Sharing-based Privacy and Availability of Cloud Data Warehouses

Varunya Attasena, Nouria Harbi and Jérôme Darmont

Université de Lyon (ERIC Lyon 2)
varunya.attasena@eric.univ-lyon2.fr
nouria.harbi@univ-lyon2.fr
jerome.darmont@univ-lyon2.fr

Abstract. Cloud computing can help reduce costs, increase business agility and deploy applications with a high return on investment such as data warehouses. However, storing and managing data in the cloud may not be fully trustworthy. In this article, we focus on both data security (data privacy, availability and integrity) and data analysis in the cloud. To solve the data security issue, we propose a new (m, n, t) multi secret sharing scheme based on block cryptography, secret sharing and hash functions. Moreover, we apply this solution onto a cloud data warehouse such that data security and data analysis are addressed. An extensive security and performance analysis shows that the proposed schemes can prevent most attacks, guarantee data availability and integrity, and allow analyzing data at low costs (data storage, data transfer and time computation) in the pay-as-you-go economic model in the cloud.

1 Introduction

Business intelligence (BI) and data analytics have been an ever-growing trend in many business (e.g., finance, telecoms, insurance, logistics. . .) and non-business (e.g., agriculture, medicine, health and environment. . .) domains for more than twenty years. The more recent advent of cloud computing now theoretically allows to deploy data analytics more easily. Data that would have earlier been too costly to process in time, money or human resources can be analyzed efficiently and at lower costs. Building a traditional BI system indeed typically necessitates an important initial investment. By contrast, with the cloud pay-as-you-go model, users punctually devote small amounts of resources in return for a one-time advantage. This trend is currently supported by numerous "BI as a service" offerings by both cloud start-ups and major BI industry vendors, with high economic stakes.

Moreover, the elasticity characteristic of cloud computing, i.e., the dynamic on-demand provisioning of resources, does not only help scale performance up or down, but also enables to dynamically bring in new data sources to meet emerging needs for new analyses. Thus, data analytics is likely to be increasingly demanded by independent actors grouping together to achieve a temporary common goal through a collaborative community effort. For instance, open data, which are easily accessed from the Web, are in high demand. They could be integrated to private data and cross-analyzed with intelligible on-line tools featuring advanced