Guest Editorial

Susan Urban, Arizona State University, and
Elisa Bertino, Universitá di Genova, Italy

Modern organizations have strong requirements for tools that support coordinated access to data stored in distributed, heterogeneous, autonomous data repositories. There are several motivations for this need. The first is that organizations evolve over time. This evolution process influences the way data is managed within an organization. Indeed, the choice of a data base management system (DBMS) depends on the application requirements and on the available technology. As those evolve over time, an organization acquires several, possibly heterogeneous, data management systems. Second, many new applications are so complex that they cannot be effectively supported by a single data management system. Therefore, different data management systems need to inter-operate in order to completely support complex applications. Finally, not all data sources belong to the organizations using them. This is the case, for example, of data banks such as library catalogs, containing on-line information about all the books and journals available at the library, or stock market information. In this situation the organization does not have any control over how the data is organized and presented to the users of an application.

Unless adequate tools are provided, applications needing data from several, heterogeneous sources, have to bridge the gap among the various data management systems. This integration task may be rather difficult due to the differences in data models between different database systems, due to the lack of data models in some systems, or due to a lack of query languages and operational capabilities, such as access control, recovery, and concurrency control. Therefore, there is a great deal of research needed to develop effective implementation of integration architectures and tools.

Looking more closely at the problem of integrating heterogeneous database management systems (HDBMSs), we can see that heterogeneity can exist at several levels.

1. The operating system and hardware level: the systems to be integrated are located on different processors using different operating systems.
2. The data management system level: the systems managing the data have different data models and/or operational capabilities.
3. The semantic level: the various databases of the environment have been independently designed; that is, the same real-world entities may have been modelled in different ways in the various databases. Therefore, two databases may be logically heterogeneous even if the DBMSs managing them are completely homogeneous.

We can say that problems arising from level (1) of heterogeneity are solved, or will be solved soon, because of adoption of standards. There is still the need, however, of developing cooperation mechanisms, tailored to the needs of heterogeneous data management systems, from general communication mechanisms offered by operating systems and communication software [1,3]. Problems arising from level (2) are partially solved. Indeed, because of efforts like the SQL Access Group [2], applications will be provided with a uniform language for accessing and manipulating data stored by different DBMSs supporting the SQL language. However, the problem of integrating non-traditional data management systems, like CAD systems, or image management systems, has not been satisfactorily tackled. In addition several issues related to transaction management, access control, replicated data management, and performance are still open [1].

Finally, problems arising from semantic heterogeneity (level (3)) are still far from being solved [4]. We note that this level will be increasingly crucial for effective use of data. Indeed, it will be necessary to provide integrated descriptions of data, with associated filtering tools, so that the complexity of the underlying distributed heterogeneous databases is hidden.

This special issue on *Heterogeneous Database Systems* focuses on some of the above open research problems. In particular, this issue features papers from industry and from academia that demonstrate the practical application and analysis of current research results.

The first paper by **J. Chen, O. Bukhres,** and **A. Elmagarmid** discusses a cooperation mechanism implemented as part of the InterBase System, an HDBMS developed at Purdue University and tested for several years at Bell Northern Research, Inc. As such the paper presents interesting insights concerning the requirements that a computational and communication model must satisfy in order to support an HDBMS.

The second paper by **Y. Breitbart** and **A. Silberschatz** discusses concurrency control mechanisms; this is an important, and very difficult, issue in HDBMSs. Several algorithms have been proposed; however no studies on performance comparison have been reported. This paper presents two concurrency control mechanisms and then compares their performance.

The third paper by **H. Srinidhi** describes an approach that has been developed at Bell Communications Research for the management of redundant data in an HDBMS. The paper identifies different categories of redundant data and describes the problems associated with maintaining semantic integrity between multiple copies of data. Two different management schemes as well as an algorithm for managing redundant data are presented.

Finally, the fourth paper by **W.-K. Whang, S. Chakravarthy,** and **S. Navathe** addresses the issue of schema integration in an HDBMS to create a global view of several databases. In particular, the paper presents an algorithm that uses schema knowledge to derive relationships that exist between schemas. The paper also presents an extended query language for expressing queries over integrated schemas in an HDBMS.

In summary, the area of HDBMS is an area that demands feasible solutions to the complex problems associated with the integration of the heterogeneous information sources that exist in today's working environments. The papers in this special issue represent a step in the direction of addressing the practical concerns of HDBMS environments.

References

1. A. K. Elmagarmid (editor), *Database Transaction Models for Advanced Applications,* Morgan Kaufmann Publishers, San Mateo, CA, 1992.
2. ISO/IEC JTC 1 /SC21/WG 3. Information processing systems—open system interconnection (OSI)—remote database access (RDA), generic model. Technical report, ISO 9579, 1991.
3. T. Ozsu and P. Valduriez, *Principles of Distributed Database Design,* Prentice Hall, Englewood Cliffs, NJ, 1991.
4. A. Sheth and J. Larson, "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases," *ACM Computing Surveys,* vol. 22, 1990, pp. 183–236.