Update Support in Component-Based Database Systems

1 Introduction to the Problem

Modern database schemata are often large and complex, and a single user typically is afforded only limited access through a window, known as a view. The problem of supporting updates systematically through such views has long been known to be a difficult problem. Traditionally, the approach has been to limit the allowed updates to those which avoid or substantially limit so-called *update anomalies*; that is, changes which are not visible within the view itself. Unfortunately, the collection of such updates, which are defined by the so-called *constant-complement strategy* [BS81], [Heg04], is severely limited, and traditional extensions afford only a moderate increase in flexibility.

In the research proposed here, a different approach to the view-update problem is taken. If a user of a given view cannot perform a given update without side effects, then that user must enlist the cooperation of other users, having update rights on other views, in such a way that all changes to the global database are visible to, and authorized by, at least one of the cooperating users.

2 Completed Work

Update by cooperation requires a way of modelling database schemata in which the users of distinct views can cooperate in a meaningful way. The key idea is to represent the main schema as the interconnection of of *database components*. The utility of such components was first forwarded by Thalheim and his coworkers [ST04] [Tha05] [ST06]. The completed work of this research program builds upon these ideas as follows.

A view-based model of database components: The components of Thalheim are based upon a *co-design* approach [Tha03] in which the database applications form an integral part of the component. While such components are useful in a process of software development for database applications, for the research program described here, a model of components which recaptures traditional database schemata, and in particular the decoupling of the schema from its applications, is essential. The foundations for such a component-based model have been developed by the author [Heg08b]. Each component corresponds roughly to a user view of the large schema which is defined by their interconnection. If an update to a given component cannot be realized while holding all other components constant (*i.e.*, via a constant-complement strategy), then that update must be realized via updates to additional components, through the cooperation of users authorized to update those components.

Models of component update via cooperation: In [HS07], a first proof-of-concept model of how such updates are managed was forwarded, consisting of a completely egalitarian model of cooperation in which all users have absolute rights over their own components, together with a global model of inter-component communication. A more sophisticated model has just been developed [HS08] which includes a hierarchy of authority (so that one component can *coerce* another to support an update), a corresponding notion of access rights for database components which is an extension of the rôle-based model of access control [Bal90] [FKC07], and a model of localized access control which resembles those used in distributed database systems [LBK02, Ch. 26].

The theoretical foundations for enforcing minimality on cooperative updates: Despite the cooperative nature of this model of update, it is important that only changes which support the initial request, without "pork-barrel" additions, be allowed; thus an appropriate notion of minimality of an extension of an update from one component to others is necessary. It is often the case that using simple measures based upon the tuples which are altered is inadequate; rather, a notion which also looks a changes within the tuples is also necessary. To this end, as part of this overall research program, a theory of *information content* and its minimization in the update process has been developed by the author during the past year [Heg08a] [Heg08c].

3 The Proposed Research

The proposed research is a continuation of a collaboration with Peggy Schmidt and Bernhard Thalheim of the Information Systems Engineering Group at Christian-Albrechts-Universität zu Kiel, Germany. It will build upon the foundational results reported in [HS07], [Heg08b], [HS08], [Heg08a], and [Heg08c] and described above, in the following ways.

Support for non-monotonic negotiation: In both [HS07] and [HS08], the negotiation process is monotonic, in the sense that an update request, as it is propagated through the components, can only be refined. There is no facility which permits an actor to present a true counterproposal which conflicts with aspects of the existing proposal, and for the initial proposer then to agree to a modification. Of course, such properties are essential for models of real-world negotiation. While the representation of such properties is simple in principle, identifying useful conditions under which the negotiation process can be proven to terminate has been difficult. Work will continue on this central topic.

The connection between update via cooperation and workflow: The classical topic of *workflow* involves the systematic modelling of processes which require the coordinated interaction of several actors [AH02]. It has long been known that such models play an important rôle in the context of information systems [FGHW88] [AAA*96] [RS95]. The model of cooperative update clearly involves a workflow, as the negotiation process proceeds and requests are passed from actor to actor. In general, a cooperative update request will not define a fixed workflow, but rather will impose constraints on the family of workflows which will lead to its realization. A goal of this part of the proposed research is to make this association explicit, as a theory of *query-based workflow construction*. In other words, an active query (*i.e.*, an update) to a single component defines a family of workflows which may be used in its realization as a cooperative update of the interconnected components.

The modelling of workflows defined by update via cooperation using BPMN: In the practical modelling of applications, BPMN (Business Process Modeling Notation) [Gro06] has seen increasingly widespread use, particularly as a standard method of representing the associated workflow [Whi04]. Since the central examples of component-based update are interactive business models, it is important to understand how BPMN relates to the workflow defined by database components, and so this topic will also be investigated as part of this phase of the project.

The application of minimality defined via information content to update by cooperation: The ideas of information content which were developed in [Heg08a] and [Heg08c] and identified in Section 2 are highly theoretical in nature and apply to general update processes. In this phase of the proposed research, these ideas will be adapted specifically to the problem of ensuring that interactive, negotiated updates do not contain any superfluous “pork-barrel” additions which were added by opportunistic users.

An executable model of database components and update by cooperation: Purely theoretical models often contain pitfalls which are difficult to detect. Therefore, to validate the theoretical underpinnings, an executable model of interconnected components is being developed using the CoreASM extensible ASM execution engine [FGG07], which is in turn based upon *Abstract State Machines* [BS03]. Work will continue on this simulation model during the course of the proposed research.

References

[AH02] Wil van der Aalst and Kees van Hee, *Workflow Management: Models, Methods, and Systems*, MIT Press, 2002.

[AAA*96] Gustavo Alonso, Divyakant Agrawal, Amr El Abbadi, Mohan Kamath, Roger Günthör, and C. Mohan, "Advanced transaction models in workflow contexts," in: Stanley Y. W. Su, ed., *Proceedings of the Twelfth International Conference on Data Engineering, February 26 - March 1, 1996, New Orleans, Louisiana*, pp. 574–581, IEEE Computer Society, 1996.

[Bal90] Robert W. Baldwin, "Naming and grouping privileges to simplify security management in large databases," in: *Proc. 1990 IEEE Symposium on Research in Security and Privacy*, pp. 116–132, 1990.

[BS81] François Bancilhon and Nicolas Spyratos, "Update semantics of relational views," *ACM Trans. Database Systems*, **6**(1981), pp. 557–575.

[BS03] Egon Börger and Robert Stärk, *Abstract State Machines*, Springer-Verlag, 2003.

[FGG07] Roozbeh Farahbod, Vincenzo Gervasi, and Uwe Glässer, "CoreASM: An extensible ASM execution engine," *Fundamenta Informaticae*, **77**(2007), pp. 71–103.

[FKC07] David F. Ferraiolo, D. Richard Kuhn, and Ramaswamy Chandramouli, *Role-Based Access Control*, Artech House, second edition, 2007.

[FGHW88] Fernando Flores, Michael Graves, Brad Hartfield, and Terry Winograd, "Computer systems and the design of organizational interaction.," *ACM Trans. Inf. Syst.*, **6**(1988), pp. 153–172.

[Gro06] Object Management Group, "Business Process Modeling Notation specification," Technical Report dtc/06-02-01, OMG, 2006.

[Heg04] Stephen J. Hegner, "An order-based theory of updates for database views," *Ann. Math. Art. Intell.*, **40**(2004), pp. 63–125.

[Heg08a] Stephen J. Hegner, "Information-optimal reflections of view updates on relational database schemata," in: Sven Hartmann and Gabriele Kern-Isberner, eds., *Foundations of Information and Knowledge Systems: Fifth International Symposium, FoIKS 2008, Pisa, Italy, February 11-15, 2008, Proceedings*, pp. 112–131, Volume 4932 of Lecture Notes in Computer Science, Springer-Verlag, 2008.

[Heg08b] Stephen J. Hegner, "A model of database components and their interconnection based upon communicating views," in: Hannu Jakkola, Yashui Kiyoki, and Takehiro Tokuda, eds., *Information Modelling and Knowledge Systems XIX*, pp. 79–100, Frontiers in Artificial Intelligence and Applications, IOS Press, 2008.

[Heg08c] Stephen J. Hegner, "Semantic bijectivity and the uniqueness of constant-complement updates in the relatiional context," in: Klaus-Dieter Schewe and Bernhard Thalheim, eds., *International Workshop on Semantics in Data and Knowledge Bases, SDKB 2008, Nantes, France, March 29, 2008, Proceedings*, Volume 5115 of Lecture Notes in Computer Science, Springer-Verlag, 2008, in press.

[HS07] Stephen J. Hegner and Peggy Schmidt, "Update support for database views via cooperation," in: Yannis Ioannis, Boris Novikov, and Boris Rachev, eds., *Advances in Databases and Information Systems, 11th East European Conference, ADBIS 2007, Varna, Bulgaria, September 29 - October 3, 2007, Proceedings*, pp. 98–113, Volume 4690 of Lecture Notes in Computer Science, Springer-Verlag, 2007.

[HS08] Stephen J. Hegner and Peggy Schmidt, "Updates to database schemata defined by interconnected components," Technical Report 0805, Institut für Informatik, Christian-Albrechts-Universität zu Kiel, 2008, to be submitted for journal publication during June 2008.

[LBK02] Philip M. Lewis, Arthur Bernstein, and Michael Kifer, *Databases and Transaction Processing: An Application-Oriented Approach*, Addison-Wesley, 2002.

[RS95] Marek Rusinkiewicz and Amit P. Sheth, "Specification and execution of transactional workflows," in: Won Kim, ed., *Modern Database Systems*, pp. 592–620, ACM Press and Addison-Wesley, 1995.

[ST06] Klaus-Dieter Schewe and Bernhard Thalheim, "Component-driven engineering of database applications," in: *APCCM '06: Proceedings of the 3rd Asia-Pacific conference on Conceptual modelling*, pp. 105–114, Australian Computer Society, Inc., Darlinghurst, Australia, 2006.

[ST04] Peggy Schmidt and Bernhard Thalheim, "Component-based modeling of huge databases," in: András Benczúr, János Demetrovics, and Georg Gottlob, eds., *Advances in Databases and Information Systems: 8th East European Conference, ADBIS 2004, Budapest, Hungary, September 22-25, 2004, Proceedings*, pp. 113–128, Volume of Lecture Notes in Computer Science, Springer-Verlag, 2004.

[Tha03] Bernhard Thalheim, "Informationssystem-Entwicklung: die integrierte Entwicklung der Strukturering, Funktionalität, Verteilung und Interaktivität von großen Informationssystemen," Technical Report 15-2003, Brandenburgische Technische Universität Cottbus, 2003.

[Tha05] Bernhard Thalheim, "Component development and construction for database design," *Data Knowl. Eng.,* **54**(2005), pp. 77–95.

[Whi04] Stephen A. White, "Process modeling notation and workflow patterns," in: Layna Fischer, ed., *The Workflow Handbook 2004*, Future Strategies, Inc., 2004.