Accepted in The International Conference on Data and Information Science (ICoDIS)
Telkom University, 5-6 December 2017

Survey Data and Metadata Modelling Using Document-Oriented NoSQL

Lutfi Rahmatuti Maghfiroh

School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
lutfirm@stis.ac.id

I Gusti Bagus Baskara Nugraha
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
baskara@stei.itb.ac.id

Abstract. Survey data that are collected from year to year have metadata change. However it need to be stored integratedly to get statistical data faster and easier. Data warehouse (DW) can be used to solve this limitation. However there is a change of variables in every period that can not be accommodated by DW. Traditional DW can not handle variable change via Slowly Changing Dimension (SCD). Previous research handle the change of variables in DW to manage metadata by using multiversion DW (MVDW). MVDW is designed using relational model. Some researches also found that developing nonrelational model in NoSQL database has reading time faster than the relational model. Therefore, we propose changes to metadata management by using NoSQL. This study proposes a model DW to manage change and algorithms to retrieve data with metadata changes. Evaluation of the proposed models and algorithms result in that database with the proposed design can retrieve data with metadata changes properly. This paper has contribution in comprehensive data analysis with metadata changes (especially data survey) in integrated storage.

1. INTRODUCTION

Data and information are important to an organization. Organization can use data analysis to make a better decision. Statistical data is usually collected through census and survey using questionnaire. To get up-to-date analysis, organization needs to integrate data storage to get statistical data faster and easier. However, data collected from year to year may have metadata change. One of ways to store the data is by using data warehouse.

Data warehouse (DW) is a subject-oriented, integrated, time-varying and nonvolatile data collection [1]. We can use data warehouse to store and organize the data so that we can use it for analyzing easily. DW can integrate all metadata and make it easy to maintain.

To manage the metadata change in DW, Kimball [2] introduce slowly changing dimension (SCD) concept to track the changes. However, it can only handle change of member attribute structure in multidimensional model. It cannot handle change of non-member attribute structure, for example, adding new attribute, removing attribute, etc. Body et al. in [3] introduce a temporal multidimensional