A Negotiation Process Approach for Building Federated Databases

Marta Oliva
Dept. d'Informàtica i Enginyeria Industrial
Universitat de Lleida
E-25003 Lleida
e-mail: oliva@eup.udl.es

Fèlix Saltor
Dept. de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya
E-08071 Barcelona
e-mail: saltor@lsi.upc.es

Abstract

The negotiation process is often referred to in the literature on federated databases, but is seldom covered in depth. This process is essential to determine data of the component schema to be integrated for building a federated schema and the access permissions to be granted. This paper presents our negotiation process approach which is incorporated in the integration schemas mechanism, so we can lie in getting the benefit of work realised in analysis and resolution phases of integration schemas needed for negotiation process.

1 Introduction

Federated databases (FDB) have arisen because of the increasing need for combining information stored in pre-existing databases. If a local database system is to join a federation, it does not want to give up its local autonomy, and integrity as well as confidentiality of data stored in the local database have to be preserved; moreover the federation also has to provide for strong security mechanisms ([JD93]).

The fact of building a federated database implies the integration of the schemas of the component databases. By integration we mean the set of necessary procedures in order to access the information stored by the federation as if it was a centralised database system. Each time that a component database wants to join a federation, it should show the information that it will be ready to share, by *export schemas* ([SL90]), with their respective access permissions.

Two kinds of federations can be distinguished [JD93], namely loosely coupled systems and tightly coupled systems. In the former case, every user has to carry out his/her own integration. On the other hand, tightly coupled systems are characterised by a global integration layer that is managed by a global or federated administrator. He/She is responsible for integrating component database systems (CDBSs) which are to join this federation. The global or federated schemas resulting from the integration process is a subject of negotiation between the federated and the local administrators. They contain the data description that could be shared by federation users, and they also contain the access authorisations to different users or groups of users.

In this paper we focus on negotiating data schemas for tightly coupled federation systems. It is organised as follows: in Section 2 we review related work about negotiation processes and schema integration; in Section 3 we present our negotiation process approach which is incorporated in the schema integration mechanism.

2 Related work

2.1 Negotiation process

Only few authors have participated in the definition and the description of negotiation process that is essential to federated database systems. Each of them has treated the negotiation process depending on the two kinds of federated architectures.

In Sheth and Larson's paper [SL90], where they present a five-level schema reference architecture for federated databases, they define the negotiation process as the dialogue between the administrator of the federation and each administrator of the component databases. They must reach an agreement about the contents of the export schemas and operations allowed on the export schemas such that the federated schemas can be defined over them to support federation users. This definition assumes the existence of a federation administrator that only appears in tightly coupled federated systems.

In loosely coupled federations, [AB89] defines a negotiation protocol where directly participate the component of the federation (importer) that wants to import data and the component of the federation (owner) that can export data. [JW89], [AJ90] describe negotiation by the establishment of bilateral agreements between nodes represented by contracts. On the same idea, Andersson et al. ([ADS$^+$93]) agree that each administrator of the local database has the responsibility to decide and define which parts of their local data will be available to external users. So each local administrator should be negotiating with the rest of local administrators. Negotiation also settles an agreement on access restrictions and the desired level of consistency between imported data (if it is the case), and the original version in the exporting system. Negotiation needs an understanding of the semantics of data.

2.2 Schema Integration

The schema integration process in federated database systems refers to the integration of existing schemas (of component databases) into a single schema called federated schema ([SL90]). According to Batini et al. ([BLN86]), the schema integration process is divided into five steps: *preintegration, comparison, conformation, merging, and restructuring.* Other integration processes have been described in the literature: [WCN92] , [SPD92], ([FN93]), [GSSC95b].

Because of the heterogeneity of data models that we can find in different component databases, it is indispensable to select a data model in order to represent the information referring to component schemas in an uniform way. This model is called *canonical data model* and it must be as semantically rich as possible. Thus, most existing approaches refer to an object-oriented data model, since it can express more information about the universe of discourse than any traditional model ([BNPS89], [SCGS91], [SPSY93], [CSGS94]).

Preintegration is the first phase of the schema integration process and it involves *semantic enrichment* of the local schema and *translation* of enriched local schemas into component schemas expressed in the canonical data model of the federation. This is a very difficult task, so it has been studied independently from schema integration ([SFL92], ([ADS+93], [SCGS93]).

The second phase of the schema integration process involves, according to Batini et al. ([BLN86]), two activities:

- analysing and comparing the objects of the schemas to be integrated, including identification of naming conflicts, domain conflicts, structural differences, constraint differences, and missing data
- specifying the interrelationships among the schema objects

[WCN92], [GSCS93], [GSSC95a], [SP91], [ADS+93] are related works on analysis process.

Both phases, preintegration and analysis, are very interesting for our negotiation process approach and we focus on the analysis phase.

3 Our Approach

Previous related work on federated databases points out that negotiation is essential to determine which data will be shared and which operations can be performed on them.

This negotiation process approach is thought for tightly coupled federated databases, where there is a federated schema (or several federated schemas) expressed in the federated canonical data model, and there is also an administrator of the federation. Along the exposition we will use Sheth and Larson's architecture as a reference ([SL90]).

When an autonomous database is to be added to a federation, it should be negotiated with the administrator of the federation. In order to perform this negotiation it will be essential for the administrator of the federation to know the semantics of data of the autonomous database. So an initial export schema should be put at the federation administrator's disposal. This initial export schema can contain a component schema subset or it can coincide with the component schema (see below), in any case after the preintegration phase (translation and enrichment) outlined before.

To carry out the negotiation it is necessary to determine the relationships between the components of the federated schema and the export schema, so they should be analysed to find out which objects have a negotiation possibility, and different possibilities for their resolution ([GSSC95b]). So far the process coincides with the analysis phase of the schema integration mechanism, according to Batini ([BLN86]) related in the previous section, and resolution subphase of the integration process.

Although most of the analysis methodologies are based on comparisons between each object/class of one component database schema (CDB schema) with all objects/classes of others CDB schemas, [GSSC95a] describes a method to detect class similarities by following a strategy and applying comparison criteria. The strategy exploits the semantically rich structures of

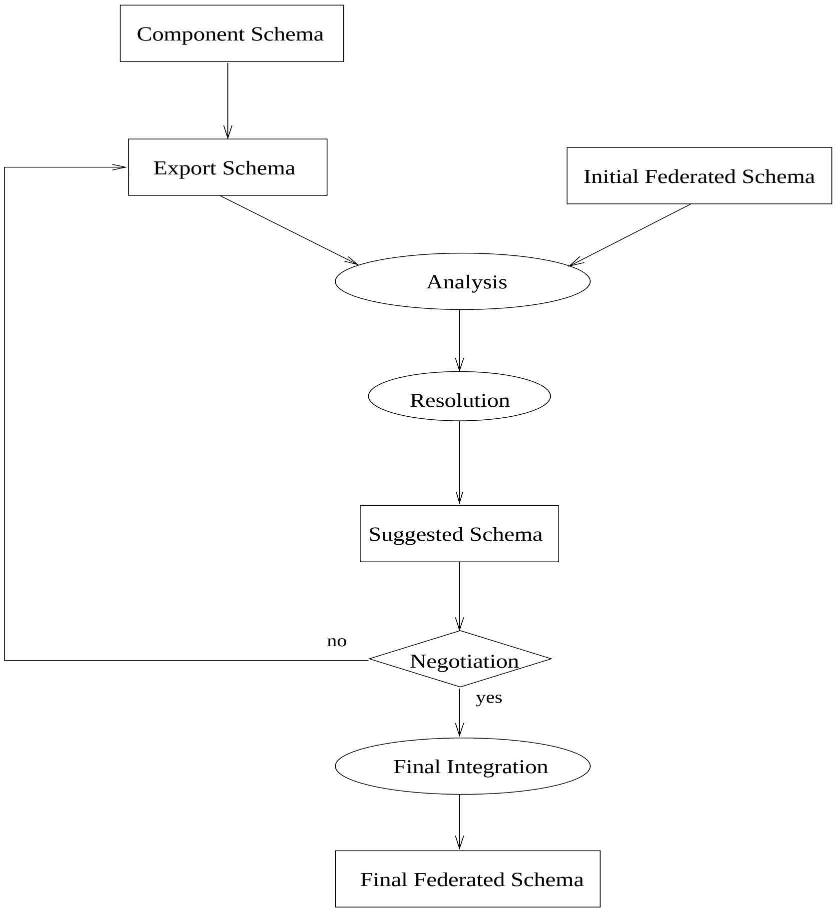

Figura 1: Negotiation process schema

the schemas (previously enriched), along both the generalization/specialization and the aggregation dimensions. Also, relaxations may be applied to conform a pair of classes, resulting in penalizations in the computation of the degree of similarity. Finally, after the comparisons the resemblances are presented to the human integrator for confirmation or refusal.

So our approach lies in getting the benefit of work realised in integration, building in the negotiation phase between the analysis and integration phases just when the human integrator must act.

With the help of the *semi-automatic analysis process* described above, where the local and federated administrators must take part as a human integrators, a proposal of federated schema is obtained which reflects the inclusion of the new database. The content of the initial export schema will depend on the initial proposal. If the federated administrator suggests the initial proposal then the initial export schema coincides with the component schema. Else, if the suggestion would be put forward by the local administrator then the initial export schema contains a component schema subset. The *semi-automatic resolution process* is needed to obtain a suggested schema. Over this suggestion the two administrators must negotiate. In case of disagreement, they must negotiate the changes that should be reflected in the export schema in order to reach an agreement. To take into account these changes it is necessary to redefine the export schema and to begin the negotiation process again (see figure 1). The final integration phase can start once both administrators reach an agreement.

To decrease the cost that each new beginning of process implies, it is necessary to implement the analysis and resolution with *differential procedures*, so that only the part or parts of export schema affected by change are taken into account. Moreover, taking into account the complexity that the negotiation of the total suggested federated schema implies, we will reduce it to the negotiation of each of its classes and attributes. So export schema changes can consist in adding or removing some classes or a group of them, or adding or removing some attributes.

Even though the description of the negotiation process has been made assuming the previous existence of the federation and the inclusion of only a new database, the same approach, with relevant modifications, can be applied to any other situation.

Next, we present the process adapted to different situations:

1. Building a federation with two autonomous databases

   The initial federated schema does not exist, so one of the two export schemas must take its place. Changes can affect both export schemas. Before the process starts it is essential to name an administrator for the new federation and define the canonical data model that the federation will use.

2. Inclusion of an autonomous database into a federation with several federated schemas

   The inclusion must make into all federated schemas following the process in figure 1 for each federated schema.

3. Inclusion of several databases

   We can make a sequential inclusion or a concurrent inclusion. The sequential inclusion consists of making all database inclusion following the process in figure 1. The concurrent inclusion consists of making all database inclusion at the same time. Changes can affect all export schemas.

The most general case shows the possibility that the result of a negotiation process involves changes in any export schema (new or old) for federation interests.

4 Conclusion

Our negotiation process approach determines the location -in a federated database building process- where the negotiation should be placed, and it identifies two very important phases, analysis and resolution, that also appear in the schema integration process. So with the inclusion of the negotiation process between the analysis and final integration phases of the schema integration mechanism, we can remove the duplicity of the analysis and resolution processes needed both for the negotiation process and for the integration mechanism.

Acknowledgements

This work has been partially supported by the Spanish Research Program PRONTIC under project TIC96-0903.

References

[AB89] R. Alonso and D. Barbara. Negotiating data access in federated database systems. In *Proceedings of the 5th International Conference on Data Engineering*, pages 56–65. IEEE-CS Press, Feb 1989.

[ADS+93] M. Andersson, Y. Dupont, S. Spaccapietra, K. Yetongnon, M. Tresch, and H. Ye. FEMUS: a Federated Multilingual Database Systems. In N.R. Adam and B.K. Bhargava, editors, *Advanced Database Systems*. Springer-Verlag, New York, 1993.

[AJ90] Matts Ahlsén and Paul Johannesson. Contracts in Database Confederations. Technical report, The Swedish Institute for Systems Development, Kista, Sweden, 1990.

[BLN86] C. Batini, M. Lenzerini, and S. Navathe. A comparative analysis of methodologies for database schema integration. *ACM Computing Surveys*, 18(4):323–364, 1986.

[BNPS89] E. Bertino, M. Negri, G. Pelagatti, and L. Sbattella. An Object-Oriented Approach to the Interconnection of Heterogeneous Databases. In *Workshop on Heterogeneous Database Systems, Chicago*, pages 81–87, Dec 1989.

[CSGS94] M. Castellanos, F. Saltor, and M. García-Solaco. A Canonical Model for the Interoperability among Object-Oriented and Relational Databases. In Ozsu, Dayal, and Valduriez, editors, *Distributed Object Management*, pages 309–314. Morgan Kaufmann, 1994.

[FN93] Peter Fankhauser and Erich J. Neuhold. Knowledge Based Integration of Heterogeneous Databases. In D.K. Hsiao, E.J. Neuhold, and R. Sacks-Davis, editors, *Interoperable Database Systems (DS-5)*, pages 155–175. North-Holland, 1993.

[GSCS93] M. García-Solaco, M. Castellanos, and F. Saltor. Discovering Interdatabase Resemblance of Classes for Interoperable Databases. In Schek, Sheth, and Czejdo, editors, *Proceedings of 3rd Int. Workshop on Research Issues in Data Engineering: Interoperability in Multidatabase Systems, Vienna*, pages 26–33. IEEE-CS Press, 1993.

[GSSC95a] M. García-Solaco, F. Saltor, and M. Castellanos. A Structure Based Schema Integration Methodology. In *Proc. 11th Int. Conference on Data Engineering, Taipei.* IEEE-CS Press, 1995.

[GSSC95b] M. García-Solaco, F. Saltor, and M. Castellanos. Extensional Issues in Schema Integration. In T. Cheung, editor, *Proc. 6th Int. HK DB Workshop: Database Reengineering and Interoperability*, pages 243–254, Hong Kong, 1995. City University of Hong Kong.

[JD93] D. Jonscher and K.R. Dittrich. Access Control for Database Federations: a discussion of the state-of-the-art. In *Proc. DBTA Workshop on Interoperability of DBSs and DB Applications*, pages 156–178, Oct 1993.

[JW89] Paul Johannesson and Benkt Wangler. The negotiation mechanism in a decentralized autonomous cooperating information systems architecture. In G.X. Ritter, editor, *Information processing 89*, pages 163–168. Elsevier Science B. V. (North-Holland) IFIP, 1989.

[SCGS91] F. Saltor, M. Castellanos, and M. García-Solaco. Suitability of Data Models as Canonical Models for Federated DBs. *ACM SIGMOD Record*, 20(4):44–48, 1991.

[SCGS93] F. Saltor, M. Castellanos, and M. García-Solaco. Overcoming Schematic Discrepancies in Interoperable Databases. In D.K. Hsiao, E.J. Neuhold, and R. Sacks-Davis, editors, *Interoperable Database Systems (DS-5)*, pages 191–205. North-Holland, 1993.

[SFL92] S.Y.W. Su, S.C. Fang, and H. Lam. An object-oriented rule-based approach to data model and schema translation. Technical Report TR-92-015, Departament of Computer and Information Sciences, University of Florida, USA, 1992.

[SL90] A.P. Sheth and J. Larson. Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. *ACM Computing Surveys*, 22(3):183–236, Sep 1990.

[SP91] S. Spaccapietra and C. Parent. Conflicts and Correspondence Assertions in Interoperable Databases. *ACM SIGMOD RECORD*, 20(4):49–54, Dec 1991.

[SPD92] S. Spaccapietra, C. Parent, and Y. Dupont. Model Independent Assertions for Integration of Heterogeneous Schemas. *Very Large Databases*, 1(1), Jul 1992.

[SPSY93] Stefano Spaccapietra, Christine Parent, Marcos Sunye, and Kokou Yetongnon. Object Orientations and Conceptual Modelling. In *GI-Emisa Fachtagung Object-oriented Methods for Information Systems, Klagenfurt*, Jun 7-9 1993.

[WCN92] W.K. Whang, S. Chakravarthy, and S.B. Navathe. Heterogeneous Databases: Inferring Relationship for Merging Component Schemas, and a Query Language. Technical Report UF-CIS-TR-92-048, Computer and Information Sciences, University of Florida, USA, Dec 1992. Submitted for publication.