DBA Companion: A Tool for Logical Database Tuning

Stéphane Lopes
Laboratoire PRiSM, CNRS UMR-8636,
45, avenue des États-Unis
78035 Versailles Cedex FRANCE
stephane.lopes@prism.uvsq.fr

Fabien De Marchi Jean-Marc Petit
Laboratoire LIMOS, CNRS UMR-6158,
Complexe scientifique des Cézeaux
63177 Aubière Cedex FRANCE
{demarchi,jmpetit}@isima.fr

1. Motivations

Understanding data semantics from existing relational databases is important for several applications such as database maintenance and analysis, database re-engeenering, data warehouse design or query optimization. Such a knowledge is carried out in particular by integrity constraints. Among these constraints, functional dependencies (FDs), which generalize the notion of key, and inclusion dependencies (INDs), which generalize the notion of foreign key, are by far the most common integrity constraints in the real world. In the best case, these constraints were specified during the database design and, hence, are available in the DBMS. Nevertheless, a seeming invariant of production databases is that they become disordered over time due, for instance, to incorrectly entered data or incorrect use of the database. In this case, integrity constraints have to be extracted from the database.

Among applications which can benefit from understanding existing databases, we have chosen to focus more particularly on *logical database tuning*, i.e. the optimization of database schema with respect to dependencies satisfied by the database. Providing knowledge about dependencies satisfied by a database can help a DBA to address several tasks such as key or foreign key enforcement or denormalized relation detection. Note that simplifying DBMS administration has been identified as a challenge for the database community [1].

2. A tool for database schema analysis

In this paper, we present a tool called *DBA Companion* which can be a help to deal with the understanding of existing relational databases. The prototype integrates algorithms dedicated to database analysis. This task rests on data mining techniques which allow to design efficient algorithms. Emphasis is put on algorithm efficiency to be able to address operational situations, and then discover FDs and INDs satisfied in a database instance.

Once integrity constraints satisfied by a given database were discovered, it remains to select the most relevant ones, i.e. eliminate accidental dependencies for instance. We propose two alternatives to provide a help to choose interesting dependencies. The former is based on example databases: an *informative Armstrong database* is an alternative representation of a set of dependencies which can be a help to visualize existing or missing dependencies. The latter is based on examining a workload of SQL statements: for instance, attributes involved in join orders convey some hints to select relevant dependencies.

For a survey of our propositions, the reader is refered to [3] and to the project web site [2].

Practically, the user opens an ODBC connection to a data source and select the relations to analyse. He can afterward perform several algorithms to study these relations (exact and approximate FD inference, key inference, exact and approximate IND inference, informative Armstrong relation (IAR) generation). The tool follows a loosly coupled approach with the underlying DBMS for algorithm execution. For instance, IARs are generated in the DBMS which allows the user to modify them and to reiterate the analysis process from these new relations.

References

[1] P. A. Bernstein, M. L. Brodie, S. Ceri, D. J. DeWitt, M. J. Franklin, H. Garcia-Molina, J. Gray, G. Held, J. M. Hellerstein, H. V. Jagadish, M. Lesk, D. Maier, J. F. Naughton, H. Pirahesh, M. Stonebraker, and J. D. Ullman. The Asilomar report on database research. *Sigmod Record*, 27(4):74–80, 1998.

[2] F. De Marchi, S. Lopes, and J.-M. Petit. The DBA Companion project. http://www.isima.fr/~demarchi/dbacomp.

[3] F. De Marchi, S. Lopes, J.-M. Petit, and F. Toumani. Analysis of databases at the logical level: The DBA compagnon project. *Sigmod Record*, 32(1):47–52, 2003.

Proceedings of the 20th International Conference on Data Engineering (ICDE'04)
1063-6382/04 $ 20.00 © 2004 IEEE