

An Object-oriented Quality Framework with Optimization Models for Managing Data Quality in Data Warehouse Applications

Chung-Yang Chen^{1,*}, Yu-Liang Chi², and Philip Wolfe³

¹Department of Information Management, Chang Gung University, 259 Wen-Hwa 1st Road, Kwei-Shan, Tao-Yuan, 333, Taiwan, R.O.C.

² Department of Management Information Systems, Chung-Yuan Christian University, Chung-Li 320, Taiwan

³Department of Industrial Engineering, Arizona State University, Tempe, AZ 85287-5906, USA

Received December 2004; Revised March 2005; Accepted May 2005

Abstract—Data quality is an important issue, especially in large-scale data applications such as data warehousing (DW). The validity (a super quality type specialized by accuracy, completeness, consistency, and currency) of data in fact has corresponding impacts on ad-hoc decisions. To ensure quality, improvement actions such as edit check, imputation, and audit et al. are applied. Yet these utilize and consume resources and time, particularly for large sets of data which get more critical as achieving zero-defects. In this paper, an object-oriented and multi-dimensional quality framework is suggested in order to comprehensively realize data quality. Two simple mixed binary integer programming optimization models based on the quality framework are presented to study the cost issues and investment allocation according to different quality aspects in DW. An example is then given to illustrate the managerial use of the models.

Keywords—Data quality, Object concept, Data warehousing, Crashing costs, Mixed binary integer programming

* Corresponding author's email: cychen@mail.cgu.edu.tw