자연수림 및 천연자원의 데이터 웨어하우스를 활용한 관리 모형

(Effectively Managing Data Warehouse System for Geographic Woods Information)

임태수*, 강석호**

(Tae-Soo Lim, Suk-Ho Kang)

요 약 자연수림과 천연자원에 대한 최신정보의 정확한 관리를 위해서 효과적인 의사결정지원 기법이 필요하다. 본 연구에서는 데이터 웨어하우스에 입각하여 오랜 기간 축적된 운영관리 데이터를 유용한 정보로 변환시키는 방법 및 그 구현에 대해 살펴보았다. 데이터웨어하우스는 데이터의 저장과 분석을 통한 정교하고도 잘 성숙된 데이터베이스 기술을 사용하므로 전략적 의사결정을 위하여 매우 적합하다고 판단되었다. 이러한 데이터웨어하우스 모델링을 활용하여 운영계 데이터를 재가공하고, 통합하고, 저장하여 조림과 천연자원에 대한 통합적 관리에 있어서 효과적으로 적용한 연구이다.

주제어 : 데이터 웨어하우스, 삼림자원 모델링, 통합자원관리시스템

Abstract A data warehouse (DW) is known to be appropriate for strategic decision-making, since it is based on large scale and mature database technology that stores and analyzes data to aid decision support. This paper introduces DW technique, which provides an enterprise solution for those companies that have collected a lot of operational data over the years and need to develop a way to turn that data into useful information. In order to manage current and accurate information about woods resources, woods organizations need to have effective decision-making techniques. This paper effectively applies DW modeling process, in which operational data is re-processed, aggregated and stored in base tables, to an integrated woods resources management

Keywords : Data warehouse, woods resources modeling, aggregated information management

1. Introduction

It is operated as a working forest whose basic function is to produce timber efficiently at sustainable levels. Within this context, the production process follows certain business rules, which influence the type of data model that is applied. GIS data such as vector maps and raster images, aerial photography and other remotely sensed data to aid in carrying various silvicultural and harvesting functions. These data are stored in various electronic and hard copy formats. It was for this reason that a decision was made to create a more modern and centralized database that would facilitate storage and retrieval of forest resource data and create a resource for woods strategic planning and decision-making. Woods companies

* LG생산기술원 설계지원그룹 선임연구원

** 서울대학교 산업공학과 교수

\+ 논문접수 : 2004년 8월 10일

\+ 심사완료 : 2004년11월 10일

ISSN 1598-9798

are in the business of logging and renewing woods resources. The law obligates them to renew the forest following harvesting through approved forest management plans.

Within these management plans, the forest practitioners based on specific objectives select various operational and silvicultural options. Forest management decisions are made on the basis of past and future trends in timber production.

In order to make effective strategic decisions about forest resources management, woods organizations need to have up-to-date and accurate information about their natural resources. This includes implementing modern and centralized databases for day-to-day management of operational resource information (timber and non-timber) and spatial data warehouses for analytical and decision support. Investments in major database management and reporting systems in woods is critical to achieving strategic goals and to making the data readily available to decision makers.

Typically, woods companies are unable to effectively manage their resources or utilize their data and knowledge bases because data are scattered across several different media in either electronic or hard copy formats. Those data that are in electronic format are supported by applications that may not be compatible and are therefore can be difficult to access or cross-reference. To compound the problem further, several different individuals maintain these documents in various departments of the organization and as such there is little or no crossover of information among departments, thus creating a very inefficient method of data extraction, transformation and processing.

Pratt and Adamski (2000) summed up a data warehouse as containing read-only views of extracted, highly aggregated and summarized data from multiple internal and external sources that are refreshed periodically.

Geospatial data warehousing is a technology that has come of age particularly in the last decade. It differs from the traditional non-spatial data warehouse in complexity in that the former requires the integration of both the spatial and the non-spatial components of the data.

Historically, data warehouses were built for and by businesses that did not have a spatial component. Data for the data warehouse were pooled from various sources and integrated for analytical processing. The sources were also traditional database with structured or unstructured data.

Technological advancement particularly in the field of geographic information systems (GIS) that integrate both spatial and non-spatial data, has led to an influx of geospatial processing of data to yield spatial information. The remarkable improvement of accessibility and usefulness of spatial data, organizations have begun to develop data warehouses that have become integrated repositories of geospatial knowledge.

2. Related Works

Confusion still exists about the meaning of a spatial data warehouse. The meaning varies depending on the application of the data contained therein. The key concepts in the definition of the traditional data warehouse are that data is extracted, summarized and stored in an easily accessible electronic environment.

Geospatial data warehouses that have cropped up to store data with a spatial component do not exactly conform to the true definition of a data warehouse as defined by Inmon (2002). The past trend has been to build geospatial storehouses that do not have analytical capabilities from a historic perspective.

In this paper, a data warehouse is a structure that contains all the elements four elements contained in the original definitions, though we recognize that variations exist. Barclay, Gray and Slutz (1999) constructed the Microsoft TerraServer, considered the largest public repository of spatial data consisting of a collection of high resolution aerial, satellite and topographic data and called a spatial data warehouse. The TerraServer as designed can be accessed simultaneously by thousands of users. Although the TerraServer manages a huge database, it differs from the classical data warehouse in one key point in that it does not summarize data to yield new trends or view points. Rather, it is a static collection of spatial data that can be accessed simultaneously accessed by thousands of users, hence behaving more like a spatial database than a data warehouse.

GeoStor (Limp, 2002 and Meredith et. al. 2002) is a web-based data search and delivery distributed database architecture that supports multiple enterprise applications in Arkansas. The terms spatial data warehouse and spatial database interchangeably, though their architecture strongly suggests a combination of a spatial database and geographic data clearinghouse. It is a centralized collection of geographic data with multi-user access as well as enabled for data discovery, visualization, re-projection, reformatting and down.

3. DW Modeling Approaches for Woods

Previous works related to the topics of geospatial modeling by using ER take two main approaches: (a) some either use ER or its extensions proposed for business applications to capture spatial semantics [7] and (b) others extend ER based on the research of some spatial characteristics [3]. Conceptual modeling is discussed in indexing in. The constraint database approach can also be used to describe spatial as well as spatiotemporal data.

The logical designs such as GeoRelational Data Model (GRDM) are presented at [9, 19]. The GRDM is an extension of the relational, based on the spatial and temporal needs at this stage of design. It provides a language for the definition of: (a) relations, used for non spatial entities and relationships, b) layers, which represent space varying attributes, c) object classes. To capture spatiotemporal information, the concepts are transformed into specific constructs of semantic models as an extended relational model to accommodate the needs of the spatial and temporal domain.

One of the fundamental mysteries and obstacles in designing a generic spatio-temporal GIS is its data model. Incorporating time and space in GIS data models increases the complexity of the data structure and has been challenging task [15]. Application oriented modeling will be more efficient if it is based on a generic model.

In traditional databases, matching is a binary operation: every item either matches the query or not [18]. Zhou et al. [22] were the first ones to propose a framework for spatial data OLAP. They extended the star schema to cube dimensions in both spatial and non-spatial and the measures can be regions in space, in addition to numerical data. They concentrate on the spatial dealings and suggest a method for selecting spatial objects for materialization. An efficient IO measure is addressed for merging spatial objects. The method is applied on aggregations for spatial measures [22]. Vagueness with a vague graph based on fuzzy graphs was raised in the particular context of spatial information [5].

The contributions of the current study is to develop a spatio-temporal data warehouse framework in respect to decision support in enterprise woods. Based on this objective, the goal is to design an integrating operational and analytical data services, and to build an operational object relational geodatabase, and decision support data warehouse.

Data warehousing provides an enterprise solution for those companies that have collected a lot of operational data over the years and need to develop a way to turn that data into useful information. A data warehouse is sophisticated technology infrastructure that stores and analyzes data to aid decision support. It is the next level beyond a database. Data warehousing is a process in which operational data is re-processed, aggregated and stored in base tables. Inmon (2002) who coined the term data warehouse, described the salient aspects of a data warehouse as a collection of data for decision support processes that is subject-oriented, integrated, non-volatile and time-variant:

4. Representing Models for the Woods

At the core of the spatial data warehouse architecture is an object-relational database model that manages and integrates both spatial and non-spatial data. The database data model is so design that there is a seamless transfer of data between the database and the data warehouse. The infrastructure of the SDB and SDW is based on the flexible relational database design language with an object-oriented data model underpinning.

A two-step approach was followed in the design process of the MKRF database. The first step was the information-level design in

(Table 1) Description of four terms contained in the definition of data warehouse

Aspects of a data warehouse	Explanation of terms for the Woods
Subject oriented	Data is grouped into major subject areas or entities rather than by application.
Integrated	This is the most important most important aspect. It brings disparate data into a cohesive whole (data warehouse), even though it originated from different departments in the organization and external sources. Data is converted, reformatted, re sequenced, summarized etc. when it passes from operational to data warehouse.
Non volatile	In regular database environment, data regularly accessed and updated. In data warehouse, data loaded periodically en mass and in a snapshot fashion (historic data). Data is read only and users cannot update data.
Time variant	Means that data in a warehouse represents snapshots and is accurate at various points in time, ie. Sophisticated series of snapshots with historic sequence of events or activities.

which the conceptual design was developed independent of a particular DBMS that will be ultimately used. In the second step, which is called the physical-level design, the conceptual design was transformed into a logical and physical design.

The building blocks of the DB are object and methods. In the woods sector, the object could be a forest cover polygon or cutblock. The methods are the actions or activities that can be imposed upon the object. The objects and the methods are connected together through a series of relationships (see Figures 1, 2, and 3).

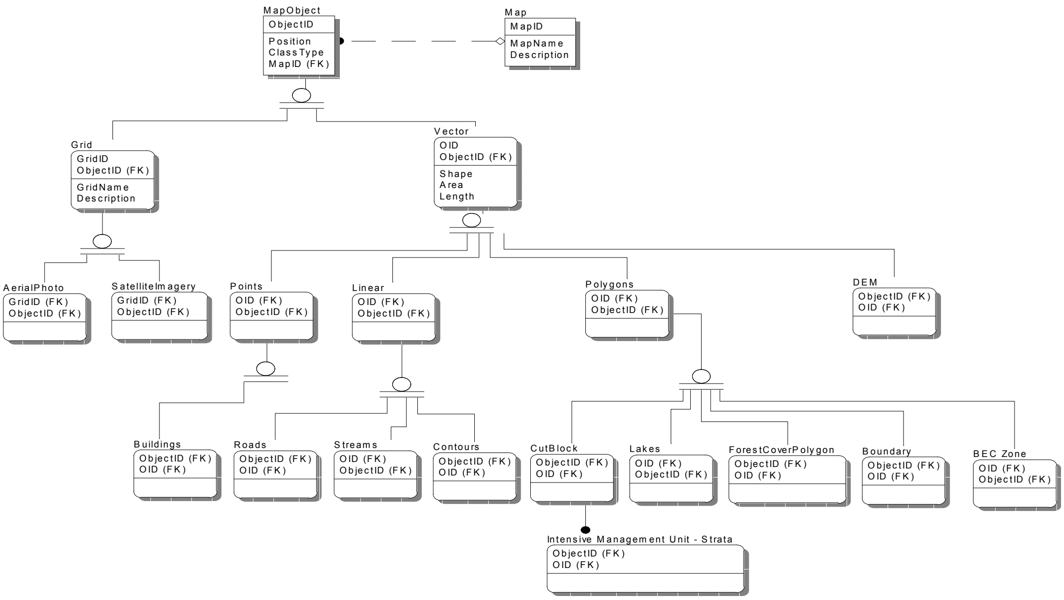
5. Implementation of DW model

With the same generic data model, we can generate woods management points of view based on activity options such as harvesting a cutblock, and prescribing management activities to rehabilitate the harvest unit, such as a planting a desired tree species, and conducting plant growth surveys.

Silviculture Management Activities allow for the planning and creation of silviculture prescriptions and tracking silviculture management activities at the cutblock or strata level. This module is linked to cutblock and strata

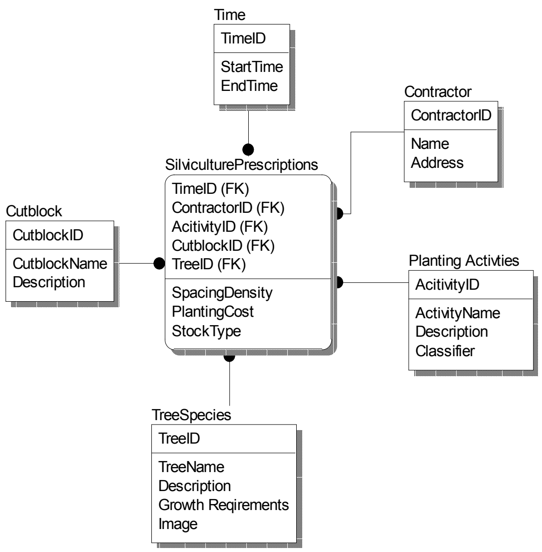
(Fig 1) Generic Model for Forestry DW Approach

(Fig 2) Map interface of the DW model



(Fig 3) A Map Schema of Spatial Objects

map layers. An illustration of a generic model on silviculture prescription activity is given in Fig 3. In the model, all the dimension factors are engaged in the instances.



(Fig 4) Modeling for Silviculture Management Activity

Timber Production Activities - this module is used to track actual results of harvested logs from cutblocks. The Timber Production Activities module also allows for the management and tracking of actual logs from the point of harvest through to delivery at a mill. This involves activities related to contractors, falling and hauling of logs and receipting scaled trucks at a log inventory location, and management of inventory.

Harvesting Activities - aid in the pre-harvest activities, proposed and approved cutblock forecast and in the logging and delivery schedule planning. Shorter-term planning such as the creation of logging and hauling plans, the monitoring of annual cut statistics for cut control management and the capture of inventory attribute and spatial data over time (broad-based, cruise, estimates, production) for the control of inventory.

Maintenance Activities - This module monitors the planning and implementation of pre- and post-planting activities in cutblocks. It allows for the creation, maintenance, digitizing of cutblocks and strata objects. It includes the management functionality for all cutblock activities.

Surveys - These surveys monitor inventories of timber from pre-harvest to post-planting levels. The surveys range in size from micro-scale (sub-management unit level) to a macro-scale whole licensed area. Survey methods utilize various techniques and tools, include remotely sensed and photogrammetry methods to field data collection scale.

Road Management Activities - provides mechanism for managing forest access and related activities. The module allows for the creation and maintenance of roads and road structures. It includes tracking and reporting of road construction, inspections, maintenance, deactivation and access control activities, as well as map display of roads.

6. Conclusion

The DW Modeling is considered an architectural framework for the Woods domain. Henceforth the model can maintain both current and historical data and it is scalable to handle a sheer volume of data that may be collected from various data sources and users. Maintaining both current and historic data in the same database structure requires a non-conventional spatially enabled object-relational data warehouse that integrates spatial, temporal, and relational objects.

In this paper, we have developed a generic model for woods management that can be applied to different management operations. The model has a fact table with five dimensions. Using the generic model, an interface has been developed that creates links to spatial, temporal and relational operators. This interface is simple and yet powerful enough to organize and maintain the integrity of the data in the Woods data.

Reference

[1] M. Akinde, M. Böhlen, T. Johnson, L. Lakshmanan, D. Srivastava: Efficient OLAP Query Processing in Distributed Data Warehouses. EDBT 2002: 336-353

[2] M. Bohlen, C. Jensen, and B. Skjellaug, Spatio-Temporal Database Support for Legacy Applications.

[3] A. Belussi, M. Negri, G. Pelagatti: An integrity constraints driven system for updating spatial databases. ACM-GIS 2000: 121-128

[4] Andrey Balmin, Thanos Papadimitriou, Yannis Papakonstantinou: Hypothetical Queries in an OLAP Environment. VLDB 2000: 220-231

[5] M. Erwig and M. Schneider. *Vague Regions*. In 5th Int. Symp. on Advances in Spatial Databases, LNCS 1262, pages 298–320. Springer-Verlag, 1997.

[6] Martin Erwig, Markus Schneider: Developments in Spatio-Temporal Query Languages. DEXA Workshop 1999: 441-449

[7] R. Fernández, M. Rusinkiewicz: A Conceptual Design of a Soil Database for a Geographical Information System. International Journal of Geographical Information Systems 7(6): 525-539 (1993)

[8] T. Hadzilacos, N. Tryfona: An Extended Entity-Relationship Model for Geographic Applications. SIGMOD Record 26(3): 24-29 (1997)

[9] T. Hadzilacos, N. Tryfona: Logical Data Modelling for Geographical Applications. International Journal of Geographical Information Science 10(2): 179-203 (1996)

[10] http://www.mkrf.woods.ubc.ca/

[11] Evaggelia Pitoura, Panos K. Chrysanthis, Krithi Ramamritham: Characterizing the Temporal and Semantic Coherency of Broadcast-Based Data Dissemination. ICDT 2003: 410-424

[12] M. Pang, and W. Shi: Development of a Process-Based Model for Dynamic Interaction in Spatio-Temporal GIS. GeoInformatica 6(4) 2002: 323-344

[13] C. Parent, S. Spaccapietra, and E. Zimanyi: Spatio-Temporal Conceptual Models: Data Structures + Space + Time. In Proc. ACM GIS 1999: 26-33

[14] D. Pfoser, N. Tryfona: Requirements, Definitions, and Notations for Spatiotemporal Application Environments. ACM-GIS 1998: 124-130

[15] A. Raza, and W. Kainsz: Cell Tuple Based Spatio-Temporal Data Model: An Object Oriented Approach. In Proc. ACM GIS 1999: 20-25

[16] S. Santini, A. Gupta: Conceptual Integration of Multiple Partial Geometric Models. ER 2002: 365-379

[17] Nebojsa Stefanovic, Jiawei Han, Krzysztof Koperski: Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes. TKDE 12(6): 938-958 (2000)

[18] S. Santini, R. Jain: Image Databases Are Not Databases with Images. ICIAP (2) 1997: 38-45
[19] N. Tryfona, T. Hadzilacos: Logical Data Modelling of Spatio Temporal Applications: Definitions and a Model. IDEAS 1998: 14-23
[20] G. Vert, A. Morris, M. Stock: Converting a Fuzzy Data Model to an Object-Oriented Design for Managing GIS Data Files. TKDE 15(2): 510-511 (2003)
[21] G. Vert, M. Stock, and A. Morris: Extending ERD modeling notation to fuzzy management of GIS data files. DKE 40(2): 163-179 (2002)
[22] X. Zhou, D. Truffet, J. Han: Efficient Polygon Amalgamation Methods for Spatial OLAP and Spatial Data Mining. SSD 1999: 167-187

임 태 수
1990년 : 서울대학교 산업공학과 (학사)
2004년 : 서울대학교 산업공학과 (공학박사)
2004년~현재 : LG생산기술원의 설계지원그룹 선임연구원
관심분야 : e-비즈니스 표준화, BPM (Business Process Management), 생산/설계 지식정보 시스템

강 석 호
1970년 : 서울대학교 (학사)
1972년 : University of Washington (공학석사)
1976년 : Texas A&M University (공학박사)
1976년~현재 : 서울대학교 공과대학 교수
1991년~1997 : 한국 O.R.학회 부회장/감사/차기회장/회장
1999년~현재 : 아시아 태평양 O.R.학회 부회장
1996년~현재 : 산업자원부 자동화 부문 심사위원
관심분야 : MIS, BPM, E-business