International Journal of Engineering & Technology, 7 (4) (2018) 4425-4428

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 16949
Research paper

Modelling of the joint evolution of analysis needs and data warehouse

Ida Brou Assie [1] *, Paul Louis Seka [1], Jérôme Kablan Adou [1], Souleymane Oumtanaga [2]

[1] *Mathematics-Informatic Department, Felix Houphouet Boigny University Abidjan-Côte d'Ivoire*
[2] *Institut Polytechnique Houphouët Boigny Yamoussoukro-Côte d'Ivoire*
**Corresponding author E-mail: ida_as09@yahoo.fr*

Abstract

The decision support system, a set of data, is specifically organized, appropriate and accessible in decision-making. It stems from the data warehouse taking into account the users in this development. However, data warehouse is a complex technology in decision-making systems. Then, we highlight a new approach to the data warehouse model in this paper about the joint evolution of user and data warehouse needs. This model is based on rules combining R-DW (Data-based Data Warehouse) and MDA (Model Driven Architecture) models. As a result, this new approach makes it possible to integrate the knowledge of the users while favoring a temporal evolution of the possibilities of analysis.

Keywords: *Decision Support System; Evolution Model; Data Warehouse.*

1. Introduction

The decision support system is a collection of data that is specifically organized, appropriate and accessible in decision-making. The fundamental element at the heart of the decision-making architecture remains data warehouse. In addition, the drawing-up of data warehouse (ED) relies on the data sources and the requirements of analysis by users on the other hand.
Moreover, it is always uneasy to identify all the users' analysis needs about the design of the ED storage of the moment. Because, this is certain in decision-making situations require decision-makers in considering one or more new parameters issued in the context and can influence the definitive decision. In this case, the evolution of warehouse model is more driven by knowledge than by data, that is, the user's analysis needs.
With the dazzling evolution of data to Big Data, we notice that all organizations are striving to reduce timeframes in decision-making systems based on users' analysis needs. As data warehousing a complex technology in decision-making systems, a question arises from the system users in this process. Our paper deals with this problem for two main reasons. First, warehousing process involves extracting data sources after study, and then users are consulted implicitly. Thus, data to be stored have subdued a series of cleaning and processing.
On the basis of this, we are vested in co-evolution of the analysis needs of the users and data warehouse. The development of data warehouse imposes on the decision-making system administrator, a centralization of the analysis needs and the data in which they are based on. But, this work remains highly dealt in the literature of the evolution of decision-making systems. For this reason, we centralize a new data warehouse approach that combines Data-based Data Warehouse (D-FWR) models of [4] and MDA (Model Driven Architecture) [6] in this paper. This approach renders it possible considering the knowledge of the users while paving the way for a gradual change of the possible analysis.

2. State of the art

Several works in the literature have dealt with the problem of the evolution of decision-making systems. In this case, the absolute necessities for analysis have been a key element in the evolution of decision-making systems through data warehouse. Lots of works were even done on this subject. We can classify the different aspects of this work according to two areas: (1) the development of data sources about warehouse model and (2) the evolution of data warehouses. A literary review of these works, it should be noted, helps us to know their contributions in order to easily position our work.
The work a part of the evolution of data sources on the warehouse model scrutinizes the impacts of the evolution of sources on warehouse model. In this case, the increase of the sources must be passed on quickly, easily and efficiently to guarantee a continuity of the services carried out by warehouse. The works presented in [6] and [1] come within this frame. Building on the Model Driven Architecture model (MDA), the former one is automated the propagation of evolutions. The implementation of the approach has stemmed from software prototype called DWE implemented under Eclipse. The second is focused on the dynamic adaptation of materialized views following data source changes with the warehouse view recount after each changing data source.
The generalization of these approaches to structurally heterogeneous sources requires the transition to a pivotal data model between the source and the evolution model. This leads to the second wave of work on the evolution of SID about the progression of data warehouse.
Today, the current change within data sources, its content and structure, significantly impacts data warehouse. In [3], authors quote this: "The DW must always contain the latest information to be able to reflect the evolving state of the real world which makes it necessary to properly manage all types of changes and appropriately update the DW."

Faced with this state of facts, the latter has thus presented the state of work about the problems to evolution of the decision-making systems based on their reflections on data warehouses. From their analysis, it appears that the current work is based on three main approaches: evaluative schema, versioning scheme and service of views.
In the same way, in [7], the authors suggest the implementation of outline of data warehouse through the semantics of a complex hierarchy. In that sense, they made their work mainly on the evolution at the level of hierarchies, and defining a set of operations of evolution and constraints which ensures the integrity of the data and development of its storage. These operations and constraints were defined in using uni-level (ULD) level description and multi-level dictionary definition (MDD) approaches. Their work resulted in adequate monitoring of warehouse schema with the applied constraints that were implemented in a SQL Server environment.
This work above mentioned on specific aspects of the evolution of decision-making systems like the evolution of data sources and the warehouse model. The aspects have resulted in applicable solutions to specific aspects of decision-making systems. But, as far as we know, a little work has addressed the issue of changing analytics needs and the data warehouse.

3. Description of the proposed approach

To alleviate the issue of evolving analysis needs and data warehouse, we propose an approach that relies on a combination of R-DW and MDA models for improved modeling of evolution in the SIDs. To achieve this, this combination integrates both the knowledge of the users and a time variation of the possible analysis at the level of the update of multidimensional model. The reasons for such a choice in the following:

- The two-part R-DW model (a fixed part containing a fact table linked to its first-level dimensions and a second evolutionary part defined by means of transformation rules) is used to dynamically create hierarchies of evolution.
- The MDA model allows our approach to separate the functional specifications of the system from the details of its implementation. This model provides an approach for developing computing systems based on models and model transformations according to a set of standards [6].

To sum up, this approach integrates both the users' analysis needs while making possible separation of the functional specifications of the SID from the details of its implementation.

4. Implementation of the proposed approach

To implement of the approach, we start from a case study in the field of study of social resilience. In a context of improving / strengthening social resilience, for example, decision-makers consider the changing context with the emergence of new parameters for the orientation of their policy of investments. However, the decisions taken by the latter come from SIDs previously designed although they know that this or these new parameters can influence the final investment decision.
The integration of these new parameters into the SID ED schema utilizes the R-DW model of (Favre et al, 2007) introducing them as new dimensions in the ED schema, when these are not a hierarchy of an existing dimension. The R-DW model of (Favre et al, 2007) is described by the following triplet:

$$\mathrm{R-DW} = (\mathcal{F}, \mathcal{E}, \mathcal{U}) \quad (1)$$

Where $\mathcal{F}$ is the fixed part (initial Schema of the ED without the evolution), $\mathcal{E}$ the evolutionary part (initial Schema of the ED with the evolution), and $\mathcal{U}$, the universe of dimensions.
Our approach also introduces the evolutionary part of the Favre model, the concept of a new table when the analysis needs do not refer to the existing dimension in the SID ED schema. However, the analysis needs can often generate a level of granularity of existing dimensions or new dimensions. To find a solution for this situation, certain rules were established. From these rules, therefore, we use that approach where the Param variable designates the users' analysis needs:

- Rule R1: If $\mathrm{Param} \in \mathcal{F}$ then "one is interested in Param like level of granularity of the dimensions".
- Rule R2: If$\mathrm{Param} \notin \mathcal{F}$then "we are interested in Param as new dimension" and Param $\in \mathcal{U}$.
- Characteristic of the R-DW model:

a) The fixed part $\mathcal{F}$ is composed of a fact table F and dimensions D represented by:

$$\mathcal{F} = < \mathrm{F}, \mathrm{D} > \quad (2)$$

b) The evolutionary part $\mathcal{E}$ composed of different levels of granularity of dimension and is noted:

$$\mathcal{E} = < \mathrm{R_{i.}}, \mathrm{L_i}^{\mathrm{sk}}.\mathrm{A} > \quad (3)$$

Where $\mathrm{R_{i.}}$is the set of aggregation rules generated by the attributes of granularity levels of dimensions $\mathrm{L_i}^{\mathrm{sk}}.\mathrm{A}$.

- Proposed approach:

Rules R1 and R2 above stated, our approach alters Favre in its evolutionary part in the addition of new dimension, so obviously it does not refer to an existing dimension which then becomes $\mathcal{E}_{\mathrm{Param}}$:

$$\mathcal{E}_{\mathrm{Param}} = < \mathrm{R_{i.}}, \mathrm{L_i}^{\mathrm{sk}}.\mathrm{A} > \cup\ \mathrm{Param} \quad (4)$$

In addition, the implementation of our approach pays attention to the MDA model providing a model-based approach to the development of computer systems, makes it possible to separate the functional specifications of the system from the details of its implementation.

5. Implementation phase of the proposed approach

The implementation of the proposed approach led to the ongoing development of a software prototype called SIDRES in the field of social resilience. This prototype is implemented on Net beans platform and interacts with Microsoft's Microsoft SQL Server decision-making suite.
Based on the MDA model of Taktak [6], our approach changes its overall architecture as a starting point of the progression of the ED as well as the analysis needs. The following figure therefore presents the functional architecture of the proposed approach:

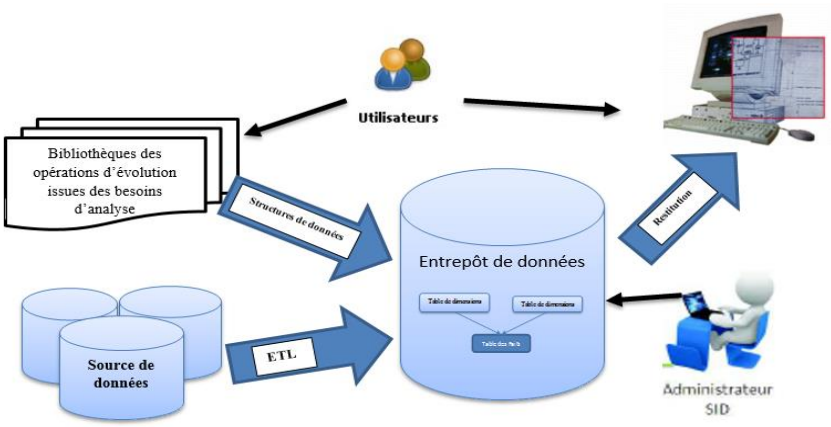
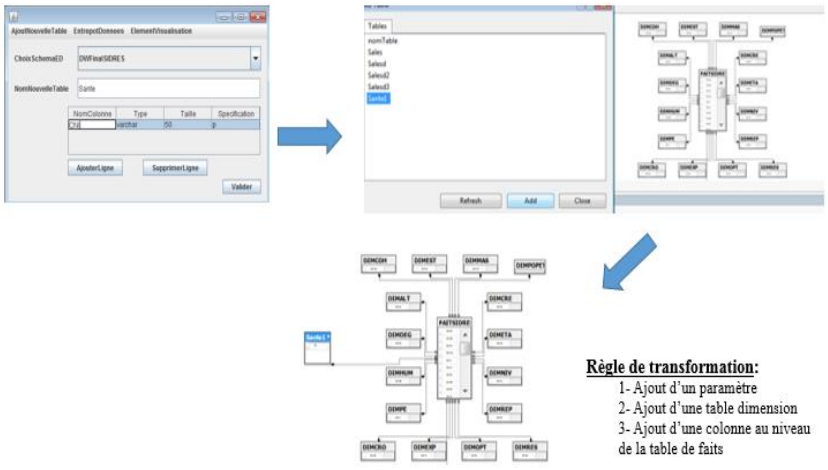


Fig. 1: Functional Architecture of the Proposed Approach.

The following diagram illustrates the ED scheme after considering the analysis of users in resilience studies.

(A) Diagram of the Initial ED.



(B) Diagram of ED in an Evolutionary Context

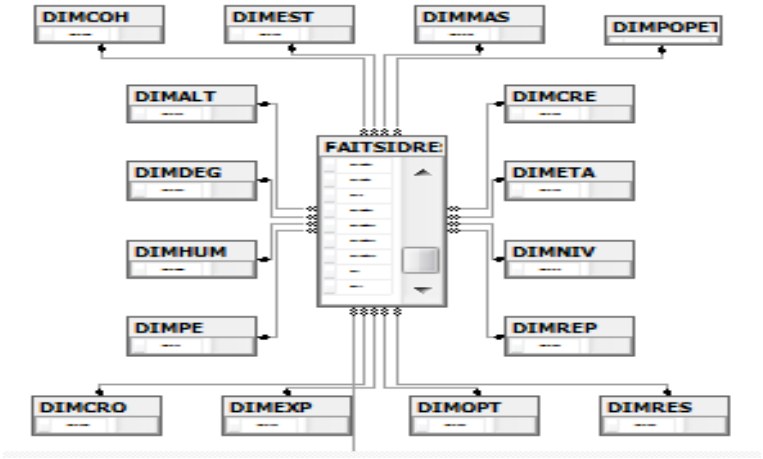


Fig. 2: ED Schema from Our Proposed Approach.

Concerning the approach, we can use the contextual data, we need changing rules that can detect and track the entire decision process in a changing environment. We formalize this by the following algorithm:

Algorithm Decision Process in an Evolving Environment
Data : F, D, $R_{i.}$, Param, $Attr_i$
Output: $\mathcal{E}_{Param}$
Begin
If $Attr_i \in R_{i.}$ then
Param, add as existing D hierarchy
Else if
Param, define as new D
End if
End.

6. Conclusion

In this paper, we have explained the problem of the co-evolution of the analysis needs of the users and the data warehouse on a pro-

posed approach based on a combination of R-DW and MDA models about decision-making systems. We mentioned their respective advantages as the analysis needs of the R-DW model that serve as a guide for the evolution of the ED schema with the MDA model. This approach led to a software prototype called SIDRES in social resilience studies being implemented. It proposes an algorithm which contributes to the decision process in an evolving environment.

The prospects for the future work appear to validate of our approach experimentally to extend all the features of our model.

References

[1] Bellahsene Z. Schema evolution in data warehouses. Knowl. Inf. Syst. 4(3), 283–304, 2002. https://doi.org/10.1007/s101150200008.

[2] Danijela, P. Patrizia, J. Vladan. Data Warehouse Schema Evolution: State of the Art. Central European Conference on Information and Intelligent Systems. 18-25, 2014.

[3] Ghozzi Faiza. Conception et manipulation de bases de données dimensionnelles à contraintes. Thèse enInformatique. Université Paul Sabatier - Toulouse III, 2004.

[4] Favre C., F. Bentayeb, O. Boussaid. Dimension hierarchies updates in data warehouses: a user-driven approach. In Proceedings of the 9th International Conference on Enterprise Information Systems, ICEIS 07, Madeira, Portugal, pp. 206–211.

[5] Jean-Fabrice Lebraty. Les systèmes décisionnels. Akoka, A, Comyn-Wattiau, I. Encyclopédie de l'informatique et des systèmes d'information, Vuibert, pp.1338-1349, 2006.

[6] Taktak Saïd, Feki Jamel, Zurfluh Gilles. Modélisation des transformations pour l'évolution de modèles multidimensionnels.EDA 2014, 5 June 2014-6 June 2014. Vichy, France.

[7] Talwar K., A. Gosain. Implementing schema evolution in data warehouse through complex hierarchy semantics. International Journal of Scientific and Engineering Research, 3, 917-922, 2012.