosis-Related Group (DRG), Data Locator Group (DLG), Data Objects Management (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.

Indexing terms: DW, STDW, STDM, OLAP

1. INTRODUCTION

Based on the DW technology, a STDW imports temporal and spatial data into DW. Utilizing this warehouse, information is abstracted 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 provide the scientific research, region economic decision-making, resource policy-constituting and so on. Motivated by the requirement of mass STDM and the progress of Data Warehouse technology, STDW came into the world. In recent years, many researchers distributed all over the world focused their energy on it, and acquired many achievements. Although the research on STDW is going deep, there are still many key problems to be solved, such as the design principle, system framework, STDM, STDP, SDM and etc. In this paper, we discuss the concept of STDW on the basis of an brief introduction of DW, and analyze the organization model of spatio-temporal data [1]. Based on the above, a framework of STDW, which is composed of data sources, management system and application tools are found.

2. SPATIO-TEMPORAL DATA ORGANIZATION

2.1. Spatio-temporal Data Model

Two concepts of time are involved in research on temporal characteristic of geographic entities, which are world and system time. The past means the time when an entity change take place in reality, and the final means the time that records the entity change in database. In general, we only take the system time into account in GIS. In STDW, we also use the system time to mark the historical changes of entity. There are three methods to depict the changes. When one or more objects change in a occasion, 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 database table. By comparing theses methods, the conclusion can be drawn 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 data in the same record. Researches on the STDM have made great progress in recent years.

2.2 Spatio-temporal Data Storage

The reason of applying STDW is to make the mass data sharing in the broadest scope. Since users have a variety of data demands, they have various requests towards STDW. To meet most users' demands and achieve fast responds, a multi-layer strategy is used to store spatio-temporal information. Commonly, there are three layers: Data Market, Department Warehouse and Whole Warehouse. Data Market is result 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. SPATIO-TEMPORAL DATA WAREHOUSE FRAMEWORK

The functions of STDW should include data analysis besides data process and data storage. When users apply certain kind of data services, STDW locates the appropriate 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 STDW is composed of Data Sources, Management System and Application Tools. Fig 1 shows the framework of STDW.

3.1. Data Sources

3.1.1. Distributed spatio-temporal databases

These databases are the information source of STDW, 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 various, but also the encoding specifications, projection systems, data formats and etc of them are in diversity.

3.1.2. Extraordinary databases

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

3.2. Data Management System

3.2.1. Data processing

Because the databases in existence provide a variety of applications, there are many differences between each other in data capture methods, encoding specifications, projection systems, data organization standards and data formats, and there are even data mistakes in some databases [2]. Therefore must be some data processing before putting them into STDW, which includes data conversion, spatial transformation and data. Data exchange means unifying the data encoding and structure, adding the temporal mark, operation and semantic exchange of data sets. Spatial transformation means unifying the data coordinate and scale. Data clean 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. STDW 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 union and data decomposition. Data integration means the overlay of multiple data that all the data remain their own characteristics after overlaying, for an e.g. using DOM and digital map to produce image map. Contrasting data integration, data union may create a new kind of data, such as the pseudo color-composite image. Data decomposition means the organization of various geographic features, for an example, the city frame data for the department of estate management which is composed of boundary, street and water system.
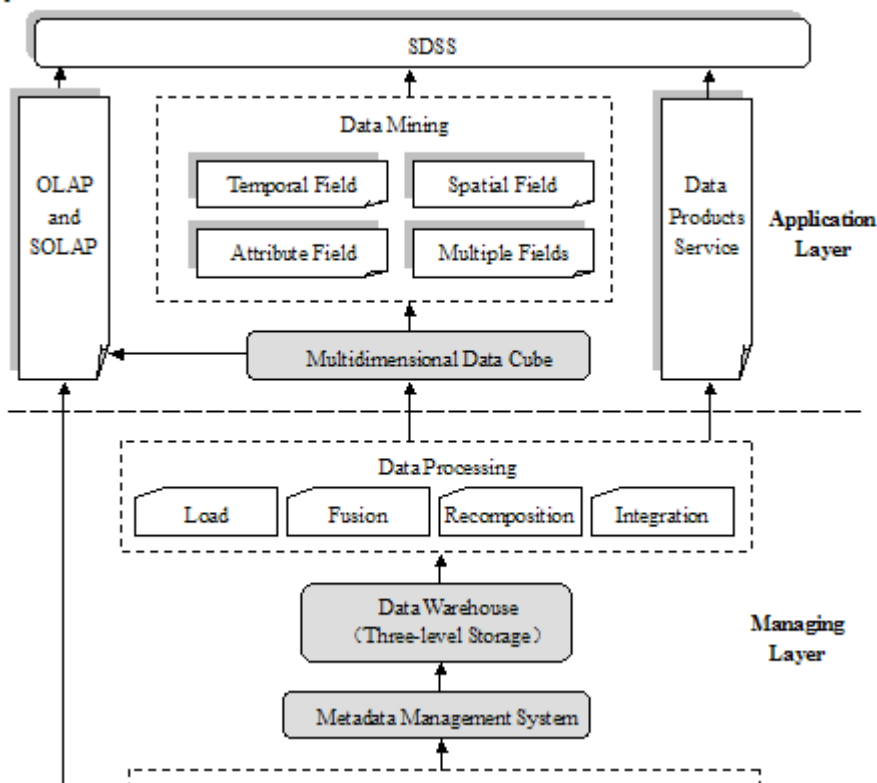
Fig 1: A Framework for Spatio-Temporal Data Warehouse

3.2.2. Metadata Management System

The data about data which is metadata is the backdrop information that explains the content, excellence, 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 (FGDC) standard for Digital Geospatial Metadata, the European organization responsible for standards CEN/TC 287. According to these standards, spatial metadata should include eleven categories information, which are identification, data quality, maintenance, spatial representation, reference system, entity and attribute , distribution, metadata reference, citation, time period 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

STDW can supply a variety of data products service through data integration, data union and data decay, such as the on-line query, scan and display, the on-line order and distribution of data products and so on. Because the STDW uses multiple dimension method 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.

3.3.2. Data Mining

Data mining is a new knowledge that supports decision by abstracting facts from databases. All over the

world there are many researchers who devote themselves to it. A prototype system is implemented based on MapInfo, which has many analysis methods of data mining including spatial comparison, spatial correlation, spatial clustering, spatial classification and so on. The 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.

3.3.3. OLAP and SOLAP

OLAP (On-Line Analytical Process) technology enables users to quickly study mass data. OLAP systems are generally based on a three-tiers architecture including a DW 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. 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 DW. Sometimes the spatio-temporal data with the special data by the unique key word in OLAP and SOLAP, which is could be the organize, place name, project number and so on in both the spatio-temporal data and the special data should be associated. 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 assist the decision-making.

4. CONCLUSION

Since the spatio-temporal data is so difficult and dynamic, the existing DW technology can't solve all the matters of STDW. The research on STDW is still in elementary stage. In this paper, a framework of STDW forms three tiers: data, management and application, which covers the whole process from building warehouse to supplying services. A deeper study on STDW technology which includes the data processing, metadata management, data storage, data mining and so on, and perfect the framework in carry outs.

ACKNOWLEDMENTS

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).

Author Biographies

Dr. C.Sunil Kumar did his B.E in Computer Science and Engineering from University of Madras, Vellore, India, in 1998, M.Tech in Computer Science and Engineering from SRM University, Chennai, India, in 2005. He is Doctorate holder in Computer Science and Engineering, JNTUH, Hyderabad 2012, India. Currently, he is **Professor & Head of CSE** at **Dhanalakshmi College of Engineering, Anna University, Chennai**, India. He has guided more than 20 M.Tech projects and 40 B.Tech projects. He has 25 research publications at International/National Journals and Conferences. His research interests are Distributed Databases, Data warehousing and Data Mining and Information Retrieval Systems.

Dr.A.Govardhan did his BE in Computer Science and Engineering from Osmania University College of Engineering, Hyderabad in 1992, M.Tech from Jawaharlal Nehru University, Delhi in 1994 and Ph.D from Jawaharlal Nehru Technological University, Hyderabad in 2003. He is presently a Professor of CSE and Director of Evaluation I/c, JNTUH, Kukatpally, Hyderabad. He has guided more than 120 M.Tech projects and number of MCA and B.Tech projects. He has 160 research publications at International/National Journals and Conferences. His areas of interest include Databases, Data Warehousing & Mining, Information Retrieval, Computer Networks, Image Processing and Object Oriented Technologies.

Mr. M.Laxmaiah is a Research Scholar in JNTUH, Kukatpally, Hyderabad. He is currently working as Professor & Head of CSE Dept in Tirumala Engineering College, Bogarm (v) Keesara (M), Hyderabad, AP, India. He has 15 years of experience in Education and 4 Years of experience in Research field. He has 3 research publications at International Journals. His areas of interest include Databases, Data Warehousing & Mining.