Generic Framework for Better Choosing Between Data Integration Types (GFCBDIT) During Build Business Intelligence Applications

[1]Ayad Hameed Mousa, [2]Norshuhada Shiratuddin, [3]Muhamad Shahbani Abu Bakar
[1]*University of Karbala*[1]*University Utara Malaysia, Iraq, maryemayad@yahoo.com*
[2]*University Utara Malaysia, Malaysia, shuhada@uum.edu.my*
[3]*University Utara Malaysia, Malaysia, shabani@uum.edu.my*

Abstract

Organizations design their own business intelligent (BI) systems to fundamentally support and improve their decision support systems. Decision-making process is indispensable in a world of constant changes; two important changes are the rapid response of organizations and companies in such process, and recognition of data integration as the backbone of most BI systems. For designers, the existence of two main types of data integration (i.e., physical and virtual integration) causes difficulty in selecting the best one during the process of designing their own BI applications. In this paper, the design and implementation of GFCBDIT, a general framework that can be used by developers of any BI application, is proposed to help in the selection of the most suitable type of data integration. Comparative analysis between two techniques related with data integration is performed, and a prototype of the proposed framework is constructed to enable its use. The prototype is based on a graphical user interface used by BI developers to select the most convenient type of data integration. The prototype is tested with a system user feedback using Computer System Usability Questionnaire, which measures satisfaction and system usability.

Keywords: *Data Integration, Data Virtualization, Data Warehouse, Business Intelligence*

1. Introduction

Data integration pertains to the exchange of meaningful information between systems regardless of whether these systems are designed to work together or not. Accordingly, this exchange can be challenging for any organization, particularly when organizations intend to develop large projects, given that data integration is a costly and labor-intensive process [1, 2]. Data integration is the process of collecting, processing, and clearing data obtained from different sources to generate a unified vision of those data delivered to BI tools to help decision makers make better decisions [3–6]. The generation of knowledge and use of data integration have induced organizations to develop a broad interest in data integration, which have resulted in the emergence of a number of techniques and algorithms, approaches, and solutions from the works of [1, 3, 6–8]. The majority of the researchers in the field of data integration, some of whom are cited in the current paper, categorized data integration into two types that are based on the nature of integration and the behavior of techniques used. The first type is physical integration, which typically uses data warehouse (DW) as a principal technique [9–15], whereas the second type is virtual integration that generally adopts data virtualization (DV) as its primary technique [14, 16–20]. Figure 1 illustrates the architecture of data integration. Subsequent paragraphs thoroughly explain these techniques to determine their respective advantages and disadvantages.

General rules, frameworks, and methods guide designers in their quest to determine the type of data integration that must be used in either forming new applications or improving existing applications for their organizations. However, these guiding elements are missing or lacking in accuracy to remove the gap, given that data integration is time consuming, complicated, and costly, which may cause a defect or failure. Consequently, the entire decision-making process may be adversely affected.