

Analysis Framework for Reduced Data Warehouse

Franck Ravat*, Jiefu Song*, Olivier Teste**

* IRIT - Université Toulouse I Capitole, 2 Rue du Doyen Gabriel Marty
F-31042 Toulouse Cedex 09

** IRIT - Université. Toulouse II Jean Jaurès, 1 Place Georges Brassens
F-31703 Blagnac Cedex

*** {ravat, song, teste}@irit.fr

Abstract. Our aim is to define a framework supporting analysis in MDW with reductions. Firstly, we describe a modeling solution for reduced MDW. A schema of reduced MDW is composed of states. Each state is defined as a star schema composed of one fact and its related dimensions valid for a certain period of time. Secondly, we present a multi-state analysis framework. Extensions of classical drilldown and rollup operators are defined to support multi-states analyses. Finally we present a prototype of our framework aiming to prove the feasibility of concept. By implementing our extended operators, the prototype automatically generates appropriate SQL queries over metadata and reduced data.

1 Introduction

Nowadays, Multidimensional Data Warehouse (MDW) is a widely used component in decision support systems. A MDW schema is based on facts (analysis subjects) and dimensions (analysis axes). The facts contain analysis indicators while the dimensions organize analysis parameters according to their level on hierarchies: from the minimal (most detailed) granularity to the maximal (most general) granularity.

In a MDW, data are stored permanently and new data is steadily added. As result, a MDW stores a huge volume of data in which the analyst may be lost during her/his analyses. On the other hand, the relevance of MDW data decreases with age: detailed information is generally considered essential for recent data Skyt et al. (2008), while more aggregated information can usually satisfy the need of analysis over older data. For instance, an analyst may have interest in analyzing published news by subthemes for the last four years. However, as most of today's subthemes did not exist before, the subtheme granularity level may be proved useless for an older period. As a result the analyst may have no more interest in analyzing published news by subtheme over the last ten years but by a higher and more stable granularity level, such as news' theme.

Facing large volumes of data among which a great amount of inadequate data are found, our aim is to both increase the efficiency of analysis and facilitate the analysts' task. To this end, we