Abstract

Data warehouse systems aim to support decision making by providing users with the appropriate information at the right time. This task is particularly challenging in business contexts where large amount of data is produced at a high speed. To this end, data warehouses have been equipped with Online Analytical Processing tools that help users to make fast and precise decisions through the execution of complex queries. Since the computation of these queries is time consuming, data warehouses precompute a set of materialized views answering to the workload queries.

This thesis work defines a process to determine the minimal set of workload queries and the set of views to materialize. The set of queries is represented by an optimized lattice structure used to select the views to be materialized according to the processing time costs and the view storage space. The minimal set of required Online Analytical Pro- cessing queries is computed by analyzing the data model defined with the visual language CoDe (Complexity Design). The latter allows to conceptually organize the visualization of data reports and to generate visualizations of data obtained from data-mart queries. CoDe adopts a hybrid modeling process combining two main methodologies: user-driven and data-

driven. The first aims to create a model according to the user knowledge, requirements, and analysis needs, whilst the latter has in charge to concretize data and their relationships in the model through Online Analytical Processing queries.

Since the materialized views change over time, we also propose a dynamic process that allows users to (i) upgrade the CoDe model with a context-aware editor, (ii) build an optimized lattice structure able to minimize the effort to recalculate it, and (iii) propose the new set of views to materialize. Moreover, the process applies a Markov strategy to predict whether the views need to be recalculate or not according to the changes of the model. The effectiveness of the proposed techniques has been evaluated on a real- world data warehouse. The results revealed that the Markov strategy gives a better set of solutions in term of storage space and total processing cost.