Chapter 26
An Encryption Methodology for Enabling the Use of Data Warehouses on the Cloud

Claudivan Cruz Lopes
Federal Institute of Education, Science and Technology of Paraíba, Patos, Brazil

Valéria Cesário-Times
Federal University of Pernambuco, Recife, Brazil

Stan Matwin
Dalhousie University, Nova Scotia, Canada & Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

Cristina Dutra de Aguiar Ciferri
University of São Paulo, São Paulo, Brazil

Ricardo Rodrigues Ciferri
Federal University of São Carlos, São Carlos, Brazil

ABSTRACT

A cloud data warehouse (cloud DW) is a subject-oriented, integrated, time-variant, voluminous, non-volatile and multidimensional distributed database that is hosted in a cloud. A solution to ensure data confidentiality for a cloud DW is cryptography. In this article, the authors propose an encryption methodology for a cloud DW stored according to the star schema, considering both the data confidentiality maintenance of the DW and the capability of processing analytical queries directly over the encrypted DW. The proposed encryption methodology comprises an encryption strategy for DW called MV-HO (MultiValued and HOmomorphic) for the definition of how the different types of DW's attributes must be encrypted. The proposed MV-HO encryption strategy was compared with encryption strategies based on symmetric encryption, order preserving symmetric encryption and homomorphic encryption. Results indicated that MV-HO is the best solution found, as MV-HO is pareto-optimal with respect to other strategies investigated.

DOI: 10.4018/978-1-7998-7705-9.ch026

INTRODUCTION

Cloud computing provides Database as a Service (DAS), where data management is outsourced to a cloud provider. This allows customers to create, maintain and query their data in the cloud using Internet connection. The storage of sensitive data in databases hosted in a cloud has made data security an essential issue for organizations. However, traditional security mechanisms of currently DBMS, which are mainly based on authentication and authorization, have become insufficient. Data confidentiality also may be affected if the data are stored in their original form, which can be read, interpreted and analyzed (Shmueli et al., 2005). Similarly, data confidentiality may be affected if the data is transmitted in their original form between the client and the cloud provider using the Internet.

A solution to ensure data confidentiality is cryptography (Vimercati et al., 2010), i.e., sensitive data are stored in an encrypted form, and even if an adversary gets access to the data, he will be unable to interpret them. However, performing queries over encrypted data requires them to be decrypted, which can pose a safety hazard if the decryption is performed in an untrusted server, such as in a DAS provider (Suciu, 2012). Also, this may lead to a high cost if all encrypted data is transferred to the client, where they are decrypted and the query is executed. Therefore, the use of encryption in databases requires a cost-benefit analysis between the guarantee of data confidentiality and the impact of encryption on query processing performance (Santos et al., 2011).

In recent years, several studies proposed the use of encryption schemes that allow the execution of operations directly on encrypted data (Liu & Wang, 2012, 2013; Popa et al., 2012; Kadhem et al., 2010, 2013; Liu, 2014; Tu et al., 2013), with the objective of reducing the overhead caused by encryption in query processing performance as well as maintaining data confidentiality. Also, an analysis of what database operations can be executed over encrypted data can be found in (Fuller et al., 2017). However, according to the literature review, only few studies address the encryption of Data Warehouses (DWs) hosted in a cloud (Lopes et al., 2014; Lopes & Times, 2015; Guermazi et al., 2015; Attasena et al., 2015).

A DW is a multidimensional database with a high redundancy degree of values and consequently, if the DW encryption is based on fixed encrypted values, a high degree of redundancy of encrypted values will be produced. This redundancy of encrypted values implies a vulnerability that can be exploited by attacks because an adversary can apply statistical measures on the encrypted values to try to infer the original values. Thus, minimizing data redundancy in an encrypted DW improves its protection against statistical attacks, by contributing to the data confidentiality. However, the literature has paid little attention to the effects of data encryption in the performance of analytical queries over non-redundant encrypted DWs (Lopes et al., 2014; Lopes & Times, 2015).

Analytical queries over the logical schema of a DW, such as the star schema, deals with the operations of projection, selection, join, aggregation, sorting and grouping of data, where the selection requires the computation of range constraints using relational operators such as $=$, $>$, $<$, $\geq$, $\leq$ and $\neq$, the join operation performs a natural join among the dimension tables and the fact table, aggregation is usually based on the sum aggregate function, and the sorting and grouping are performed over the projection values. Thus, the processing of analytic queries over an encrypted DW necessarily implies computing such operations over the encrypted data.

This paper focuses on how to process range constraints, equality constraints, data groupings and sorting operations over a DW hosted in a cloud. For this purpose, the proposed work uses multivalued encrypted values to minimize data redundancy. The motivation for addressing this problem is illustrated by Example 1.