Guest editorial: special issue on physical data warehouse design

Ladjel Bellatreche

Since fifteen years ago, data warehousing and On-Line analysis (OLAP) technologies proved their effectiveness in managing and analyzing huge amounts of structured and semi structured data. A data warehouse is usually queried in various combinations involving many tables. The most used operations are joins, aggregations and selections. Joins are well known to be expensive operations, especially when the involved relations are substantially larger than the size of the main memory which is usually the case of business intelligence applications. Without efficient optimization techniques, such queries may take hours or days, which is unacceptable in most cases. As a consequence, the physical design needs to address the challenges in efficient processing of complex decision support queries [1]. To optimize these queries, in addition to the existing ones (developed for OLTP applications), a large spectrum of optimization techniques were proposed in the literature and mostly supported by commercial database systems (ORACLE, SQL Server, DB2, etc.). We can cite materialized views, partitioning, advanced indexing schemes, parallel processing, etc. The aim of this special issue of the Distributed and Parallel Databases is twofold. First is to present new and challenging issues in the field of physical data warehouse design (in centralized, distributed and parallel environments). Secondly is to present the current research and technological development in this field.

There was great response to the call for papers, but due to the limited space only five papers were accepted and selected for this special issue. These papers are authored by an outstanding roster of experts in their respective fields, and tackle various issues from different angles, requirements and interests. Their topics include data partitioning (relational and XML), processing queries on private data in untrusted environments, parallel processing, load balancing management. The evaluation process

L. Bellatreche (✉)
LISI/ENSMA–Poitiers University, Poitiers, France
e-mail: bellatreche@ensma.fr

was done by the following reviewers: Ashraf Aboulnaga (Canada), Zohra Bellahsene (France), Ladjel Bellatreche (France), Omar Boussaid (France), Alfredo Cuzzicrea (Italy), Vivekanand Gopalkrishnan (Singapore), Jérôme Darmont (France), Karla-palam Kamalakar (India), Sanjay Madria (USA), Mukesh Mohania (India), Stefano Rizzi (Italy), David Taniar (Australia), Panos Vassiliadis (Greece).

The five selected papers are summarized as follows:

The first article *Revelation on Demand* by Anciaux, Benzine, Bouganim, Pucheral and Shasha presents a framework for processing queries on private data in untrusted environments. The basic idea is that the private data resides on a secure USB key, and never leaves the USB key except in aggregated form. Hence, all queries processing on private data must occur in the USB key, which has limited memory and CPU capacity, and has unique I/O characteristics. At a high level, the paper addresses two questions: (1) what is the best physical design for the database on the USB key, given the characteristics of this device, and (2) how to take advantage as much as possible of the powerful but untrusted workstation in which the USB key is plugged.

The second paper *Data Warehouse Access using Multi-agent Systems* by Kolsi, Abdellatif and Ghedira describes an approach to dynamic distribution of data ware-house on multiple servers using a multi-agent system coupled by the scalable distrib-uted data structures. This approach ensures the scalability and exploits the storage and processing resources available in an organization. This paper proposes a novel research direction on the contribution of multi-agent system in managing data ware-house.

The third paper *Partitioning Methods for Multi-version XML Data Warehouses* by Rusu, Rahayu and Taniar explores partitioning methods for multi-version XML data warehouse. This paper proposes a nice state of art on different partitioning tech-niques proposed in traditional data warehouses and Semantic Web. Three main parti-tioning techniques were proposed: document partitioning, schema-based partitioning and cascaded (mixed) partitioning. In order to evaluate the effectiveness of each par-titioning method, a cost model calculating number of pages required for executing a query. This cost model is given for a partitioned and non partitioned multi-version XML data warehouse.

The fourth paper *Model and Procedure for Performance and Availability-wise Par-allel Warehouses* by Furtado presents a model and a procedure for chunk-wise on de-mand processing over parallel data warehouse in order to ensure load and availability balancing. The proposed work considers a relational data warehouse modeled using a star schema with a fact table and small dimension tables which are copied (dupli-cated) into all nodes. The roles of Partition-handling, load-balancing and availability-related flexibility are well described along this paper.

Finally, *Parallel OLAP Query Processing in Database Clusters with Data Replica-tion* by Lima, Furtado, Valduriez and Mattoso presents a distributed database design alternatives which combine physical fragmentation with a partial replication. A new load balancing strategy taking advantages of replicas to redistribute the load is pro-posed. The fact table is horizontally partitioned and allocated across various nodes using chained declustering, and the dimension tables are fully replicated. This en-ables the middleware layer to do load balancing among the replicas for improving query response time.

We hope readers will find the content of this special issue interesting and will inspire them to look further into the challenges that are still ahead before designing data warehouse applications. We would like to thank all the authors who submitted their papers to this special issue. In addition, I am grateful for the support of various reviews that ensured the high quality of this special issue. Last but not least, I would like to thank the two EICs of this journal: Professor Elmagarmid for accepting to run this special issue, and Professor Amit Sheth for overseeing the review process and his relevant advise for finalizing this issue. I also appreciate the assistance of Valerie Schofield from Springer.

References

1. Chaudhuri, S., Narasayya, V.R.: Self-tuning database systems: a decade of progress. In: Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB'07), September 2007, pp. 3–14