Personalized Online Analytical Processing in Big Data Context Using User Profile and Search Context

Menaceur Sadek, Laboratory of Mathematics, Informatics and Systems (LAMIS) University of Larbi Tebessi, Tebessa, Algeria
Makhlouf Derdour, Computer Sciences Department University of Larbi Tebessi, Tebessa, Algeria
Bouramoul Abdelkrim, MISC Lab & Fundamental Computer Science and its Applications Department Constantine2 University, Constantine, Algeria

ABSTRACT

This article is part of the field of analysis and personalization of large data sets (Big Data). This aspect of analysis and customization has become a major issue that has generated a lot of questions in recent years. Indeed, it is difficult for inexperienced or casual users to extract relevant information in a Big Data context, for volume, the velocity and the variability of data make it difficult for the user to capture, manage and process data by methods and traditional tools. In this article, the authors propose a new approach for personalizing OLAP analysis in a Big Data context by using context and user profile. The proposed approach is based on five complementary layers namely: Extern layer, layer for the formulation of the contexts defined in the system, profiling and querying layer and layer for the construction of personalized OLAP cubes and a final one for multidimensional analysis cubes. The conducted experiment has shown that taking context and user profile into account improves the results of online analytical processing in the context of Big Data.

KEYWORDS

Big Data, OLAP, Personalization, User Profile, User Query

INTRODUCTION

Big Data (Salloum, 2016) is one of the most sought-after research areas in the world. This term refers to datasets whose volume is more than 500 terabytes of data per day. In literature, the term Big Data has been characterized as having one or more of the four dimensions: Volume, variety, velocity, and veracity. These dimensions make it difficult to capture, manage, process and analyze two million records per day, which is considered a big challenge to Online Analytical Processing (OLAP). This analytical processing provides better support for decision making when it comes to data warehouses. Users in this case typically perform an OLAP request sequence called an OLAP session to analyze the results.

The processing of OLAP data cubes in a Big Data context is not efficient enough because of two characteristics that are related to the Big Data repositories: (i) the size; which becomes really explosive, (ii) the complexity; which can be very high in such data sets. As a result, strong demand is emerging for the design of new models, techniques, algorithms, and compute platforms to support the problem of computing OLAP data cubes on Big Data.

DOI: 10.4018/IJSITA.2017100106

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
The size problem mentioned above, which arises in the traditional OLAP cube calculation, could possibly be overcome by the reduction of large data sets using personalization techniques. The latter consists in involving the context of the user in the storage process to avoid all sorts of unnecessary or noisy information and thus make the users’ performances very strong on the creation or exploitation of OLAP data cubes.

In this article, we propose a new architecture dedicated to the creation of personalized OLAP cubes according to the context and the profile of the user. The contributions of this document can be summarized as follows: (1) Capture the contexts necessary for the categorization of the user; (2) Capture the context of the initial query; (3) Use the user context through the personalization module to generate new queries from the initial query and (4) Exploit new queries in Hadoop, Hive and Apache Kylin for creating and storing OLAP cube in HBase (Hadoop Base).

The rest of this paper is organized as follows: Section 2 provides the motivations of research and related works. Section 3 describes online analysis and massive data personalization. Section 4 presents the general architecture of our system and the different layers. Section 5 presents the implementation and experiments of the system. Finally, Section 6 gives the conclusion and future work to be done.

MOTIVATION AND RELATED WORK

Motivation

Today, OLAP analysis from big data is entering a new phase called personalized data. Modern Big Data systems, such as Hadoop are used as an alternative for data warehousing (Hollingsworth, 2012), (Kuldeep et al., 2014). It collects and stores inherently complex data streams due to volume, velocity, value, variety, variability and veracity (Russom, 2011). According to these reasons, the traditional OLAP implementation, namely, the RDBMS based ROLAP system, seems to be inadequate, because the new massive data architectures and analytics tools go beyond SQL Datawarehouses and OLAP engines (Jie et al., 2015). On the other hand, the work (Cuzzocrea et al., 2013) highlights the raised problems and research tendencies in the field of data warehousing and OLAP analysis on Big Data. On the basis of this observation, we seek through this study to design a new architecture to solve problems when calculating OLAP data cubes.

Related Work

OLAP systems have become among the promising solutions to improve the decision-making process. They are unsuited to the needs and contexts of analysis of the decision makers for the diversification of the peculiarities of the users and the enormous increase of volume of data. As a result, research is directed towards the challenges of large data architectures (Jie, 2015), where new techniques have been used for the storage and calculation of OLAP data cubes on Big Data.

The work of (Tian, 2008), who presented an OLAP analysis approach based on the simplified parallel programming model MapReduce (Dean & Ghemawat., 2008). The system’s input is in the form of Short Message Service (SMS), which is operated by the two MapReduce functions that produce very small popular messages with a high send coverage rate and meet the actual requirements.

The authors in (Wang et al., 2014) presented a data cube model for XML documents. They proposed an optimized algorithm implemented by MapReduce on Hadoop and that responds to the growing demand for XML massive data analysis in OLAP.

The paper (Jie et al., 2015) presents the design, implementation and evaluation of HaoLap, which is an OLAP system for Big Data, it is built on Hadoop based models and algorithms well specified. HaoLap has been tested on several large data sets and OLAP applications and compared these performances with other tools such as: Hive, HadoopDB, HBaseLattice and Olap4Cloud.

Recently, the authors of (Supriya et al., 2017), proposed a system primarily intended for the construction of OLAP cubes on the Hadoop ecosystem. The internal architecture of the system
Strategic Overhaul of Government Operations: Situated Action Analysis of Socio-Technical Innovation in the Public Sector
www.igi-global.com/chapter/strategic-overhaul-of-government-operations/208727?camid=4v1a

K-link+: A P2P Semantic Virtual Office for Organizational Knowledge Management
www.igi-global.com/chapter/link-p2p-semantic-virtual-office/36783?camid=4v1a