Chapter XVI

Analyses and Evaluation of Responses to Slowly Changing Dimensions in Data Warehouses

Lars Frank
Copenhagen Business School, Denmark

Christian Frank
Copenhagen Business School, Denmark

ABSTRACT

A Star Schema Data Warehouse looks like a star with a central, so-called fact table, in the middle, surrounded by so-called dimension tables with one-to-many relationships to the central fact table. Dimensions are defined as dynamic or slowly changing if the attributes or relationships of a dimension can be updated. Aggregations of fact data to the level of the related dynamic dimensions might be misleading if the fact data are aggregated without considering the changes of the dimensions. In this chapter, we will first prove that the problems of SCD (Slowly Changing Dimensions) in a datawarehouse may be viewed as a special case of the read skew anomaly that may occur when different transactions access and update records without concurrency control. That is, we prove that aggregating fact data to the levels of a dynamic dimension should not make sense. On the other hand, we will also illustrate, by examples, that in some situations it does make sense that fact data is aggregated to the levels of a dynamic dimen-
sion. That is, it is the semantics of the data that determine whether historical dimension data should be preserved or destroyed. Even worse, we also illustrate that for some applications, we need a history preserving response, while for other applications at the same time need a history destroying response. Kimball et al., (2002), have described three classic solutions/responses to handling the aggregation problems caused by slowly changing dimensions. In this chapter, we will describe and evaluate four more responses of which one are new. This is important because all the responses have very different properties, and it is not possible to select a best solution without knowing the semantics of the data.

INTRODUCTION

A data warehouse is an OLAP (On Line Analytical Processing) database (Codd, 1993 and Codd et al. 1993), where the data is loaded/updated periodically. In other words, a data warehouse is not an OLTP (On Line Transaction Processing) database (e.g., Gray and Reuter, 1993). The data warehouse drill functions (e.g., Kimball et al., 1998) have been developed to accommodate the special needs for aggregating the data stored in the fact table of a data warehouse.

The traditional drill-down functions use the one-to-many relationships of the data warehouse to find more detailed information. If we take accumulated data as an example, the drill-down function will show the more detailed data elements of the accumulated data. The roll-up function can use the one-to-many relationships of the data warehouse to generate an overview of the detailed information. However, the aggregating drill functions may give misleading results as old fact data may be aggregated to dimension levels that have changed since the fact data were created. In this paper, we will evaluate the following 7 different techniques for handling the aggregation problems of slowly changing dimensions (Responses to SCD). The first three responses are the classic techniques described by Kimball et al. (1998).

- **Type 1 response:** Overwrite the dimension record with the new values by which historical information is lost.
  - **Type 2 response:** Create a new additional dimension record with the current information.
  - **Type 3 response:** Create a “Previous” field in the dimension record to store the immediate previous attribute value.
  - **Type 4 response:** Create a new dimension using a dynamic dimension hierarchy
  - **Type 5 response:** Store dynamic dimension data as fact data
  - **Type 6 response:** Use finer granularity in combination with response 1 or 3
  - **Type 7 response:** Store the dynamic dimension data as static facts in another data mart

Table 1 gives an overview of the most important properties of the different responses, and in section 4 the evaluation is described in more details. The most important evaluation criterion for responses to Slowly Changing Dimensions is whether the responses preserve historical information. In the next section, we will illustrate by examples that it is the semantics of the data that determine whether historical information should be preserved. That is, both history preserving and history destroying responses are important. Therefore, it is important to know that only the Type 1 and 3 responses can overwrite historical information in an effective way. The Type 3 and 6 responses do not preserve all the details of historical information but may be a good choice in special situations. Response Types 2, 4, 5, and

---

**Analyses and Evaluation of Responses to Slowly Changing Dimensions in Data Warehouses**
12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product’s webpage: www.igi-global.com/chapter/analyses-evaluation-responses-slowly-changing/28174?camid=4v1


Related Content

Discovering Surprising Instances of Simpson's Paradox in Hierarchical Multidimensional Data
www.igi-global.com/chapter/discovering-surprising-instances-simpson-paradox/7831?camid=4v1a

Semantic Data Mining
www.igi-global.com/chapter/semantic-data-mining/7847?camid=4v1a

Text Mining by Pseudo-Natural Language Understanding
www.igi-global.com/chapter/text-mining-pseudo-natural-language/11085?camid=4v1a

Text Categorization
www.igi-global.com/chapter/text-categorization/11084?camid=4v1a