GENERATING DIMENSIONAL DATA MARTS FROM META-DATA:

A PROOF-OF-CONCEPT PROTOTYPE

by

Abdellatif Lghali

A THESIS

Submitted to the Faculty of Sciences of the VU University
in partial fulfillment of the requirements for the degree of

MASTER OF BUSINESS ANALYTICS

Abdellatif Lghali, Candidate

ADVISORY COMMITTEE

Hans van Leijen, Supervisor

Wan Fokkink, Reader

Peter Boncz , Second reader

VRIJE UNIVERSITEIT AMSTERDAM

De Boelelaan 1105
Amsterdam, 1081 HV
2014

Throughout my life some people have always been there during those difficult and trying time. I lovingly dedicate this thesis and everything I do to Sidi HAMZA al Qadiri al Boutchich and My Mother. For their constant love, encouragement and support .

Preface

The purpose of this report is to fulfill the internship requirement for the Master of Business Analytics degree at the VU University of Amsterdam. While I was employed at the Group IT Business Application department at TomTom as an intern within the Business Intelligence team, I focused on the requirement specification and the design of a Data Mart Code generation tool. This tool will assist end-users to reduce the effort needed to develop SQL code to transform data from Warehouse to Data Marts.

Glossary:

Aggregation: One way of speeding up query performance. Facts are summed up for selected dimensions from the original fact table. The resulting aggregate table will have fewer rows, thus making queries that can use them go faster.

Attribute: Attributes represent a single type of information in a dimension. For example, year is an attribute in the time dimension.

Conformed Dimension: A dimension that has exactly the same meaning and content when being referred to from different fact tables.

Data Mart: Data marts have the same definition as the data warehouse (see below), but data marts have a more limited audience and/or data content.

Data Warehouse: A warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process (as defined by Bill Inmon).

Data Warehousing: The process of designing, building, and maintaining a data warehouse system.

Dimension: The same category of information. For example, year, month, day, and week are all part of the Time Dimension.

Dimensional Model: A type of data modeling suited for data warehousing. In a dimensional model, there are two types of tables: dimensional tables and fact tables. Dimensional table records information on each dimension, and fact table records all the "fact", or measures.

Dimensional Table: Dimension tables store records related to this particular dimension. No facts are stored in a dimensional table.

Drill Across: Data analysis across dimensions.

Drill Down: Data analysis to a child attribute.

Drill Through: Data analysis that goes from an OLAP cube into the relational database.

Drill Up: Data analysis to a parent attribute. ETL: Stands for Extraction, Transformation, and Loading. The movement of data from one area to another.

Fact Table: A type of table in the dimensional model. A fact table typically includes two types of columns: fact columns and foreign keys to the dimensions.

Grain: The grain of a fact table represents the most atomic level by which the facts may be defined. **Hierarchy:** A hierarchy defines the navigating path for drilling up and drilling down. All attributes in a hierarchy belong to the same dimension.

Meta-data: Data about data. For example, the number of tables in the database is a type of metadata.

Metric: A measured value. For example, "Total Sales" is a metric.

MOLAP: Multidimensional OLAP. MOLAP systems store data in the multidimensional cubes.

OLAP: On-Line Analytical Processing. OLAP should be designed to provide end users a quick way of slicing and dicing the data.

ROLAP: Relational OLAP. ROLAP systems store data in the relational database.

Snowflake Schema: A common form of dimensional model. In a snowflake schema, different hierarchies in a dimension can be extended into their own dimensional tables. Therefore, a dimension can have more than a single dimension table.

Star Schema: A common form of dimensional model. In a star schema, each dimension is represented by a single dimension table.