

Cloud-Based Data Warehousing Application Framework for Modeling Global and Regional Data Management Systems

Thanh Binh Nguyen

International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361
Laxenburg, Austria
nguyenb@iiasa.ac.at
<http://www.iiasa.ac.at>

Abstract. In this paper, a Cloud-based Data Warehousing Application (CDWA) Framework is proposed to handle multi levels in structures of data management systems, i.e. global data warehouse and its data marts on the clouds. First, a Cloud-based Multidimensional Model, i.e. dimensions and theirs related concepts, variables or facts, data cubes, is specified in a very formal manner. Afterwards, to fulfill global and regional specific requirements of data management systems, the Cloud-based Data Warehousing Application Framework and its classes of services are modeled by using UML (Unified Modeling Language) to design a centralized global data warehouse and its local regional data marts. In this context, an implementation example is presented in order to proof of our conceptual framework.

Keywords: Data Warehousing (DWH), OLAP, CDWA Framework, Multidimensional Data Model, UML, ETL.

1 Introduction

Multidimensional data analysis, as supported by OLAP (Online Analytical Processing), requires the computation of many aggregate functions over a large volume of historically collected data [15]. In this context, as a collection of data from multiple sources, integrated into a common repository and extended by summary information, data warehousing (DWH) workloads usually consist of a class of queries typically interleaved with group-by and aggregation OLAP operators [10].

There arise a few data cubing architecture variations which make Web reporting and data analysis deployment scenarios an integral part of workgroup collaboration [15], i.e. cubes can be dynamically generated, using parameters specific to the user when the request is submitted. However, user requirements and constraints frequently change over time and ever-larger volumes of data and data modification, which may create a new level of complexity [16], maintaining multiple copies of the same data across multiple cubes for different kinds of business requirements. In order to achieve a truly open data place, we need

standardized and robust descriptions of data cubing services [16], making easier the processes of configuration, implementation and administration of data cubes in heterogeneous environments with various deployment support.

To solve the above challenge, cloud computing, which is a new computing paradigm to provide reliable, customized and QoS guaranteed dynamic computing environments for end-users [3,6], is considered as an option. Cloud computing delivers infrastructure, platform, and software (application) as services, which are made available as subscription-based services in a pay-as-you-go model to consumers. These services in industry are respectively referred to as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [12]. According to [3], cloud computing, the long-held dream of computing as a utility, has the potential to transform a large part of the IT industry, making software even more attractive as a service.

In this paper, to handle multi levels in structures of data management systems, i.e. global data warehouse and its data marts on the clouds, a cloud-based data warehousing application framework, namely CDWA, is proposed. First a multi-dimensional data model is formulated in a very formal manner in the context of cloud-based data warehousing systems. Afterwards, to fulfill global and regional specific requirements of a class of data management systems, the Cloud-based Data Warehousing Application Framework and its classes of services are modeled by using UML (Unified Modeling Language) to design a centralized global data warehouse and its local regional data marts. In this context, an implementation example is presented in order to proof of our conceptual framework.

The rest of this paper is organized as follows: section 2 introduces some approaches and projects related to our work; in section 3, an introduction of cloud-based multidimensional data models and CDWA framework concepts will be presented. Section 4 will show our implementation results in term of typical case studies. At last, section 5 gives a summary of what have been achieved and future works.

2 Related Work

Our approach is rooted in several areas of cloud intelligence research, including the trends and concepts, the combined use of cloud computing and data warehousing technologies in supporting cloud intelligent systems. With the amount of data generated on the clouds increasing continuously, delivering the right and sufficient amount of information at the right time to the right business users has become more complicated and critical [13,14].

Cloud computing has developed quickly, with companies such as Amazon, Google and Salesforce.com getting ahead of the information technology infrastructure stalwarts such as HP, IBM, Microsoft, Dell, EMC, Sun and Oracle [3,12]. The latter are certainly participating, but doing so more behind the scenes notwithstanding some high profile press releases. More and more enterprise solutions and platforms for Business Intelligence have been developed such as IBM DB2 with Business Intelligence Tools, Microsoft SQL Server, Teradata Warehouse, SAS, iData Analyzer, Oracle, Cognos, Business Objects, etc.[3,4,11,13].