

Open issues in Big Data Warehouse design

Sandro Bimonte*

*TSCF, Irstea

9 Avenue Blaise Pascal CS20085, 63178 Aubière, France
sandro.bimonte@irstea.fr

Abstract. Data Warehouse and OLAP systems allow analyzing huge volumes of data represented according to the multidimensional model. In the era of Big Data, NoSQL systems have been proved to be an effective Business Intelligence solution. Some works recently study warehousing and OLAPing Big Data. (Un)Lucky these works exclusively investigate time performance related to the Volume and Velocity features of Big Data. Therefore, in this paper we investigate the impact of other Big Data features: Variety, Veracity and Value on warehousing and OLAP analysis. Then, we go beyond computation performance and we highlight new Big Data Warehouses design issues.

1 Introduction

Data Warehouses (DWs) and OLAP systems allow analyzing huge volumes of data represented according to the multidimensional model, which defines the concept of dimension (the analysis axes) and fact (the analysis subject) (Kimball, 1996). OLAP relational and multidimensional architectures have been widely studied in the last 30 years (Kimball, 1996).

Conceptual, logical and physical design issues have been extensively investigated by academic and industrial communities (Malinowski and Zimányi, 2006), (Kimball, 1996). Nowadays, DWs and OLAP systems have reached a great maturity for the analysis of Small Data (Miller, 2010). They have been successfully applied in several domains such as marketing, health, agriculture, etc.

However, with the advent of Big Data (Davis, 2012)(Media, 2014) (new) analytical possibilities are offered to decision makers for (new) application domains. In the era of Big Data NoSQL systems have been proved to be an effective Business Intelligence solution (Chen et al., 2012). Different types of NoSQL systems exist: Key value, Extensible record, and Document, Graph (Bugiotti et al., 2014) (Stonebraker et al., 2007), (Floratou et al., 2012). A key value database is a collection of data without a schema and organized as a collection of key value pairs. Data is accessed using the key and its value represents data. Extensible record databases represent data with tables where each row can present different attributes (different columns). Document databases store information as documents having a complex structure. In particular, some works recently study warehousing and OLAPing data using NoSQL systems, since they allow scaling in time and space (e.g. (Dehdouh et al., 2014a) (Dehdouh et al., 2014b) (Chevalier et al., 2015a) (Chevalier et al., 2015b)). Although these works show the feasibility