An Approach for Alert Raising in Real-Time Data Warehouses

Maximiliano Ariel López*, Sergi Nadal**,
Mahfoud Djedaini***, Patrick Marcel***, Verónika Peralta***, Pedro Furtado****

*École Centrale Paris, maxilopez@economicas.uba.ar,
**Universitat Politècnica de Catalunya, snadal@essi.upc.edu
***Université de Tours, firstname.lastname@univ-tours.fr
****University of Coimbra, pnf@dei.uc.pt

Abstract. This work proposes an approach for alert raising within a real-time data warehouse environment. It is based on the calculation of confidence intervals for measures from historical facts. As new facts arrive to the data warehouse on a real-time basis, they are systematically compared with their appropriate confidence intervals and alerts are raised when anomalies are detected. The interest of this approach is illustrated using the particular real world use case of technical analysis of stock data.

1 Introduction

Traditional data warehousing architectures assume offline periods in order to run the costly ETL processes that move and transform data coming from operational sources. Currently, many organisations have the requirement of analysing their information in a real-time manner. For instance, in the stock markets domain, technical analysis allows modelling market behaviour and predicting tendencies. Monitoring markets and quickly detecting deviations from the expected behaviour allow analysts to face abrupt changes. Other domains where real-time analysis is desirable include energy production, traffic and network monitoring.

In most cases, the aforementioned requirement cannot be satisfied with the classic data warehouse and ETL architectures. To enable near real-time analysis based on the most recent information, data warehouse architectures have been extended or adapted (Ferreira and Furtado, 2013; Ferreira et al., 2013; Waas et al., 2013; Jörg and Dessloch, 2009; Santos and Bernardino, 2008). In such systems, how data is loaded and the frequency in which this process is executed change. In (Ferreira and Furtado, 2013), an integrated architecture is proposed, which implements a real-time data warehouse without data duplication. It is composed of three main components: the Dynamic Data Warehouse (D-DW), the Static Data Warehouse (S-DW) and the Merger, with additional components providing real-time ETL. The main idea is to load new data into the D-DW, an in-memory database that holds the most recent information and provides fast integration. On the other side, the largest volume of historical data is stored in the S-DW. Integration between D-DW and S-DW is done through classical offline procedures. After this integration is done, the D-DW is emptied. Queries are handled by the Merger com-