



International Conference on Intelligent Computing, Communication & Convergence
(ICCC-2015)

Conference Organized by Interscience Institute of Management and Technology,
Bhubaneswar, Odisha, India

Literature Review of Data model Quality metrics of Data Warehouse

Anjana Gosain^a, Heena^a

^aGGSIPIU, Sector 16C Dwarka, New Delhi, India

Abstract

Quality of data warehouse is very crucial for managerial strategic decisions. Multidimensional data modeling has been accepted as a basis for data warehouse, thus data model quality has a great impact on overall quality of data warehouse. Metrics act as a tool to measure the quality of data warehouse model. Various authors have proposed metrics to assess the quality attributes of conceptual data models for data warehouse such as understandability, maintainability etc. All the related research work inspires us to investigate the metrics proposed to measure data warehouse data model quality, the various quality factors assessed and to provide a ground work for research advancement in this field. A total of 22 studies were selected and analyzed to identify the various validation techniques used to prove usage and practical utility of metrics and the quality factors measured by these metrics. Opportunities for future work lie in the gaps that were found in the validation of the metrics and the lack of quality factors measured.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of scientific committee of International Conference on Computer, Communication and Convergence (ICCC 2015)

Keywords: Literature review, Data warehouse Data Model Quality Metrics, Conceptual Data model quality metrics, Understandability

1. Introduction

A Data Warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of

management's decision making process¹. But lack of quality in the information being provided by data warehouse can lead to bad strategic decisions. Thus, information quality in data warehouse needs to be assured which further depends on presentation quality, data quality and data model quality (both physical and logical model². It has been widely accepted that data modeling for data warehouses is based on multidimensional modeling and OLAP tools directly access multidimensional schemas of data warehouses to help decision makers. Thus, quality of data model has great influence on the overall quality of data warehouse. Multidimensional schemas organize data in facts and dimensions. Facts contain the measures for analysis along the dimensions. Dimensions contain attributes, granularity levels and hierarchies for better representation of data. All these structural elements and the relationship between them affect the complexity of the schema and this structural complexity notably determines the data model quality. Earlier in²¹ quality of data warehouse conceptual models has been considered as intuitive notions but these intuitive notions of quality needs to be substituted by a set of quantitative measures to increase objectivity. Metric provides a way to measure a quality factor in a consistent and objective manner². Various authors have proposed several metrics to measure the quality factors of conceptual data models such as maintainability, understandability etc^{6,8,10,14,15,18,22,24} for data warehouse. These metrics are validated theoretically and empirically. Therefore, in this paper we present a Systematic literature review of data model quality metrics of data warehouse in order to know the current state of art and explore the opportunities for further research.

The remainder of the paper is organized as- Section 2 describes the steps taken to conduct SLR for our research field. Section 3 summarizes various metrics proposed to access the quality of conceptual data models of data warehouse. Section 4 deliberates about the findings of the SLR elaborating empirical evidences, data warehouse quality factors and existing gaps in the research area. Section 5 concludes the work.

2. Systematic Literature Review (SLR) Process

SLR is a form of secondary study that uses a well-defined methodology to identify, analyze and interpret all available evidence related to a specific research question in a way that is unbiased and (to a degree) repeatable³. Following the guidelines stated by Kitchenman³, series of action followed to direct our research are as shown in Fig -1 and are detailed in the following subsections.

2.1 Research Questions

Research questions forms the important part of SLR as they only direct the research with the intent of finding the appropriate studies, analyzing them and extracting useful data to find their answers. Our work intend to find out answers for the following research questions -

RQ 1: What are the existing empirical evidences for various metrics proposed to measure conceptual data model quality for data warehouse?

RQ 2: What are the quality factors that have been related to proposed metrics?

RQ 3: What are the existing gaps in the current research in the use of quality metrics in data model of data warehouse?

2.2 Search Process

Sources of Information

Following 5 online databases were used to extract required studies for review-

- a) IEEE Xplore (<http://ieeexplore.org/>)
- b) Springer LNCS (<http://www.springer.com/gp/>)
- c) Science Direct (<http://www.sciencedirect.com/>)
- d) ACM digital library (<http://dl.acm.org/>)
- e) Google Scholar (<http://scholar.google.co.in/>)

Initial search string was needed to extract relevant studies from the databases and search string defined for our work was:

("data quality" OR "data warehouse quality" OR "multidimensional data quality" OR " data warehouse conceptual model") AND ("metrics" OR "measures").

Download English Version:

<https://daneshyari.com/en/article/489959>

Download Persian Version:

<https://daneshyari.com/article/489959>

[Daneshyari.com](https://daneshyari.com)