17th International CODATA Conference-
"Data and Information for the Coming Knowledge Millennium"
Symposium I: Data, Information and Knowledge
Theme 1-3: Integration of Heterogeneous Databases and Data Warehousing

BUILDING SECURE DATA WAREHOUSE SCHEMAS FROM FEDERATED INFORMATION SYSTEMS

Fèlix Saltor [(a)] Marta Oliva [(b)] Alberto Abelló [(a)] José Samos [(c)]
[(a)] U. Politècnica de Catalunya [(b)] U. de Lleida [(c)]U. de Granada
+3493 401 7023, +34973 703227, +3493 401 7332, +34958 243176
{saltor,aabello}@lsi.upc.es, oliva@eup.udl.es, jsamos@ugr.es

Introduction

There are many *heterogeneities* among preexisting informations sources, such as DataBases (DBs), particularly *semantic heterogeneities* (see our chapter "Semantic Heterogeneity" in [1]). Any secure *Federated Information Systems* (FIS) must solve them, and deal with a): how schema levels are related to security, and b): operations on schemas to convert from level to level. We present the approach of our integrated schema architecture for both FIS and secure *Data Warehousing*,

Terms of reference

Security model We will be assuming Mandatory Access Control, specifically Multi Level Security (MLS)

Federated Information Systems (FIS) For concepts see for example [1]. We will be using a 7-level schema architecture [presented at EFDBS'97], as an extension of the 5-level schema reference architecture in [3], One of the two additional levels contains *Authorization Schemas*, representing subsets of the Federated Schema, for a class of federated users with a certain Clearance Level.

Data Warehousing and Data Marts We will be using Inmon's definitions [2]: *Data Warehouse* (DW) is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management's decision-making process. *Data Mart* (DM) is a subset of a DW that has been customized to fit the needs of a department or subject area.

Our previous work This work (supported by CICYT as TIC99-1078-C02) is based upon earlier results of the BLOOM research group, including (see page *http://www-lsi.upc.es/bloom/* for details):

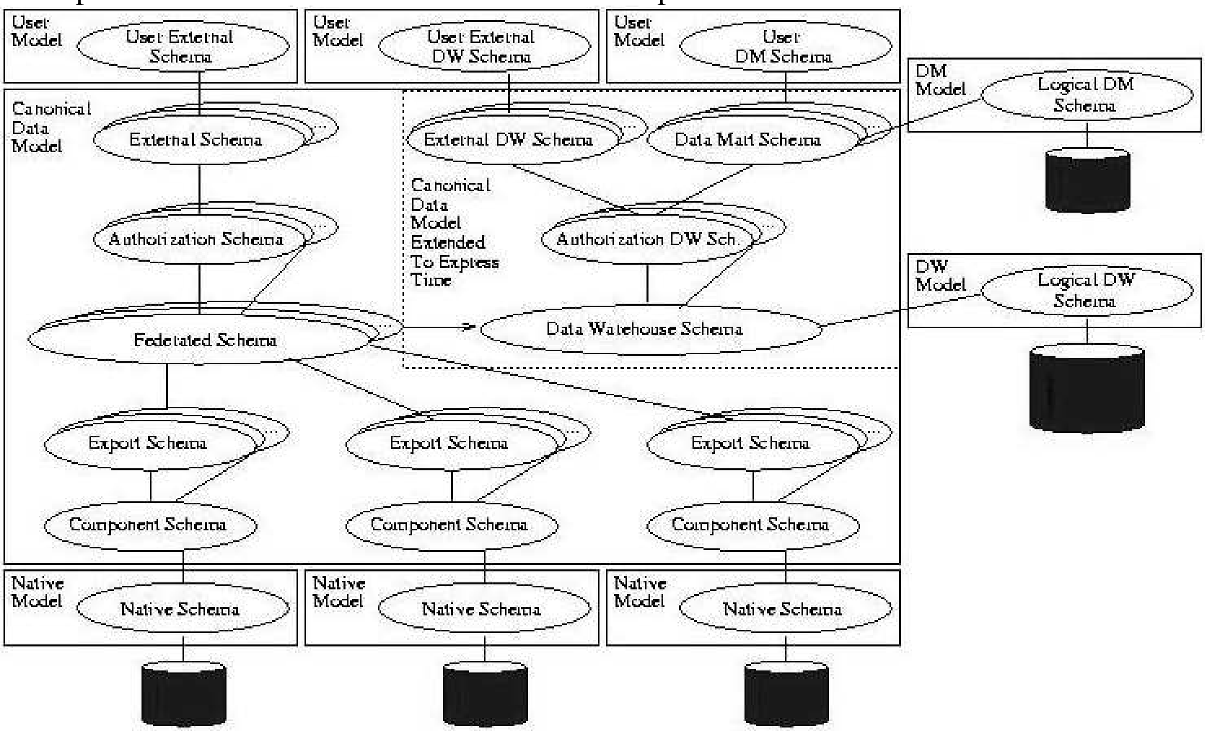
Development of the BLOOM data model as Canonical Data Model (CDM).
Classification, detection and resolution of semantic heterogeneities.
Development of a 7-level schema architecture for FIS.
Development of an integrated architecture for both FIS and DW [paper at EFIS'00].
Integration of different security models [paper at DBSec'00].
Adequacy for DW of O-O data models [paper at ECOOP'00 Symp. Objects & Databases].

Results

Secure DW schemas

Authorization Schemas help federated databases in MLS data protection, because each Authorization Schema defines the subset of information that a class of users/applications can access. In our integrated architecture, Authorization DW Schemas help Data Warehousing in the same way.

Another characteristic of Authorization Schemas is that they are also used to assist information inference control. Some abstractions of BLOOM are so rich semantically that the description of some characteristics of a class could point out other information.



In the figure we can see the result of integrating the 7-level architecture for federated databases, and the schema levels for Secure Data Warehousing.

Operations on Schemas We need different kinds of operations in the CDM (BLOOM):

1) *conforming* operations to transform the (export) schema of one DB to a form more suitable for integration into a Federated schema.

2) *generalization* of classes from different DBs to a superclass in a Federated schema. Discriminated generalization is preferred [our paper at ICDE'95].

3) *object identification* function (oid) to assert when an object O1 in one DB represents the same real world object as an object O2 in another DB.

4) *collapse* two objects into one using a particular oid [our paper at CoopIS'95].

5) *dealing with value discrepancies*, from preserving all values to keeping just one, including all intermediate aggregating options [our paper at KRDB'97].

6) *protecting security* by hiding relationships between BLOOM abstractions that could reveal confidential information.

7) *incorporating time* into BLOOM schemas.

8) *transforming* the structures of the DW data model to the Data Mart model.

Conclusions

We have shown how our 7-level integrated architecture supports both FIS users, who access the component databases in real time, and DW/DM users, who access data stored in the DW or the DMs. Moreover, Multi Level Security is assured in both cases by an additional schema level and by adequate operations.

References

[1] Bukhres & Elmagarmid (ed.): *Object-Oriented Multidatabase Systems.* Prentice-Hall 1996

[2] W. H. Inmon: *Buiding the Data Warehouse.* John Wiley & Sons, 1996.

[3] Sheth & Larson: Federated database systems for managing distributed, heterogeneous and autonomous databases. *ACM Computing Surveys, 22*(3):183-236 (Sept 1990).