Entity Relationship Stored Procedure

CĂLIN-ADRIAN COMES	NICOLAE GHIŞOIU	VASILE PAUL BREŞFELEAN	IOAN RUS
Petru Maior University	Babeş-Bolyai University	Babeş-Bolyai University	Petru Maior University
Nicolae Iorga 1, 540088	Mihail Kogălniceanu 1, 400084	Mihail Kogălniceanu 1, 400084	Nicolae Iorga 1, 540088
Târgu-Mureş, ROMÂNIA	Cluj-Napoca, ROMÂNIA	Cluj-Napoca, ROMÂNIA	Târgu-Mureş, ROMÂNIA

Abstract: ER-SP(Entity Relationship - Stored Procedures) model is proposed like an extension to ER (Entity Relationship) model for conceptual, syntactic and semantic modeling of RDBMS in idea to create a link between external to physical models. ER-SPL (Entity Relationship - Stored Procedures Language) is proposed to be independent by RDBMS team players in idea to support their own SQL dialects. The model proposes a new entity type called Stored Procedures type and the language behind him in idea to be platform independent in relation with SQL dialects and their Procedural Languages. The concept of stored procedures type is abstract, helps in syntactic and semantic modeling, and is required in physical implementation. A stored procedures entity type along with his language captures the syntactic and semantics of an RDBMS schema in his dynamical evolution.

Key–Words: Entity Relationship(ER), Entity, Attribute, Relationship, Relational Database Management System(RDBMS), Data Definition Language (DDL), Data Manipulation Language (DML), Data Control Language(DCL)

1 Introduction

Peter Pin-Shan Chen with ER in [6] describe a model that serves as the foundation of many systems analysis and design methodologies, repository systems, and CASE tools from commercial software vendors to free source world.

Any important article from journals and books from respectable publishing houses where the **modeling** word is a key word the ER Model or ER Model flavors are just in place. During the time because there is no standard for ER Model there are a lot confusions regarding the definition of the entity. Bernhard Thalheim in [20] cite more than twelve different approach for defining the entity notion with the following remark ”The confusion is almost since most of the database and software engineering books do not define at all the notion of entity ”.

OCL was formally developed as a business language with roots in Syntropy, second generation language object-oriented analysis and software design method developed at Object Designers Limited in the UK during the early 1990s, used to describe UML models.

In this paper we try to present the Entity Relationship-Stored Procedure Language in Object Constraint Language style, used to describe the Entity Relationship-Stored Procedure.

2 State of the art of ER Metamodels

ER Model has been redefined because of his popularity with ER Metamodels like EER (Extended Entity Relationship)[10] , CSL (Conceptual Schema Language) [5], HERM (Higher-Order Entity Relationship Model) [20], REMORA [17], and Barker [1].

2.1 Extended Entity Relationship (EER) Metamodel

David W. Embley and Tok Wang Ling in [10] extend the ER Model with EER Meta-model with the following approaches: capturing the real world semantic, transform the EER Metamodel in to a normalized EER Metamodel and generate normalized relations.

Entity-Relationship (ER) models and extended Entity-Relationship (EER) models has limitations and problems:

- Entity-Relationship (ER) models and extended Entity-Relationship (EER) models require designers to distinguish between attributes and entities. This can cause downstream redesign when attributes and entities are mismatched.

- Before the transformation of the ER model to relations designer work with ER diagrams, but after the transformation they work with relation schemes. These relations may need normalization. Normalization is done through a combination of decomposition and synthesis techniques.

Their approach are to capture the real-world semantics in an improved EER model, transform the EER model to a normalized EER model, and generate normalized relations.

2.2 A Language for Defining Conceptual Schema (CSL)

In [5] Breutmann, Falkenberg and Mauer describe a high level data definition language, CSL((Conceptual Schema Language)), for defining conceptual schemes. The language provides descriptive and procedural elements, so static aspects and dynamic behavior of data can be described. The proposal CSL provides: standard types, object types and association types:

- Standard operations are: $=, <, >, +, -, *$ and standard types are like INTEGER, REAL, CHAR, STRING, DIGIT
- Within an object type only those characteristics can be described which hold for all objects of a certain type
- The association type definition consists of the unique association type name, the participating object types together with their roles and simple occurrence frequencies and the identifier of the association type together with the candidate identifiers.

2.3 HERM (Higher-Order Entity Relationship Model) Metamodel

Bernhard Thalheim introduces the Higher-Order Entity Relationship Metamodel in [20], called HERM, providing an interesting conceptual data model, but it is strictly founded in theory.

The Higher-Order Entity Relationship Model is an extension of the Entity Relationship Model. Schemata in the Higher-Order Entity Relationship Model can be mapped automatically to relational database schemata. One key feature of the HERM is the nesting of attributes.

2.4 REMORA Metamodel

In [17] the REMORA methodology and an expert design tool, that supports the design of information systems, are described.

The REMORA methodology provides a consistent set of models, languages, methods and software tools to design and implement large and semantically complex information systems. The conceptual models use, there is a more direct and efficient interaction between the designers and the users, this allows the definition of the information system conceptual schema.

2.5 Barker Metamodel

The Barker model was introduced by Richard Barker in [1] and modified slightly by the Oracle Corporation.

The pedantic problem with the Barker model is that one needs to fully understand relational database theory to understand why the Barker model is done the way it is. We present the Barker model here because the way it unfolds is a bit different from the Chen model. The Chen model focuses on modeling data, whereas the Barker model adapts the data to the relational database concurrently with the design. Therefore, the ER design methodology for the Barker model will develop differently from the Chen model. Further, the Barker model does not have some of the conventions used in the Chen model. The Barker model does not directly use the concept of composite attributes, multi-valued attributes, or weak entities, but rather handles these concepts immediately in light of the relational model. Because the Barker model is so close to the relational model to begin with, the mapping rules are trivial the mapping takes place in the diagram itself.

A Barker model uses soft boxes for entities (with the entity name in capital letters), and there is a line separating the entity name from the attributes (and the attribute names are in lowercase letters).

A Barker model does not place the attributes in ovals (as the Chen model does), but rather lists the attributes below the entity name. All attributes in a Barker model are considered simple or atomic, as in relational databases.

The model does not have the concept of composite attributes. In Barker model, the primary key has a # in front of the name of the attribute.

A primary key has to be a mandatory attribute in a relational database, but again, all mandatory attributes here are not necessarily unique identifiers.

In the Barker model, a relationship is represented by a line that joins two entities together. There is no diamond denoting the relationship (as we saw in the Chen model). The relationship phrase for each end of a relationship is placed near the appropriate end (entity) in lower case.

2.6 Data Schema Integration Methodology

In 1983 many problems of data integration are addressed but not all of them and the whole area of data integration is not at a mature stage.

In [3] Batini and Lenzerini describe a methodology for data schema integration consists of three steps: conflict analysis, merging and final enrichment and restructuring of the schema.

In the conflict analysis phase all incoherences should be detected and solved. The main tasks are naming conflicts analysis, resolution of synonymy and homonyms and the modeling compatibility analysis. In the merging phase schemata are merged into a unique draft integrated schema. The main tasks of the third phase are analysis of interschema properties, analysis of redundant paths or schema restructuring. In this paper an effort is made to face all the relevant issues that can arise in the integration of several conceptual schemata and provide for all of them a design strategy

2.7 DATAID

DATAID-1 [7, 8] is a manual methodology dealing with the design of centralized databases. DATAID-1 is described by Valeria De Antonellis and Antonio Di Leva in 1985 and extends the basis of practical experience of several projects in banks and government offices. They also illustrate a case study of database design using the DATAID approach. The case study refers to a banking environment.

DATAID consists of the following methodological phases: requirement collection and analysis, local views design, local views integration, logical design, physical design. These phases are followed in the development of the case study of database design.

3 UML Metamodel

The Unified Modeling Language(UML)[13, 14] is a non-proprietary object modeling and specification language used in software engineering. UML includes a standardized graphical notation that may be used to create an abstract model of a system: the UML model. While UML was designed to specify, visualize, construct, and document software-intensive systems is not restricted to modeling software.

UML has its strengths at higher, more architectural levels and has been used for modeling hardware (engineering systems) and is commonly used for business process modeling, systems engineering modeling, and representing organizational structure.

4 Quality measures and Transformation of conceptual schemes – ANNAPURNA

In the 1980'ies conceptual schemes are recognized and accepted as an important tool for the design and evolution of integrated databases and knowledge-based systems, but the question of quality of conceptual schemes is largely ignored by researchers. In relational database theory quality is defined by the presence or absence of certain normal forms; the definition of quality was very restrictive because a conceptual schema is either good or bad. Quality measures are also ignored, but transformation of conceptual schemes are explored systematically.

Christoph F. Eick represents in his paper [9] the back end of a conceptual schema design methodology, called **ANNAPURNA**. This methodology aims to automate conceptual schema design focusing on the transformation and evaluation of conceptual schemes based on quality measures and transformations that has a theoretical foundation. A general framework for the specification of conceptual schema transformations is proposed and algorithms for evaluation and transformations are provided.

5 ER-SP and ER-SPL our Approach regarding ER Model

5.1 Entities and attributes

At this level we consider entities and relationships.

Definition 1(ENTITY)

The set of entities is a finite set of names

$$ENTITY \subseteq N \tag{1}$$

In [6] Chen argue that "An entity is a thing which can be distinctly identified."

Our approach is that each entity e_i belongs to a Entity Set E_i, $e_i \in E_i$, induces *entity type's* $t_{e_i} \in T_{E_i}$. T_{E_i} is a set of domain type names. All entities have a distinct name; in particular, an entity name may not be used again to define another entity with a different type.

$$\forall e_1, e_2 \in T_E : (e_1 : t_{e_1} \rightarrow E_i, e_2 : t_{e_2} \rightarrow E_i) \Rightarrow e_1 = e_2$$

Entities with the same name may, however, appear in different ER Models that are not related by generalization.

Definition 2(ATRIBUTES)

An attribute a_i, could be formally defined as a function which maps from an entity, E_i or a relationship, R_i into a value set, V_i or a cartesian product $\prod_{1 \le j \le n} V_{ij}$ of value sets:

$$a : E_i \, or R_i \ \rightarrow V_i \, or \, V_{i_1} \times V_{i_2} \times V_{i_3} \times \ldots \times V_{i_n} \quad (2)$$

5.2 Relationships

A relationship is an association among entities.

Definition 3(RELATIONSHIP)

A relationship R_i, is a mathematical relation among n entities each taken from an entity set:

$$\{[e_1, e_2, \cdots, e_n] | e_1 \in E_1, e_2 \in E_2, \cdots, e_n \in E_n\} \quad (3)$$

and each tuple of entities, $[e_1, e_2, \cdots, e_n]$, is a relationship. Note that the E_i in the above definition may not be distinct.

5.3 Primary Key and Foreign Key

Definition 4(CANDIDATE KEY)

An attribute or set of attributes that uniquely identifies individual occurrences of an entity type.

$$V(a_1^i, \ldots, a_k^i) \neq V(a_1^j, \ldots, a_k^j), \, \forall \, i \neq j \quad (4)$$

where $(a_1^i, \ldots, a_k^i)$ is candidate key, a subset of entity $e_i = (a_1^i, \ldots, a_n^i) \, with \, k \ \le n$

Definition 5(PRIMARY KEY)

A unique identifier for a row in a table in relational database; A selected candidate key of an entity.

$$V(a_1^i, \ldots, a_k^i, \ldots, a_n^i) \neq V(a_1^j, \ldots, a_k^j, \ldots, a_n^j) \quad (5)$$

$\forall \, i \neq j, with \, k \ \le n$

Definition 6(FOREIGN KEY)

An attribute that is a primary key of another relation (table). A foreign key is how relationships are implemented in relational databases.

$$e_i = (a_1^i, \ldots, a_k^i, \ldots a_n^i) \quad (6)$$

where $(a_1^i, \ldots, a_k^i)$ is a primary key, a subset of entity $e_i = (a_1^i, \ldots, a_n^i) \ with \ k \ \le \ n$ and $e_j = (b_1^j, \ldots, b_l^j, \ldots b_m^j, a_1^i, \ldots, a_k^i)$ is the corespondent entity from relationship and $(a_1^i, \ldots, a_k^i)$ represents the foreign key of $e_j, \forall \, i \neq j$.

5.4 Stored Procedures

Stored Procedures:(**procedures** and **functions**) are SQL subroutine statements in a RDBMS, for use by all application including the control statements that allow repetition(LOOP) and conditional execution(IF and CASE statements)

A procedure is performed using the SQL CALL statement.

5.5 From Entity Relationship Model to Relational Model

Entity Relationship Model(conceptual level) could be translate in Relational Model(physical level) in a natural mod using the following transformation [11]:

- any entity became a table;
- any attribute from entity became a field in the repectiv table;
- any relationship became a special table or an primary key in one of two tables and referencing in the corespondent table.

6 Conclusion

The model proposes a new entity type called Stored Procedures type and the language behind him in idea to be platform independent in relation with SQL dialects and their Procedural Languages. The concept of stored procedures type is abstract, helps in syntactic and semantic modeling, and is required in physical implementation. A stored procedures entity type along with his language captures the syntactic and semantics of an RDBMS schema in his dynamical evolution.

Acknowledgments: The author would like to acknowledge the financial support by "Petru Maior" University of Târgu-Mureş, through the Grant no. 1783/2006).

References:

[1] R. Barker, *CASE*METHOD Entity Relationship Modeling*, Workingham, England: Addison-Wesley, 1990.

[2] C.A. Comes, L.O. Marian, N. Ghişoiu, I. Bircea, *Business Process Management with Unified Modeling Language*, WSEAS TRANSACTIONS ON COMPUTERS, pp 361-366, Issue 2, WSEAS Press, Vol. 6, February 2007,

[3] C. Batini, M. Lenzerini, *A Methodology for Data Schema Integration in the entity relationship model*, ER 1983: pp 413-420

[4] H. Bedoya, F. Cruz, D. Lema and S. Singkorapoom, *Stored Procedures, Triggers, and User-Defined Functions on DB2 Universal Database*, International Technical Support Organization, October 2006

[5] Breutmann, Falkenberg, Mauer: *CSL: A Language for Defining Conceptual Schema*, North Holland Publ.Company, 1979

[6] P. P. Chen, *The Entity-Relationship Model: Toward a Unified View of Data*, ACM Trans. Database Systems, Volume 1, Number 1,1976, pp 9-36, ACM Press, New York, NY, USA.

[7] V. De Antonellis, A. Di Leva, *A Case Study of Database Design using the DATAID Approach*, Information Systems, Vol. 10, pp 339 359, 1985.

[8] V. De Antonellis, A. Di Leva, *DATAID-1: A database design methodology, Information Systems* 10(2), pp 181 195, 1985.

[9] C. F. Eick, *A Methodology for the Design and Transformation of Conceptual Schemes*, Proceedings of the 17th International Conference on Very Large Data Bases, Barcelona, September 1991

[10] D. W. Embley and T. W. Ling, *Synergistic Database Design with an Extended Entity-Relationship Model*, Proceedings of the Eight International Conference on Entity-Relationship Approach to Database Design and Querying,1990, pp 111-128, North-Holland

[11] F. E. Ipate and M. Popescu, *Dezvoltarea aplicaţiilor de baze de date în Oracle8 şi Forms6*, Editura BIC ALL, 2000, pp 25, Bucureşti

[12] *MySQL 5.1 Reference Manual*, Document generated on: 2006-10-29 (revision: 3792), Copyright ©1997-2006 MySQL AB

[13] OMG. Object Management Group, *Unified Modeling Language (UML) Specification: Infrastructure*, version 2.0 (ptc/03-09-15)

[14] OMG. Object Management Group, *Unified Modeling Language (UML) Specification: Superstructure, version 2.0*, Revised Final Adopted Specification (ptc/0410-02)

[15] *PostgreSQL 7.2 Programmers Guide*, Copyright ©1996-2001 by The PostgreSQL Global Development Group

[16] *PostgreSQL 7.2 Tutorial*, Copyright ©1996-2001 by The PostgreSQL Global Development Group

[17] C. Rolland: *An Information System Methodology supported by an Expert Design Tool*, University of Paris, 1988, Elsevier Science Publishers

[18] Adaptive ®*Server AnywhereSQL Reference*, Part number: DC38129-01-0902-01 Last modified: October 2004, Copyright ©19892004 Sybase, Inc.

[19] Teradata Database, *SQL Reference*, Fundamentals, Release V2R6.1, B035-1141-115A May 2006, NCR, Teradata and BYNET are registered trademarks ™of NCR Corporation.

[20] B. Thalheim, *HERM: Putting theory into practice* IDS-92, IFIP-Workshop on Database Intellectualization, May 1992, Control Systems and Machines, 1992, 5/6, pp 85-93, Kaliningrad