Research Issues in Federated Database Systems

Report of EFDBS '97 Workshop

S. Conrad[1], B. Eaglestone[2], W. Hasselbring[3], M. Roantree[4], F. Saltor[5], M. Schönhoff[6], M. Strässler[7], M. Vermeer[8]

[1] Technical Information Systems, University of Magdeburg, Germany.
[2] Department of Computing, University of Bradford, UK.
[3] Software Technology, University of Dortmund, Germany.
[4] School of Computer Applications, Dublin City University, Dublin, Ireland.
[5] Universitat Politecnica de Catalunya, Catalunya.
[6] Department of Computer Science, University of Zurich, Switzerland.
[7] ABB Corporate Research Ltd., Baden-Dättwil, Switzerland.
[8] Centre for Telematics and Information Technology, University of Twente, Netherlands.

Abstract *In June 1997, an international workshop on engineering of federated database systems has been held in Barcelona in conjunction with the 9th Conference on Advanced Information Systems Engineering (CAiSE'97). This paper reports on the results of this workshop and summarises the identified open issues for future research in this area.*

1 Introduction

A federated database system (FDBS) is an integration of autonomous heterogenous software systems (often database systems), where both global applications accessing multiple systems, and local applications are supported. The techniques for FDBS are employed in various application areas such as engineering design, healthcare, banking and so on.

The workshop proceedings are available online at [EFDB97] and in a published format [CHHS97]. The topics were:

- Software architectures for FDBS
- Design and implementation of FDBS
- Influence of the application domain on the design of FDBS
- Evaluation of FDBS approaches

The workshop reflected the high level of research activity in the area of FDBS. This activity is a response to real benefits delivered by FDBS technology countered by the open problems of reconciling conflicting requirements of integrated global access and local autonomy. Much of the work addressed consequential problems of semantic and implementation heterogeneities.

Work presented concerned:

- integrity constraint integration [CST97] and enforcement across replicated data [Has97];
- FDBS architecture extensions, to include management of schema evolution through ECA rules [GIMB97] and enhanced global and local metadata [RM97];
- algorithms for schema derivation [ROSC97] and mapping between local and global data models [HST97]
- security [ROSC97];
- data versioning [SSD97];
- semantic analysis of schemas in FDBS design [EM97];
- location-specific FDBS querying [LCC97];
- Case study FDBSs in the areas of healthcare [RM97] and engineering [SSD97] demonstrated strong influence of the application domain on both FDBS design and functionality.

2 Workshop Results

The following results are in the form of a number of observations on the emerging trends and consensus. These comments are not meant to be a definitive perspective on FDBSs but merely represent the opinions of those present at the workshop.

- The 5-level-schema-architecture of [SL90] is generally accepted as the basic structure of FDBS. However,
  - several extensions have been proposed [ROSC97, HST97, Has97, EM97];
  - component schemas usually represent only a subset of the conceptual schemas of component databases. As a result, export and component schemas are sometimes merged [SSD97].
- Object models, mainly ODMG [RM97, EM97, SSD97] and to a lesser extent EXPRESS [HST97], are generally used as canonical data models but integration problems such as those listed below are still a major priority.
- The integration of *legacy systems* often requires 'ad hoc' engineering to get access to their data stores. No general rules to get into legacy systems have been identified.
- Several prototype implementations exist.
These systems are usually 'hand made' with little or no CASE tool support. This situation is unlikely to change soon as there is little evidence of major focus on such tools.
- One of the *hot* topics of this workshop was integrity control [CST97] among component databases (data replication, referential integrity etc.). There exists some consensus on the use of active mechanisms for integrity control. To achieve integrity control, compromises (trade-offs) between integration and autonomy of component systems are necessary.
- More applications experience with implemented systems is needed.
- Integration of, and mapping between schemas resolves both semantic and implementation heterogeneities. In most of the work, the former was not identified as a problem. However, some authors [EM97] felt semantic analysis to be critical to successful FDBS design and identified a need for methods and tools.

3 Research Issues

The identified open issues for future research are:

- Many authors were attempting to use the ODMG Object Model as a canonical data model primarily because it represented a *standard* object model. However, some participants [Roa97] felt that there were major difficulties in using this model. The main reasons cited were its failure to provide a relational style view mechanism such as [KK95, Ber92, SS96] and problems with its query language (OQL). Other authors [EM97] expressed similar scepticism towards using another emerging *standard* SQL3 because of its ad hoc treatment of types and consequential lack of orthogonality and excessive complexity.
- It is unclear whether specific CASE tool support is required or whether it is possible to use existing tools for, e.g., schema integration. FDBS design methods, methodologies and tools have not been addressed by most authors (with the exception of [EM97]), and there is a clear need for investigation in this area.
- The semantics and implementation techniques for inter- and intra-model mappings need further investigation. In particular, adequate view mechanisms for the (object-oriented) canonical data models are required, as are tools and methods for identifying and resolving the heterogeneities that the mappings must traverse.
- The integration of semi-structured data that is stored in files. This is particularly needed if FDBS technology is to be relevant to WWW applications.
- The integration of behaviour in addition to data. For example, the definition of methods on the federated schema should involve the reuse of behaviour implemented in participating databases. Pre-existing method definitions at the component level can be seen as sub-routines of methods at the federated level. Issues include dealing with equivalences and discrepancies between methods defined in different participating databases [VA97].
- Support for versioning in FDBS. A global data model with support for version management is regarded as useful for global applications in certain FDBS environments such as engineering design [SSD97]. The federated system itself may also make use of a global versioning mechanism to support administrative tasks like schema evolution.
- Technological support for the integration of legacy systems.

References

[Ber92] E. Bertino. A View Mechanism for Object Oriented Databases. *Proc. Intl. Conference on Extending Database Technology*, Springer Verlag, 1992.

[CHHS97] S. Conrad, W. Hasselbring, A. Heuer and G. Saake (eds). Engineering Federated Database Systems (EFDBS97). *Proceedings of the International CAiSE97 Workshop*, Computer Science Preprint 6/1997, University of Magdeburg, 1997.

[CST97] S. Conrad, I. Schmitt and C. Türker. Dealing with Integrity Constraints During Schema Integration. *In [CHHS97]*.

[EFDB97] http://wwwiti.cs.uni-magdeburg.de/~conrad/EFDBS97/

[EM97] B. Eaglestone and N. Masood. Schema Interpretation: An Aid to the Schema Analysis in Federated Database Design. *In [CHHS97]*.

[GIMB97] A. Goni, A. Illarramendi, E. Mena and J. Blanco. Monitoring the Evolution of Databases in Federated Relational Database Systems. *In [CHHS97]*.

[Has97] W. Hasselbring. Extending the Schema Architecture of Federated Database Systems for Replicating Information in Hospitals. *In [CHHS97]*.

[HST97] T. Härder, G. Sauter and J. Thomas. Design and Architecture of the FDBS Prototype INFINITY. *In [CHHS97]*.

[KK95] W. Kim and W. Kelley. On View Support in Object-Oriented Database Systems. *In Modern Database Systems: The Object Model, Interoperability and Beyond*, Won Kim (ed), Addison-Wesley, 1995.

[LCC97] E. Lim, Y. Cao and R. Chiang. Source-aware Multidatabase Query Processing. *In [CHHS97]*.

[RM97] M. Roantree and J. Murphy. An Architecture for Federated Database Metadata. *In [CHHS97]*.

[Roa97] M. Roantree. Evaluating the ODMG Object Model for Usage in a Multidatabase Environment. *Technical Report No. CA-2597*, Dublin City University, 1997.

[ROSC97] E. Rodriguez, M. Oliva, F. Saltor and B. Campderrich. On Schema and Functional Architectures for Multilevel Secure and Multiuser Model Federated DB Systems. *In [CHHS97]*.

[SL90] A. Sheth and J. Larson. Federated Database Systems for Managing Distributed, Heterogenous, and Autonomous Databases. *ACM Computing Surveys*, vol. 22, no. 3, September 1990.

[SS96] J. Samos and F. Saltor. External Schema Generation Algorithms for Object Oriented Databases. *Proc. Intl. Conference on Object Oriented Information Systems*, Springer, 1996.

[SSD97] M. Schönhoff, M. Strässler, K.R. Dittrich. Data Integration in Engineering Environments. *In [CHHS97]*.

[VA97] M. Vermeer and M. Apers. Behaviour Specification in Database Operation. *Proceedings of CAiSE 97*, LNCS no. 1250, 1997.