Chapter 1
Current Approaches, Challenges, and Perspectives on Spatial OLAP for Agri-Environmental Analysis

Sandro Bimonte
National Research Institute of Science and Technology for Environment and Agriculture (IRSTEA), France

ABSTRACT

Spatial OLAP (SOLAP) systems are powerful GeoBusiness Intelligence tools for analysing massive volumes of geo-referenced datasets. Therefore, these technologies are receiving considerable attention in the research community and in the database industry as well. Applications of these technologies are current in several domains such as ad marketing, healthcare, and urban development, to name a few. Contrary to other application domains, in the context of agri-environmental data and analysis, SOLAP systems have been underexploited. Therefore, in this paper, the author makes an exhaustive survey of most of the published studies in the domain of the SOLAP analysis of agri-environmental data with an emphasis on the reasons why only few recent works investigate the use of SOLAP systems in the agri-environmental context. In particular, the author focuses on the complexity of the spatio-multidimensional model and its implementation. Moreover, based on surveying the state of the art in this domain, this paper identifies some general guidelines that must be considered by the scientific community to design and implement efficient SOLAP approaches to the analysis of geo-referenced agri-environmental datasets. Finally, open issues about warehousing and OLAPing agri-environmental data are also shown in the paper.

1. INTRODUCTION

The concept of On-Line Analytical Processing (OLAP) was presented by Codd et al. (2013), who largely demonstrated that On-Line Transaction Processing (OLTP) cannot satisfy the expectations of end users on the query analysis of databases and that the use of SQL queries is not an efficient tool to support a user's analysis. Therefore, Codd et al. presented the concept of multidimensional databases

DOI: 10.4018/978-1-5225-7033-2.ch001

and multidimensional analysis, called OLAP. OLAP can be defined as a software technology allowing users to access information immediately, consistently and interactively to gain a deeper understanding of data for different types of analysis. Thus, OLAP technology is one of the driving forces behind Business Intelligence (BI). At the heart of this technology is the concept of "dimension", which represents a point of view or a type of classification of a high level of data considered in the study (i.e., "fact"). In other words, a dimension represents a data attribute (time, location and product type as examples) of the problem under study. Usually, the dimensions are represented as cubes that the user can create, update, manipulate and query. The cube could be considered as an *abstract data type* offering high-level operations on cubes or between cubes such as slice, dice, drill-down, roll-up, and drill-across. Over the years, several variants of OLAP have been introduced such as multidimensional OLAP (MOLAP) and relational OLAP (ROLAP). The main differences between the two technologies concern data storage processing capability and the currency of data.

Spatial OLAP systems present the conventional OLAP systems provided with spatial capabilities to address problems occurring in time and space. Spatial Data Warehouses (SDWs) and Spatial OLAP (SOLAP) systems allow for analysing massive volumes of georeferenced datasets (Bédard et al., 2001). SOLAP has been successfully applied to several domains such as marketing (Glorio et al., 2012) and healthcare (Bernier et al., 2009). Usually, warehoused (spatial) data are cleaned and transformed after being extracted from existing external data sources. These data sources are transactional systems (i.e., databases), text files, paper forms, and sometimes also sensor data (Bédard et al., 2001). Spatial OLAP systems allow decision-makers to explore warehoused (spatial) data using SOLAP operators that aggregate fact data along hierarchies, allowing for selection of a subset of interesting information (Bimonte, 2010). The success of (S)OLAP technologies is mainly based on the capabilities of such systems to: (i) avoid quality problems related to the integration of heterogonous data sources (Jarke et al., 2013), (ii) interactively analyse large datasets, and (iii) provide simple and user-friendly visual analytic interfaces (Bédard et al., 2006). Applications of the SOLAP technology field have caught the attention of the research community. There are publications in different scientific journals addressing this issue (e.g., Pestana et al., 2005, Di Martino et al., 2009, Zaamoune et al., 2013). However, contrary to other application domains, in the context of agri-environmental data and analysis, SOLAP systems have been underexploited. Therefore, in this paper, we conduct an exhaustive survey of published studies in the domain of SOLAP analysis of agri-environmental data with an emphasis on the reasons why only few recent works investigate the use of SOLAP systems in the agri-environmental context. In particular, we focus on the complexity of the spatio-multidimensional model and its implementation. For that goal, we propose a set of new criteria, extending existing (S)OLAP metrics, to evaluate SOLAP applications. Moreover, based on surveying the state of the art and our 10 years of experience in this domain, we identify some general guidelines that must be considered by the scientific community to design and implement efficient SOLAP applications for the analysis of geo-referenced agri-environmental datasets. Finally, we investigate issues about warehousing and OLAPing agri-environmental data that remain unexplored and that we consider to be fundamental for an effective SOLAP analysis of agri-environmental data. To summarize, the main contribution of this paper is to provide readers with a set of guidelines, based on existing works, for the effective design and implementation of SOLAP applications for agri-environmental data.

The paper is organized as follows: Section 2 introduces main concepts of SOLAP and compares SOLAP with Geographical Information Systems (GIS). Section 3 presents a systematic review of the topic. Section 4 presents some open issues, and Section 5 presents some conclusions we have reached in our study.

Related Content