Geo-Data Fusion Integrator for Object-Oriented Spatiotemporal OLAP Cubes

Chun-Kit Ngan
Engineering and Information Science Division
Great Valley School of Graduate Professional Studies
The Pennsylvania State University
Malvern, PA, USA
cxn20@psu.edu

Abstract— We first propose a Geo-Data Fusion Integrator. Specifically, we design a sequential-parallel-modularized (SPM) approach to integrate different datasets into a geo-data object, i.e., a multidimensional unified-OLAP cube, archived in a geo-data warehouse for decision-making analysis. Different datasets of geo-data objects are processed in parallel across multi-stages in sequence, and then integrated into a well-defined OLAP cube. Each SPM component is a self-contained, modularized unit that processes the data. The technical merits of this SPM approach include fast manipulations, error minimization, and easy maintenance. Second, to create a unified geo-data object, we extend the object-oriented spatiotemporal data model as a multidimensional OLAP cube, i.e., a Star-based Geo-Object-Oriented Spatiotemporal (S-GOOSE) data model, which combines the advantages of both OLTP and OLAP approaches. This S-GOOSE data model is an object-relational-based cube that enables military operators to analyze unified geo-data objects from multiple dimensions, such as time, space, and location, to help them make a better decision on paths.

Keywords— data processing; data integration; data modeling; OLAP cube

I. INTRODUCTION

In the past decade, military units in nations successfully analyzed integrated geo-data objects to execute decision-making operations. Consider one critical operation [1] that military troops need to determine is to find the shortest path to reach a target location for rescue and recovery missions so that medical and healthcare services can be efficiently delivered to victims who can receive supports and supplies in the aftermath of natural disasters. To support such a decision-making operation, it is important for military researchers to develop a geo-data analytical methodology that is reliable and useful for the operation. The main challenge in such a methodology is how to accurately, expressively, and effectively integrate and model these geo-data objects such that diverse military units can analytically make a concerted decision together. This is exactly the focus of this paper.

Currently, a number of data integration approaches has been proposed and developed to combine different sets of geo-data objects into a single composite unit. Chen, et al. [2] proposed a geospatial information integration approach, named Automatic Multi-Source conflation, to automatically integrate vector data and imagery together. Shahabi, et al. [3] proposed a Geospatial Decision Making system that applies the relevant methodologies developed in the fields of databases, artificial intelligence, computer graphics, and computer vision to integrate satellite imagery, road data, maps, point data, and temporal data into unified geo-data sets. However, these techniques require significant quality and accuracy of each dataset, such as spatial consistencies, high pixel resolutions, data completeness, time precision, and location exactness, to generate an accurate, unified view of data. If one of these datasets is corrupted, or some data is missing, the integrated geo-data objects are inaccurate and unsuitable for decision-making analysis.

Another challenge is to model unified geo-data objects that require a precise expression to depict their properties and relationships. Presently, a number of researchers has proposed and developed different approaches to model geo-data objects, which can be roughly divided into two categories: On-Line Transaction Processing (OLTP) and On-Line Analytical Processing (OLAP). The former OLTP approach mainly focuses on facilitating and managing transaction-oriented applications that are used for data entry and retrieval processing. For example, Li and Cai [4] proposed an OLTP-based object-oriented spatial temporal data model, from which geo-data objects can be described by a theme, space, and time. Although this data model is efficient for a fast transactional processing analysis, it is not designed for supporting decision-making analytics. The latter OLAP approach enables military operators to analyze multidimensional data interactively from multiple perspectives. For instance, Vaisman and Zimanyi [5] proposed a MultiDim model based on a Spatial On-Line Analytical Processing (SOLAP) cube that can enable military operators to use multi-dimensional views to analyze geo-data objects, from which the operators can gain a more complete decision-making insight. However, since this spatial OLAP solution does not model real-world objects’ properties and operations based on the object-oriented paradigm, it is hardly suitable for modeling objects’ interactions.

II. OUR METHODOLOGY

Thus this paper focuses on bridging the above research gaps, i.e., the geo-data integration and modeling. First, we propose a Geo-Data Fusion Integrator (GDFI) shown in Fig. 1 of Appendix. Specifically, we design a sequential-parallel-modularized (SPM) approach to integrate different datasets
into a geo-data object, i.e., a multidimensional unified-OLAP cube, archived in a Geo-Data Warehouse (GDW) for decision-making analysis. Different datasets of geo-data objects are processed in parallel across multi-stages in sequence and then integrated into the well-defined OLAP cube. Each SPM component is a self-contained, modularized unit that processes the data. The technical merits of this SPM approach are summarized in the following: (1) Fast manipulations, as multiple dataset objects are processed simultaneously; (2) Error minimization because a logical progression in the stage-by-stage process guarantees that the next-stage processing units receive a valid data input for the final cube generation; and (3) Easy maintenance in that each self-contained, modularized component enables domain experts to discover and fix faults.

Second, to create a unified geo-data object, we extend the object-oriented spatial-temporal data model as a multidimensional OLAP (star-schema) cube, i.e., a Star-based Geo-Object-Oriented SpatioEmporal (S-GOOSE) data model, which combines the advantages of both OLTP and OLAP approaches. This S-GOOSE data model is an object-relational-based cube that enables domain experts, e.g., military operators, to analyze unified geo-data objects from multiple dimensions, such as time, space, and location, to help them make a better decision on paths.

III. GEO-DATA FUSION INTEGRATOR

Technically, this GDFI integrator is composed of six types of processing units, including an intermediate storage area, vectorizers, cleansers, group mediators, a centralized mediator, and a data cube layer, which operate in six respective stages, to create a S-GOOSE OLAP cube archived in a geo-data warehouse.

In general, the GDFI transforms the same type of datasets from multiple military units’ data sources to the same data format, i.e., vector data, and then integrates them together into a single unit. The main purpose of this process is to create a single integrated dataset from multiple data sources of the same data type. The reasons to represent all the data in a vector format instead of other possible data types are because (1) the vector data structure produces a smaller file size in that each file stores numerical values; (2) the vector data is easier to be processed on a computer, as it has fewer data items and is more flexible to be adjusted for different scales; and (3) the graphical objects are much easier to be represented using vector forms, since commonly shared edges are defined as polygons. At the end, the GDFI integrator creates a S-GOOSE OLAP cube, based upon all the vectorized data created from multiple data types and data sources of diverse military units, to support decision-making analysis. More specifically, the six stages are described as follows:

Stage 1: The intermediate storage area is a temporary location that assembles all the geo-data objects collected from different military units’ data sources. The data objects may include temporal data, satellite images, raster data, 2D/3D objects, vector data, location data, and non-spatial information. The data objects are classified into the four categories used for building the dimensions of a S-GOOSE OLAP cube. The dimensions include (1) TimeDim for supporting time-based analysis at different periods, (2) VisualDim for visualizing objects’ shapes, (3) LocatDim for depicting objects’ location and position, and (4) ObjDim for describing geo-data objects’ contents and properties.

Stage 2: Vectorizers transform the geo-data objects received from Stage 1 to a standardized vector data format. Each vectorizer is incorporated with an existing, specific vectorization algorithm (e.g., [6]), based upon the data objects being processed, to perform the data transformation.

Stage 3: After the vectorization, the data objects are sent to their corresponding cleansers (e.g., [7]) to remove dirty data that may be misspelled, corrupted, inaccurate, or inconsistent.

Stage 4: Using the methodology proposed by [8], each group mediator takes advantage of the same data type from different data sources to complement one another to create an integrated geo-data object.

Stage 5: Once the centralized mediator receives the unified geo-data objects from all the group mediators, it conflates the received objects into a single vectorized unit by using the [8] approach and then creates a S-GOOSE OLAP cube based upon the dimensions of interest, e.g., time, location, visualization, and other objects’ descriptive non-spatial information.

Stage 6: The S-GOOSE OLAP cubes temporarily reside in a staging area until the GDW is ready to receive them for the permanent storage. The S-GOOSE OLAP cubes are then sent to the data cube layer for further processing. This data cube layer segregates the S-GOOSE OLAP cubes into the smaller cubes according to the different themes such as buildings, trees, and mountains. All the smaller cubes are also archived in the GDW for the future decision-making analysis.

Towards the end, the domain experts can use the Geo-Query Language (GQL), i.e., Object-Relational Query [9], to construct different views, e.g., reporting, simulation, visualization, and decision optimization [10], to support the military operations, e.g., determining the shortest path for rescue and recovery missions.

IV. S-GOOSE DATA MODEL BY EXAMPLE

The S-GOOSE data model is an extension of the object-relational database model with a specialized schema, i.e., a S-GOOSE OLAP cube with a number of dimension schemas shown in Fig. 2 of Appendix. The dimension schemas include Time, Location, Visualization, and Building Object in a terrain over a time horizon. More specifically, the S-GOOSE database schema is a set of object-relational schemas, which include a number of S-GOOSE OLAP cubes. For an elucidation purpose, let us reconsider the military case of determining the shortest path to reach a target building for a rescue and recovery mission after a natural disaster. To demonstrate the concepts, we assume that there is a building layer in a disaster region and the military operational unit passes through a sequence of buildings in an area to rescue a group of victims, who are trapped inside the target building.

Using the extended S-GOOSE OLAP cube of the building layer, the military forces can (1) create the S-GOOSE tables and views, including reporting, simulation, visualization, and decision optimization, (2) store the tables and views with the
data in the GDW, and (3) execute the views to perform the analysis in different combinations of dimensions.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we first propose a GDFI integrator. Specifically, we design a SPM approach to integrate different datasets into a geo-data object, i.e., a multidimensional unified-OLAP cube, archived in a GDW for decision-making analysis. Different datasets of geo-data objects are processed in parallel across multi-stages in sequence, and then integrated into the well-defined OLAP cube. Each SPM component is a self-contained, modularized unit that processes the data. The technical merits of this SPM approach include fast manipulations, error minimization, and easy maintenance. Second, to create a unified geo-data set, we extend the object-oriented spatial-temporal data model as a multidimensional OLAP cube, i.e., a S-GOOSE data model, which combines the advantages of both OLTP and OLAP approaches. This S-GOOSE data model is an object-relational-based cube that enables military operators to analyze unified geo-data objects from multiple dimensions, such as time, space, and location, to help them make a better decision on paths. There are still many open research questions, particularly how to present and visualize integrated data objects to support decision-making.

REFERENCES


APPENDIX

Fig. 1. Geo-Data Fusion Integrator.

Fig. 2. S-GOOSE OLAP Cube for the Building Layer.