Facilitate Effective Decision-Making by Warehousing Reduced Data:
Is It Feasible?

Faten Atigui, Centre d'Etude et De Recherche en Informatique et Communications (CEDRIC), Conservatoire National des Arts et Métiers, Paris, France

Franck Ravat, Institut de Recherche en Informatique de Toulouse (IRIT), Université Toulouse I Capitole, Toulouse, France

Jiefu Song, Institut de Recherche en Informatique de Toulouse (IRIT), Université Toulouse I Capitole, Toulouse, France

Olivier Teste, Institut de Recherche en Informatique de Toulouse (IRIT), Université Toulouse II Jean Jaurès, Toulouse, France

Gilles Zurfluh, Institut de Recherche en Informatique de Toulouse (IRIT), Université Toulouse I Capitole, Toulouse, France

ABSTRACT

The authors' aim is to provide a solution for multidimensional data warehouse's reduction based on analysts' needs which will specify aggregated schema applicable over a period of time as well as retain only useful data for decision support. Firstly, they describe a conceptual modeling for multidimensional data warehouse. A multidimensional data warehouse's schema is composed of a set of states. Each state is defined as a star schema composed of one fact and its related dimensions. The derivation between states is carried out through combination of reduction operators. Secondly, they present a meta-model which allows managing different states of multidimensional data warehouse. The definition of reduced and unreduced multidimensional data warehouse schema can be carried out by instantiating the meta-model. Finally, they describe their experimental assessments and discuss their results. Evaluating their solution implies executing different queries in various contexts: unreduced single fact table, unreduced relational star schema, reduced star schema and reduced snowflake schema. The authors show that queries are more efficiently calculated within a reduced star schema.

Keywords: Conceptual Modeling, Data Reduction, Experimental Assessment, Multidimensional Design, Reduction Operators

DOI: 10.4018/ijdsst.2015070103

INTRODUCTION

Nowadays, decision support systems are mostly based on Multidimensional Data Warehouse (MDW). A MDW schema is based on facts (analysis subjects) and dimensions (analysis axis). A fact includes analysis indicators while dimensions contain analysis parameters. The analysis parameters are organized according to their hierarchical level in order to classify the parameters from the lowest granularity to the highest granularity.

In a MDW, data is stored permanently and new data is steadily added. Hence, a MDW stores a huge volume of data in which the analyst may well get lost during her/his analyses. On the other hand, the relevance of MDW data decreases with time: detailed information is generally considered relevant for recent data (Skyt, Jensen, & Pedersen, 2008), while more aggregated information can usually satisfy the needs of analysis with older data. For instance, a decision-maker may have interest in analyzing published news by subthemes for the last four years, but this level of granularity may be proved inappropriate if the analysis was carried out over an older period because most of today's subthemes did not exist before: it is impossible to carry out analyses over published news by subthemes but by a higher granularity level which stays stable over time, such as news' theme.

Facing large volumes of both relevant and irrelevant data, our aim is to increase the performance of query treatment and especially to facilitate the analysts' task by providing only pertinent data over time. Our aim is to provide a multidimensional analysis framework adapted to analysts' needs, allowing them to remove the temporal granularity levels which are irrelevant for their analyses. As detailed data value decreases with time, we implement selective deletions at low levels of granularity. This reduction is achieved mainly through progressive data aggregation: older data is synthesized.

This paper is composed of the following sections: Section 2 describes a state of the art of data reduction. Section 3 defines our conceptual model of multidimensional data based on reductions. Section 4 presents a meta-model for managing MDW composed of states. Section 5 provides experimental assessments to evaluate our solution in various implementation environments.

RELATED WORK

Reducing data allows us both to decrease the quantity of irrelevant data in decision making and to increase future analysis quality (Udo & Afolabi, 2011). In the context of decision support, data reduction is a technique originally used in the field of data mining (Okun & Priisalu, 2007; Udo & Afolabi, 2011).

In the DW context, (Garcia-Molina, Labio, & Yang, 1998) were the first to define solutions for data deletion. More precisely, they study data expiration in materialized views so that they are not affected and can be maintained after updates with the help of a set of standard predefined views.

In the multidimensional area, (Chen et al., 2002) propose an architecture allowing the integration of data streams into a MDW and reduce the size. The size reducing is predefined and automatically executed by partially aggregating the data cube; it makes sure the detailed information is only available during a time interval. Nevertheless, this work only focuses on the fact table. (Skyt et al., 2008) presents a technique for progressive data aggregation of a fact. This study intends to specify data aggregation criteria of a fact due to higher levels of dimensions. The authors also provide techniques to query reduced multidimensional objects. As mentioned in (Iftikhar & Pedersen, 2011), this work is highly theoretical but it fails to provide us a concrete example of implementation strategy. In (Iftikhar & Pedersen, 2011), a gradual data aggregation

A Decision-Making Support for Business Process Outsourcing to a Multi-Cloud Environment

Karim Zarour and Djamel Benmerzoug (2019). *International Journal of Decision Support System Technology (pp. 66-92).*

www.igi-global.com/article/a-decision-making-support-for-business-process-outsourcing-to-a-multi-cloud-environment/216942?camid=4v1a

The Traveling Salesman Problem, the Vehicle Routing Problem, and Their Impact on Combinatorial Optimization

Gilbert Laporte (2012). *Decision Making Theories and Practices from Analysis to Strategy (pp. 342-352).*

www.igi-global.com/chapter/traveling-salesman-problem-vehicle-routing/65970?camid=4v1a