A Report on the Eighth ACM International Workshop on Data Warehousing and OLAP (DOLAP'05)

Juan Trujillo
Dept. of Language and Information Systems
University of Alicante, Spain
Apto. Correos 99. E-03080
{jnmazon,jtrujillo}@dlsi.ua.es

Il-Yeol Song
Drexel University
Colleague of Information Science & Technology
Philadelphia, PA 19104
songiy@drexel.edu

1. Introduction

Research in data warehousing and OLAP has produced important technologies for the design, management and use of information systems for decision support. Much of the interest and success in this area can be attributed to the need for software and tools to improve data management and analysis given the large amounts of information that are being accumulated in corporate as well as scientific databases. However, in spite of the maturity of these technologies, new data needs or applications currently run at companies not only demand more capacity, but also new methods, models, techniques or architectures to satisfy these new needs.

This report focuses on the *Eighth ACM International Workshop on Warehousing and OLAP (DOLAP'05)* held in conjunction with the 14th International Conference on Information and Knowledge Management (CIKM'05) in Bremen, Germany, on November, 1st-5th, 2005. A summary of the accepted papers, the invited keynote speaker and the Panel discussions held at the end of the workshop is given.

In the call for papers, papers focused on new research directions and emerging application domains in the areas of data warehousing and OLAP were especially encouraged. In response to the call for papers, the workshop received 31 submissions from 18 different countries and only 12 papers were selected by the Program Committee, making an acceptance rate of 38.7%.

The accepted papers were organized in five different sessions: (i) querying OLAP databases, (ii) data warehouse models, (iii, iv) data warehouse design, and (v) query processing and view maintenance.

2. Querying OLAP Databases

Pu [1] focuses on the problem of representing OLAP databases and their query language. To this aim, the author first defines a framework based on functional symbols annotated by typing information. Then, once the basic multidimensional database has been defined, query constructs are specified as higher-order polymorphic functions, and queries are expressed as complex functional expressions. The author argues that its query language is flexible enough to represent useful OLAP queries as well as new user defined functions can also be easily expressed. Moreover, the presented equational specifications can help designers to automatically reason about problems such as summarizability of OLAP views. Finally, Pu argues that static type analysis of OLAP queries has not been dealt with existing formalisms.

Ladjel *et al.* [2] argue that in most cases, end users cannot properly query the DW mainly because (i) sometimes the most interesting information is not correctly shown, and (ii) the OLAP answers provided by the corresponding OLAP tools cannot be properly visualized. For this reason, the authors present a framework in which end users (i) specify their information preferences by ordering the different parts of an OLAP query (e.g. dimensions, classification hierarchy levels, and so on), and (ii) define their visualization constraints (mainly imposed by the limitations of the device used for the OLAP queries). The authors argue that their framework takes into account both personalization and visualization at the same time.

3. Data Warehouse Models

Perez *et al.* [3] start by arguing that many data contexts of current corporate data warehouses can be found in both external and internal documents. Thus, the authors propose a Relevance-Extended Multidimensional Model to combine structure of DWs and documents in what they call Data Warehouse Contextualized with Documents. To this aim, authors propose the *R-cube*, a type of OLAP cube based on (i) specifying the relevance of each fact in a query, and (ii) defining the related documents that provide information on the selected facts. In this way, users can query a traditional DW (with the corresponding MD terms) and obtain further information stored in related documents.

Jones and Song [4] claim that designing dimensional models can be complex, costly and time consuming. For this reason, they apply well-known software pattern techniques in the developing dimensional models. Authors identify and classify main dimensional patterns that normally occur in specifying dimensions. Then, they provide a metamodel to help designers in specifying Dimensional Design Patterns (DDP) and apply it to the

design of dimensional models. The initial results of the experiment reveal a significant increase in the efficiency of their students in designing dimensional models.

Bimonte *et al.* [5] present a multidimensional model in order to correctly represent spatial data for data warehouses. Spatial data need complex structures that cannot be correctly managed by traditional data warehouses. Thus, the authors claim that supporting and exploiting the particular nature of spatial data into the MD analysis implies the re-thinking of basic OLAP concepts, such as spatial dimensions or facts, or the logical and physical representation of these spatial dimensions and facts. Therefore, the authors propose a Spatial Multidimensional Model that allows us to represent complex (spatial) objects following the MD paradigm in order to handle geographical data. They show that spatial data warehouses can be modeled based on the traditional conception of the MD paradigm.

4. Data Warehouse Design

Giorgini *et al.* [6] claim that a significant percentage of data warehouses fails to meet business objectives. The authors argue that requirement analysis is typically overlooked in real world DW projects. In this paper, they propose a goal-oriented approach to requirement analysis for DWs, based on the Tropos methodology. The authors integrate in the same analysis requirement approach both (i) organizational modeling, centered on stakeholders, and (ii) decisional modeling, focused on decision makers. The authors argue that their approach can be used in both a demand-driven and a mixed supply/demand-driven design framework. Finally, the authors provide some snapshots of the prototype that implement their methodological approach and apply it to a real case study.

Mazon *et al.* [7] argue that most existing modeling approaches do not provide designers with an integrated and standard method for designing the whole DW (ETL processes, data sources, DW repository and so on). In this work, the authors present a novel approach to align the whole DW development process to MDA (Model Driven Architecture). Then, they focus on the MD analysis and define the MD2A (Multidimensional Model Driven Architecture) as an approach for applying the MDA framework to MD modeling. They first describe how to build the different MDA artifacts (i.e. models) by using extensions of the Unified Modeling Language (UML). Then, transformations between models are clearly and formally established by using the Query/View/Transformation (QVT) approach. Finally, the authors provide an example on how to apply MDA and its transformations to the MD modeling.

Simitsis [8] describes the mapping of conceptual models into logical models for ETL processes. The author starts by identifying how a conceptual entity is mapped to a logical entity. Next, he determines the execution order in the logical workflow by using information adapted from the conceptual model. Finally, he provides a methodology for the transition from the conceptual to the logical model.

Nguyen *et al.* [9] start by arguing that traditional business intelligence (BI) architectures lack the support of real-time BI and closed-loop decision making. In this work, authors present a real-time BI architecture called SARESA. The main aim of the BI architecture is to provide continuous, real-time analytics in order to enable proactive responses to a business environment for effectively managing and controlling time-sensitive business processes. Then, the authors describe the *Sense & Respond* loops and a service-oriented architecture that is able to detect situations and exceptions, perform complex analytical tasks and reflect on the gap between current situations and desired management goals. Finally, they apply it to a mobile phone fraud detection case study.

5. Query Processing and View Maintenance

The work presented by Dehne *et al.* in [10] relies on supporting efficient indexing facilities for M cube queries. The authors argue that the complexity and difficulty of the indexing problem is exacerbated by the existence of attribute hierarchies that sub-divide attributes into aggregation layers of varying granularity. Thus, they present a hierarchy and caching framework that supports the efficient and transparent manipulation of attribute hierarchies within a parallel ROLAP environment. They also provide experimental results that verify that very little overhead is required to handle streams of arbitrary hierarchical queries.

Cuzocrea's work [11] focuses on the efficient execution of approximate answers for OLAP applications. The author claims that the scalability of the query techniques and the accuracy of the answers are recognized as important limitations of state-of-the-art approximate query answering proposals in OLAP. In this paper, Alfredo presents a statistical framework that covers the limitations related to the accuracy of queries. This is done by ensuring the probabilistic bounds on the retrieved answers, and tailored for the specific OLAP context. Within this framework, the author defines the KSyn synopsis data structure, which efficiently supports approximate query answering in OLAP. Finally, Alfredo presents and discusses encouraging preliminary experimental results stating the goodness of his proposal.

Lee and Kim [12] start by defining the multiple view maintenance problem and arguing that materialized views are still commonly used in data warehouse environments. The authors claim that although there has been much work on efficient maintenance of a single view, maintenance of multiple views has not been sufficiently investigated. Then, in this paper they propose an efficient incremental maintenance of multiple join views. Basically, they propose the delta propagation strategy that computes the change of multiple join views in a recursive manner. Then, the authors provide a heuristic algorithm that finds a global maintenance plan for the given views. Finally, they present

an experimental result that shows the efficiency of the proposed method.

6. Keynote address

In DOLAP'05, we had a highly interesting keynote address entitled *"My Favorite Issues in Data Warehouse Modeling"* given by Jens Lechtenbörger [13]. He started by sketching major achievements and trends in conceptual DW modelling and pinpointed open problems along the way. Then, he took a closer look at the overall design process focusing on the transformation of conceptual data warehouse schemata into logical ones, and arguing that there is still a semantic gap between advanced conceptual data models and relational or MD implementations, which needs to be bridged. Finally, he turned to one specific aspect of the DW lifecycle, namely schema changes, and highlighted challenges in DW schema versioning.

7. Panel Discussions

At the end of DOLAP'05 we organized a Panel discussion among all participants and identified new research directions of data warehouses and OLAP technologies. Some ideas came out for possible future collaboration between different research groups. Some of the hot topics identified were DW security, data quality, visualization, ETL processes, distributed DWs, advanced OLAP for business intelligence, web warehouses, DWs for new applications such as XML documents, stream data, spatial or GIS data or biomedical data.

8. Conclusions / Summary

DOLAP'05 continued with the successful series of DOLAP workshops, being an international forum where both researchers and practitioners can share their findings in theoretical foundations, current methodologies, and practical experiences. As seen throughout this report, this year, DOLAP'05 was specially focused on new research directions and emerging application domains in the areas of data warehousing and OLAP. Thanks to the high quality of the presented papers, we were able to host a special issue by extending the best papers in the *Decision Support Systems* journal [15].

9. Acknowledgments

We would like to express our gratitude to the Program Committee members and the additional external referees for their hard work in reviewing papers. We also thank all the authors for submitting their papers and the CIKM2005 organizing committee for all their support. Another gratitude is for our technician Miguel A. Varo, who developed the web site and the review system (http://gplsi.dlsi.ua.es/congresos/dolap05/). This workshop was organized with the support from the following projects: METASIGN (TIN2004-00779), DADASMECA (GV05/ 220) and DADS (PBC-05-012-2).

10. References

[1] K. Q. Pu "Modeling, Querying and Reasoning about OLAP Databases: a Functional Approach", in [14].

[2] L. Bellatrach *et al.* "A Personalization Framework for OLAP Queries", in [14].

[3] J. M. Pérez *et al.* "A Relevance-Extended Multi-dimensional Model for a Data Warehouse Contextualized with Documents", in [14].

[4] M.E. Jones and I-Y. Song. "Dimensional Modeling: Identifying, Classifying, and Applying Patterns ", in [14].

[5] S. Bimonte *et al.* "Towards a Spatial Multidimensional Model", in [14].

[6] P. Giorgini *et al.* "Goal-Oriented Requirement Analysis for Data Warehouse Design", in [14].

[7] J-N. Mazón *et al.* "Applying MDA to the Development of Data Warehouses", in [14].

[8] A. Simitsis. "Mapping Conceptual to Logical Models for ETL Processes ", in [14].

[9] T. M. Nguyen *et al.* "Sense & Response Service Architecture (SARESA): An Approach Towards a Real-time Business Intelligence Solution and Its Use for a Fraud Detection Application", in [14].

[10] F. Dehne *et al.* "Parallel Querying of ROLAP Cubes in the Presence of Hierarchies", in [14].

[11] A. Cuzzocrea. "Providing Probabilistically-Bounded Approximate Answers to Non-Holistic Aggregate Range Queries in OLAP", in [14].

[12] K.Y. Lee and M.H. Kim. "Optimizing the Incremental Maintenance of Multiple Join Views", in [14].

[13] Jens Lechtenborger. "Issues in Data Warehouse Modeling", in [14]

[14] J. Trujillo, I-Y. Song (eds.). "Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005.

[15] *Decision Support Systems* journal. Ed: Elsevier. ISSN: 0167-9236. In press.