Chapter 6
Built-In Indicators to Support Business Intelligence in OLAP Databases

Jérôme Cubillé
EDF R&D, France

Christian Derquenne
EDF R&D, France

Sabine Goutier
EDF R&D, France

Françoise Guisnel
EDF R&D, France

Henri Klajnmic
EDF R&D, France

Véronique Cariou
ENITIAA, France

ABSTRACT
This chapter is in the scope of static and dynamic discovery-driven explorations of a data cube. It presents different methods to facilitate the whole process of data exploration. Each kind of analysis (static or dynamic) is developed for either a count measure or a quantitative measure. Both are based on the calculation, on the fly, of specific statistical built-in indicators. Firstly, a global methodology is proposed to help a dynamic discovery-driven exploration. It aims at identifying the most relevant dimensions to expand. A built-in rank on dimensions is restituted interactively, at each step of the process. Secondly, to help a static discovery-driven exploration, generalized statistical criteria are detailed to detect and highlight interesting cells within a cube slice. The cell’s degree of interest is determined by the calculation of either test-value or Chi-Square contribution. Their display is done by a color-coding system. A proof of concept implementation on the ORACLE 10g system is described at the end of the chapter.

DOI: 10.4018/978-1-60566-748-5.ch006
INTRODUCTION

Context

Most of large companies have identified the importance and the strategic value of the information contained in their Data Warehouses. But these Data Warehouses are of interest only if the knowledge they contain, is correctly extracted, formatted, summarized, presented and shared by business analysts in order to create added value. Different technologies exist that can organize, present and make data available, from simple request or visualization tools to simulation applications or multidimensional analysis.

In particular, On-Line Analytical Processing (OLAP) applications are powerful decision support tools. They offer a multidimensional view of the data, by calculating and displaying indicators either in detail or with some level of aggregation. Given the fact table of a cube, the multidimensional model consists in dimensional attributes (dimensions) and measures attributes (measures) defined using an aggregate function like sum, count or average.

In practice, a data cube is built from a single large table which contains the most detailed data and we assume that the structure of this underlying table (called DETAIL) is the following: DETAIL(#ID,

\[ d_1, d_2, \ldots, d_z, M_1, M_2, \ldots, M_q, W_1, W_2, \ldots, W_l \]

where \( d_1, d_2, \ldots, d_z \) are called dimension attributes and form the dimensions of the data cube, \( M_1, M_2, \ldots, M_q \) are called measure attributes and form the measures, and \( W_1, W_2, \ldots, W_l \) are a set of weights of each tuple. Usually one of them equals 1 and is used for the count measure. Tuples of this detailed table are called statistical units.

The measures functionally depend on the set of dimensions they are associated with. In this chapter, we consider a sales data cube example in order to analyze customers energy consumption per year and market segmentation according to the following relation scheme:

\[ \text{CUSTOMER(#Cust, Tariff, Dwelling, Heating, Occupation, Water Heating, Oldness Of House, Date Of Contract Subscription, Type Of Dwelling, Consumption) } \]

Figure 1 presents, in a simplified data cube, the mean electrical consumption (per year) and the number of customers according to three characteristics.

In the underlined cell on Figure 1, 3 133 corresponds to the number of customers with Tariff 2 subscribed between 1985 and 1995 and using fuel oil for their energy heating. These customers have consumed 29 560 930 KWh per year.

In many company departments, such as Marketing, Sales or Human Resources, business analysts take advantage of OLAP products for