Analyzing Mappings and Properties in Data Warehouse Integration

Domenico Beneventano, Marius Octavian Olaru, Maurizio Vincini[*]

Department of Engineering "E. Ferrari", University of Modena and Reggio Emilia, Italy.

Abstract

The information inside the Data Warehouse (DW) is used to take strategic decisions inside the organization that is why data quality plays a crucial role in guaranteeing the correctness of the decisions. Data quality also becomes a major issue when integrating information from two or more heterogeneous DWs. In the present paper, we perform extensive analysis of a mapping-based DW integration methodology and of its properties. In particular, we will prove that the proposed methodology guarantees coherency, meanwhile in certain cases it is able to maintain soundness and consistency. Moreover, intra-schema homogeneity is discussed and analysed as a necessary condition for summarizability and for optimization by materializing views of dependent queries.

References

[1] W. H. Inmon, Building the data warehouse, 3rd ed. John Wiley & Sons, Inc., 2002.

[2] M. Preis and J. Seitz, "Challenges and conflicts integrating heterogeneous data warehouses in virtual organisations," International Journal of Networking and Virtual Organisations, vol. 11, no. 3/4, pp. 329-335, 2012.

[3] A. P. Sheth and J. A. Larson, "Federated database systems for managing distributed, heterogeneous, and autonomous databases," ACM Computing Surveys, vol. 22, no. 3, pp. 183-236, September 1990.

[4] S. Abiteboul, I. Manolescu, and N. Preda, "Constructing and querying peer-to-peer warehouses of XML resources," Proc. 21st International Conference on Data Engineering (ICDE 2005), IEEE Press, April 2005, pp. 1122-1123.

[5] S. Bergamaschi, M. O. Olaru, S. Sorrentino, and M. Vincini, "Dimension matching in peer-to-peer data warehousing," Proc. IFIP Working Group 8.3 International Conference on Decision Support Systems, 2012, pp. 149-160.

[6] F. Guerra, M. O. Olaru, and M. Vincini, "Mapping and integration of dimensional attributes using clustering techniques," E-Commerce and Web Technologies, Springer press, 2012, pp. 38-49.

[7] R. Torlone, Interoperability in data warehouses, Encyclopedia of Database Systems, Springer, pp. 1560-1564, 2009.

[8] D. Beneventano, S. Bergamaschi, G. Gelati, F. Guerra, and M. Vincini, "MIKS: an agent framework supporting information access and integration," Intelligent Information Agents, vol. 2586, pp. 22-49, 2003.

[9] S. Bergamaschi, G. Gelati, F. Guerra, and M. Vincini, "An intelligent data integration approach for collaborative project management in virtual enterprises," World Wide Web, vol. 9, no. 1, pp. 35-61, March 2006.

[10] A. Halevy, A. Rajaraman, and J. Ordille, "Data integration: the teenage years," Proc. of the 32nd international conference on Very large data bases, September 2006, pp. 9-16.

[11] R. Torlone, "Two approaches to the integration of heterogeneous data warehouses," Distributed and Parallel Databases, vol. 23, no. 1, pp. 69-97, February 2008.

[12] R. Kimball and M. Ross, The data warehouse toolkit: the complete guide to dimensional modeling, New York: John Wiley & Sons, Inc., 2002.

[13] M. Banek, B. Vrdoljak, A. M. Tjoa, and Z. Skocir, "Automated integration of heterogeneous data warehouse schemas," International Journal of Data Warehousing & Mining, vol. 4, no. 4, pp. 1-21 October-December 2008.

[14] D. Beneventano, S. Bergamaschi, F. Guerra, and M. Vincini, "The SEWASIE network of mediator agents for semantic search," Journal of Universal Computer Science, vol. 13, no. 12, pp. 1936-1969, January 2007.

[15] W. Lehner, J. Albrecht, and H. Wedekind, "Normal forms for multidimensional databases," Proc. International Conference on Scientific and Statistical Database Management, IEEE Press, July 1998, pp. 63-72.

* Corresponding author. E-mail address: Domenico.Beneventano@unimore.it

[16] C. A. Hurtado, C. Gutierrez, and A. O. Mendelzon, "Capturing summarizability with integrity constraints in OLAP," ACM Transactions on Database Systems, vol. 30, no. 3, pp. 854-886, September 2005.

[17] H. V. Jagadish, L. V. S. Lakshmanan, and D. Srivastava, "What can Hierarchies do for data warehouses," Proc. 25th International Conference on Very Large Data Bases, September 1999, pp. 530-541.

[18] L. Cabibbo and R. Torlone, "On the integration of autonomous data marts," Proc. International Conference on Scientific and Statistical Database Management, IEEE Press, July 2004, pp 223-234.

[19] L. Cabibbo and R. Torlone, "Integrating heterogeneous multidimensional databases," Proc. International Conference on Scientific and Statistical Database Management, IEEE Press, June 2005, pp. 205-214.

[20] M. Golfarelli, F. Mandreoli, W. Penzo, S. Rizzi, and E. Turricchia, "OLAP query reformulation in peer-to-peer data warehousing. Information Systems," vol. 37, no. 5, pp. 393-411, July 2012.

[21] V. Harinarayan, A. Rajaraman, and J. D. Ullman, "Implementing data cubes efficiently," Proc. ACM SIGMOD international conference on Management of data, ACM Press, June 1996, pp. 205-216.

[22] M. Rafanell and A. Shoshani, "Storm: a statistical object representation model," Proc. Statistical and Scientific Database Management, vol. 420 of Lecture Notes in Computer Science, Springer Press, 1990, pp. 14-29.

[23] M. Banek, B. Vrdoljak, A. M. Tjoa, and Z. Skocir, "Automating the schema matching process for heterogeneous data warehouses," Proc. 9th International Conference on Data Warehousing and Knowledge Discovery, Springer, 2007, pp. 45-54.

[24] S. Bergamaschi, C. Sartori, F. Guerra, and M. Orsini, "Extracting relevant attribute values for improved search," IEEE Internet Computing, vol. 11, no. 5, pp. 26-35, September 2007.