Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 5, May 2016, pg.794 – 802

Represent Aggregate Knowledge in Data Warehouse and OLAP Systems by Gathering bases with Objects

Yasir ali mmutni Alanbaky[1], Akeel jawad Alameri[2], Waseem saad nsaif[3], Hassan hadi Saleh[4]
[1]Dept. of computer science, basic education college, Diyala University, IRAQ, ee22a12345@gmail.com
[2]Dept. of computer science, basic education college, Diyala University, IRAQ, akeelthm1@yahoo.com
[3]Physical education & sport college, Diyala University, IRAQ, Waceem81@gmail.com
[4]Physical education & sport college, Diyala University, IRAQ, hassan@sport.uodiyala.edu.iq

Abstract: Data warehouse is founded on modeling of multi-dimensional. By using (OLAP) Online Analytical Processing tool, and analyzes multi-dimensional data. In common users must analyzed data at a various overall standards, That why aggregated knowledges must be appropriate represented in the model of multi-dimensional, and planned in depended of physical and logical models. Any ways, existing "conceptual-multi-dimensional-models" weakly represented overall knowledges, that (1) is highly contextual and, (2) has a dynamics structure and complex. To account merits of this knowledge, we proposed to represent the merits of this knowledge with bases in the (PRR) "Production Rule Representation" and objects (UML class diagrams), represent static aggregate knowledges in the class diagram. However bases represented the dynamic, we prepare a typology and associated bases as examples and the class diagrams, We discussed that this aggregation knowledge representation help the requests of user in a data warehouses projects as an early modeling.

Keywords: Data warehouse, Conceptual multi-dimensional model, (OLAP) On-line Analytical Processing, Aggregation, Production base, Unified Modeling Language (UML), (PRR) "Production Rule Representation"

1. Introduction

(DSS) Data-driven decision Support System relies mainly in Data warehouses (March and Hevner 2007). They providing business-oriented view of data for user depend on a multi-dimensional model , which it organize data in hypercube , it is simply represented as a cubes. May then decision makers analyzed and navigate through multi-dimensional data by using

(OLAP) On-Line Analytical Processing tool . changing the granularity of dimension that is relevant it is necessary , to achieve aggregations built from a cube . Analyzing data at different aggregation standards that users needed , which is attained by roll-up operator means (drill-down operator, and its converse operator) . (COUNT, SUM, MAX, MIN , AVG) are aggregation operators. Some aggregations may be not pertinent and the aggregation operator may be different for each dimension standard and measure . For example, if using the AVG operator the aggregation will performed at a given dimension standard , thus letters can't be used at a higher aggregate standard . When the cube including valueless the aggregations have to transact with informations that is uncompleted.

Then, to guarantee flexible and correct aggregation, it should mapped in subsequent and sufficiently represent on "conceptual-multi-dimensional-models" physical and logical model. Aggregate knowledges relate to aggregate functions that could be hierarchies along which they are applicable , as well as to applied , etc.

The major features of a base-based represented we suggested to represent this knowledge using bases , include the following sides : (1) bases have a relatively and formal semantics and precise simple ,(2) it is possible to an axiomatic way to formulated aggregate , (3) some base languages encompass a procedural expressiveness, (4) base languages has an easy maintenance to knowledge , (5) base language is supple to integrate and correspond different kinds of knowledges.

Representations easily considering static information , like relationship and data structure . Particular conceptual model should define to symbolize multi-dimensional data (Abello , Samos and Saltor 2006, Torlone 2003). Beside that , should proposed enriched UML-based model (Abello , Samos and Saltor 2006, Lujan-Mora , Trujillo and Song 2006, Prat , Akoka and Comyn-Wattiau 2006). Although however, introduce dynamic informations like aggregate operation in conceptual standard isn't instantaneous . Therefor it need to linked constraints , high-standard abstract languages enable aggregation representative and chain on aggregation . We debate that the base -based language is the most efficient formalism to achieved like an objective (Prat , Comyn-Wattiau and Akoka 2010). Conceptual that operation might be mapping in next step automatically in the designing processes.

OMG lay the foundations of data warehouses models and concepts through the popular Warehouses "Metamodel" . Therefore it is substantial to indicate to these calibration pains in the meaning of a base languages to represented aggregate knowledges . We followed OMG's recommendation , should based our bases on the PRR (" Production Rules Representation ") formalisms confirming by Object Management Group (OMG) . Thus formalisms are independent tools base languages , high-standard and perfectly merge with UML language (represent static aggregation knowledge we used in our process).

2. Related work

Many researches were in subject of multi-dimensional aggregate in various outlooks. Research devote to aggregations along hierarchy had a lot of benefit over the last conclude. several researches study how (along hierarchies how data may be aggregated), also recognized as compendium ability. Aggregations are achieved using many operators , like MAX, SUM, MIN, AVG, RANGE, COUNT , VAR, etc. . Motivation to learning aggregations over hierarchy is to increasing control role model of multi-dimensional which important in systems to make decision support . (Rafanelli, and Shoshani, 1990) refer to that the concept of epitomizability pointed to the calculation of aggregation of values . It has been observed that epitomizability demand two needful condition : disjointnes and completenes. Disjointnes require in order to categories do not interfere (Rafanelli, and Shoshani, 1990). In other word , completeness cases should be no losted value . These condition is equal to the next assurance.

Many ”conceptual multi-dimensional metamodels” include aggregation knowledge (Abello , Samos and Saltor 2006 , Lujan-Mora , Trujillo and Song 2006, Prat , Akoka and Comyn-Wattiau 2006, Hüsemann,. Lechtenbörger, and. Vossen 2000). They take in consideration the three following cases namely completeness, type compatibility, and disjointness. (Lechtenbörger, and. Vossen 2003) suggest to define three ordinary shape for multi-dimensional scheme ,whose in specific , guarantee epitomize when considering schema problem . However, no guarantee multi-dimensional ordinary form that design method are given to get better of our knowledge, (Hüsemann,. Lechtenbörger, and. Vossen 2000). Depend on the typologies of aggregation functions presents in describes aggregation using 4 limitation standards: standard one , may be all aggregations operator will be used , standard tow, cannot be use SUM operation . Standard three permit on measures COUNT operations only . Standard four, is devoted to measurement that cannot be at all aggregated. define exception in aggregation hierarchies by using intended bases: they modeling aggregation knowledge by explain the applicability of bases, but it not considered as a aggregation functions . other problem is related to the possible chaining of operator. (DeHaan, Toman, and. Weddell 2003) description “cleanly compose aggregate function” that be restricted to some extent , but contain minimal MAX–MAX, SUM–SUM, MIN–MIN, and SUM–COUNT.

Aggregate knowledges are incompetently or poorly represents in existing “conceptual multi-dimensional metamodels” in spite of the assistances of preceding researches . Aggregate knowledges are hard to represented in a simple method . That knowledges (i) in nature is highly contextual ,and (ii) it own a complex framework and dynamic .We propose represented it with objects , in order to consider the feature of aggregate knowledge , (class diagrams in UML and bases (represented language in the Production Base). UML class diagrams represented static aggregation knowledge , put together with PRR bases which represent the dynamic.

3. UML class diagrams for representing static aggregation knowledge

Clearly distinguish between content (instances) and structure (schema) that should be at a "conceptual multi-dimensional metamodel" (Torlone 2003). In our method ,is decisive distinction, ever after will be performed aggregations (roll-ups) at the standard example.

Therefore we distinguish between the data cube "metamodel", and the essence "conceptual multi-dimensional metamodel" (used to represent data warehouse conceptually). User view is the data cube on multi-dimensional data . Result in data cubes and aggregations operate on . We will defined the essence "conceptual multi-dimensional metamodel" and data cubes "metamodel" successively . We using an example in media-planning field to illustrate these "metamodels" ,that will be an example used throughout the paper . We concentrate on past decelerating expeditions, cost, efficiency ,and their duration .

3.1 Essence "conceptual multi-dimensional metamodel"

Figure. 1 (class diagram in UML) is static view of the essence "conceptual multi-dimensional metamodel" represented in thus Figureure . This models emphases in ideas associated with aggregation , and draws on past work (Prat , Akoka and Comyn-Wattiau 2006, Akoka, Comyn-Wattiau, and Prat 2001) . A conceptual multi-dimensional schemas is collection of dimensions and facts. The reality expedition is dimensioned by the dimension of Time, Product ,and Media in our media- pattern example, (an declaration campaign started at a certain date for a product, in a media). Dimensions are collected of hierarchies and facts are collected of measures. Hierarchy is made of sequential warp relationships among dimensions standards. A hierarchies are truly an aggregation way between successive dimensions standards. dimension members are Instances of dimension standards. a combination of n-coordinates of dimensions member are a specific measurement in n-dimensional cubes .

The role of rollup relationship is described by their pluralism . For example, we have asymmetrical hierarchies (Zimanyi , and Malinowski 2006), where the lower pluralism of the source role is (0) , also known as "drill-down uncompleted" hierarchies (Mazón, Lechtenbörger, and Trujillo 2009).

We have an irregular hierarchy if some category is not subdivided into sub-category like the hierarchy Product→Sub--category→Category→All . We have a non-strict hierarchies when the upper pluralism of the target role of a rollup is (*), (Mazón, Lechtenbörger, and Trujillo 2009), may be involved a number definition .

We will have a non-strict hierarchies if considering that the similar product may be classified in numerous category ; we will need to describe a factor to determining how the estimate expedition for a products are share among different category to same product to measuring the cost . We have a "rollup incomplete" hierarchies when the lower pluralism of the aim role is 0.

We considering fact-dimensions relationships are uncommon ("we called dimensioning in our model"), may be it beneficial to representing many situation . Likewise to rollup , may be a "dimensioning" is incomplete (plural) or non_strict . Non_strict dimensioning may need to

definition a factor. We will explain non-strict (plural) dimensioning . At present we suppose that an advertising expeditions which may begin in various media together. at this situation, the dimensioning between Expedition and Media is the collection. The cost assigned to the different media of an advertising campaign based on a coefficient. in a same way , the sales raise is assigned to the various media depend on a coefficients (ascription rule is not should be the same rule for the cost).

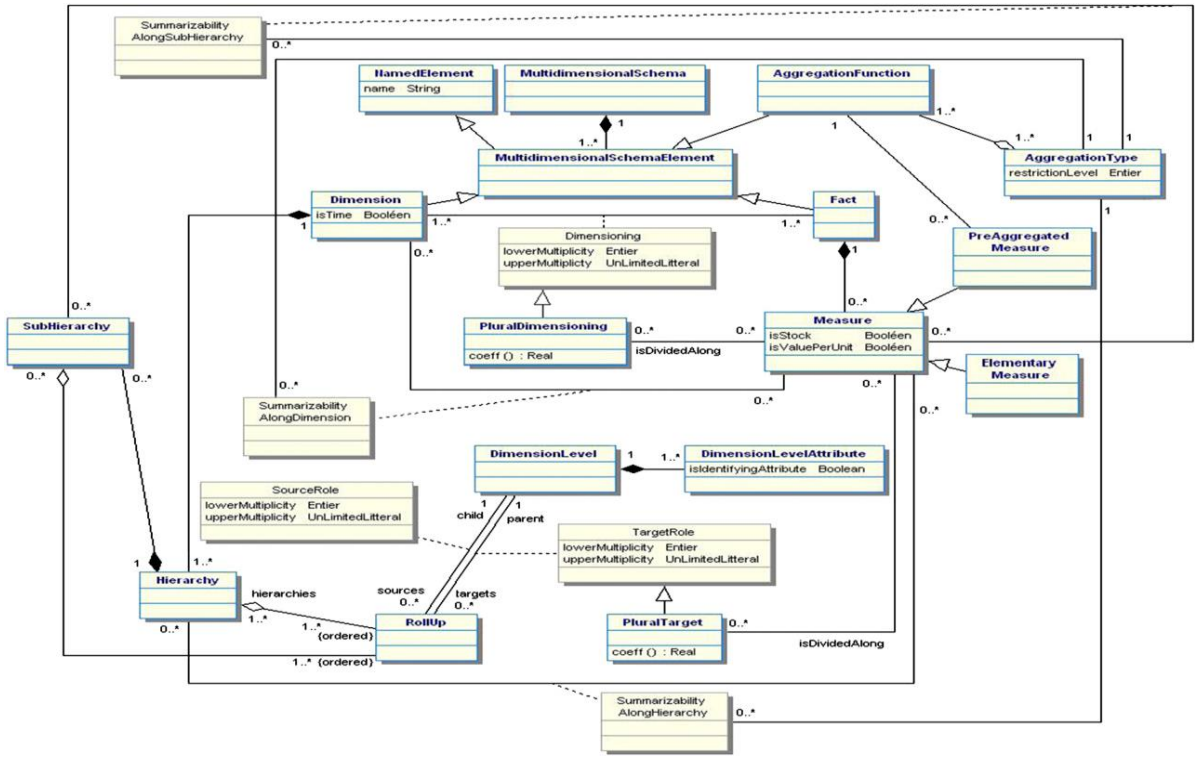

Figure. 1. (static view) Essence "conceptual multi-dimensional metamodel" (MM)

In our "Metamodel" the designer of data warehouse can specify appropriate aggregate task for various measure . In our research, looked at the generality popular aggregate function COUNT, SUM, MAX ,MIN and AVG (that used in SQL).

The applicable aggregate function may be indicated for a measures along a sub hierarchy, along a dimension, or along a specific hierarchies . In same way with preceding worked , we recognize three aggregation kinds. The first aggregation kind (limitation standard 1), whole aggregate functions (COUNT, SUM, MAX ,MIN and AVG) might use . The second kind (limitation standard 2), (COUNT, SUM, MAX ,MIN and AVG) apply. And in limitation standard 3, COUNT function apply only. In same way to (Hüsemann,. Lechtenbörger, and. Vossen 2000), we suggest a separate table to explain applicable aggregate functions.

Note that viable aggregate functions fundamentally static which defined in a multi-dimensional schema . May be , they don't consider the reality that a measures may not be collected after it were averaged.

Following (Lenz, and Shoshani 1997), the essential "conceptual multi-dimensional metamodel" recognize ratios (value-per-units) and stock (a measure is either a flow or a stock) . These merits will have a straight effect on viable aggregate function. We too required to recognized among dimension that are temporal and other dimension . Lastly , we differentiate among pre-aggregated measure and elementary measure . This uniqueness account for the reality that data warehouse oftentimes do not stored data in standard of package , but pre-computed several aggregate.

3.2 "Metamodel" data cube

In Figure. 2 we representing a data cube in static view ("MM" refer to class of essences "Conceptual multi-dimensional meta-model") . The data cubes are collection of axes & cells. Thus axes match to (one-and-only-one) dimensions in the essence "conceptual multi-dimensional metamodel" . (The cell is collected of cells values (every measures have one value) that why a default hierarchy for each axis .

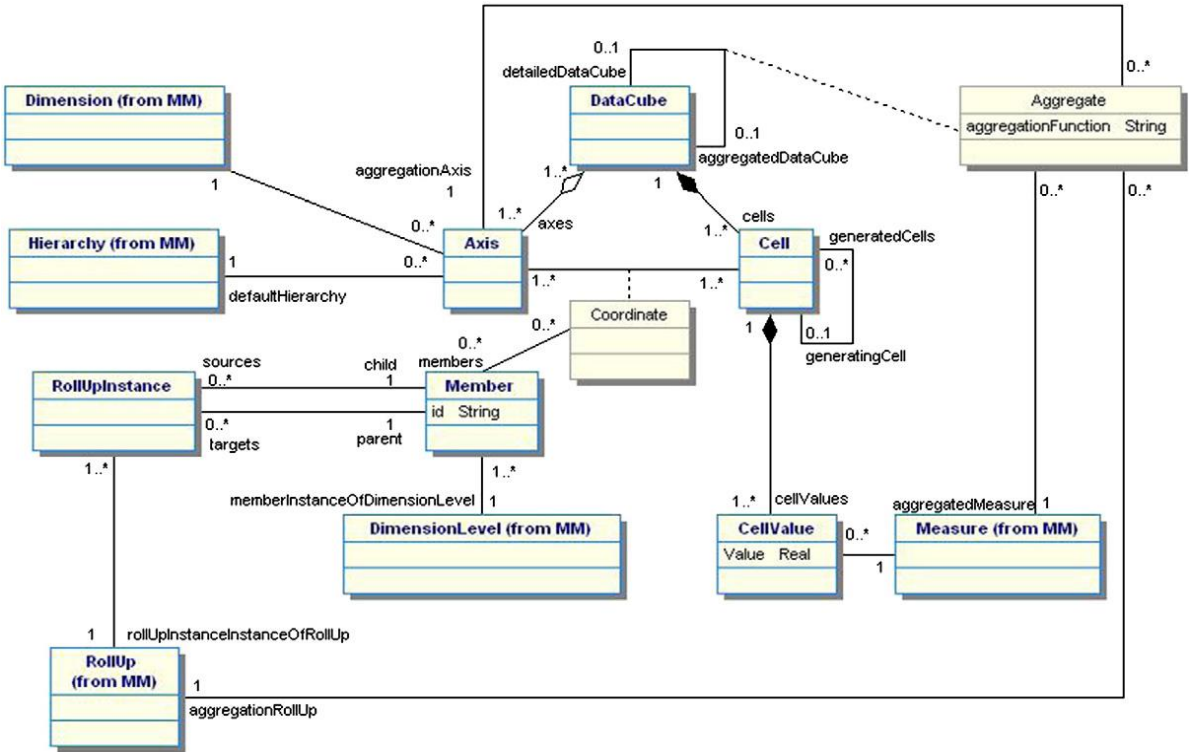

Figure. 2. (static view) Data cube "Metamodel".

Expedition is an example of data cubes these are in our media-planning example. These cubes are tri-dimensional's (one axis for each dimensions Time , Media and Product). Same the beyond debate, this dimension have two , one hierarchies respectively . An cells in the cubes had coordinate over each axis (for example , coordinate "2015-02-01" ,"Channel 2", & "Mini 16") . The defaults hierarchies are ("Product→Sub--category→Category→All") for dimension Product, this we supposed. These three members are instances of these three member is example

of the dimensions standard(Product , Day and Media) respectively. Every cell have the value of each cell measures (for example, the adverting expedition for product Mini 16 started on 2015-02-01, lasted 10 day, in Channel 2 ,cost 100 $, and resulted in a sales rise to 75 unit).

4. PRR bases for representing dynamic aggregation knowledge

We should determine how to fully represent aggregation knowledge , aggregations may be performing in particular data cubes (that mean in what way achieved the aggregate once the aggregate functions that are chooses , and how to choose the aggregations functions) that we explain in before . Ever after this knowledge is highly contextual and complex (count on user priority , data cube …), it is representing with bases appropriately. A base-based approach too ease traced : it enables us to explain how it has been applied and why a specific aggregate functions had chosen in a given situation .

We have chosen the UML to represent the "Conceptual multi-dimensional meta-model", with (PRR) "Production Rule Representation". Representing bases associated with class diagram in UML , enables the "Production Rule Representation" independently of next implementations.

5. Conclusion

Through representation of aggregate knowledge and bases, we find this features:

• taken into consideration , in multi-dimensional diagram the Aggregation is constraints, by purpose of the specific-model bases (if such constraints have been defined).
• Bases enable to represent different kinds of aggregate knowledge .
• Specify aggregation execution knowledge by use aggregation rules .

Acknowledgements

We wish to thanks the unknown reader's for the comments in our research which was helpful to us.

References

* A. Abello, J. Samos, F. Saltor, YAM2: a multi-dimensional conceptual model extending UML, Information Systems 31 (6) (2006) 541–567.
* R. Torlone, "conceptual-multi-dimensional-models", in: M. Rafanelli (Ed.), Multi-dimensional Databases: Problems and Solutions, Idea Group, Hershey, PA, 2003,pp. 69–90.
* S. Lujan-Mora, J. Trujillo, I.-Y. Song, A UML profile for multi-dimensional modeling in data warehouses, Data & Knowledge Engineering 59 (3) (2006)
725–769.
* A. Hevner, S. March, Integrated decision support systems: a data warehousing perspective, Decision Support Systems 43 (3) (2007) 1031–1043.

* N. Prat, I. Comyn-Wattiau, J. Akoka, Representation of aggregation knowledge in OLAP systems, Proc. of ECIS 2010 (18th European Conference on Information
Systems), Pretoria, South Africa, June, 2010.
* N. Prat, J. Akoka, I. Comyn-Wattiau, A UML-based data warehouse design method, Decision Support Systems 42 (3) (2006) 1449–1473.
* M. Rafanelli, A. Shoshani, STORM: a statistical object representation model, Proc. of the 5th International Conference on Scientific and Statistical Database
Management (SSDBM 1990), Charlotte, North Carolina, USA, Springer-Verlag, April 1990, pp. 14–29, LNCS 420.
* H.J. Lenz, A. Shoshani, Summarizability in OLAP and statistical data bases, Proc. of the Ninth International Conference on Scientific and Statistical Database
Management (SSDBM 1997), Olympia, Washington, USA, August 1997, IEEE, 1997, pp. 132–143.
* R. Kimball, M. Ross, The Data Warehouse Toolkit, The Complete Guide to Dimensional Modeling, 2nd Edition, John Wiley and Sons, 2002.
* J. Horner, I.-Y. Song, A taxonomy of inaccurate summaries and their management in OLAP systems, Proc. of the 24th International Conference on Conceptual
Modeling (ER 2005), Klagenfurt, Austria, Springer-Verlag, October 2005, pp. 433–448, LNCS 3716.
* J.-N. Mazón, J. Lechtenbörger, J. Trujillo, Solving summarizability problems in fact-dimension relationships for multi-dimensional models, Proc. of the ACM
11th International Workshop on Data Warehousing and OLAP (DOLAP'08), Napa Valley, California, USA, October 2008, pp. 57–64.
* J. Horner, I.-Y. Song, P. Chen, An analysis of additivity in OLAP systems, Proc. of the 7th ACM International Workshop on Data warehousing and OLAP
(DOLAP'04), Washington, DC, USA, November 2004, pp. 83–91.

* W. Lehner, Modeling large scale OLAP scenarios, Proc. of the 6th International Conference on Extending Database Technology (EDBT '98), Valencia, Spain,
March 1998, LNCS 1377, Springer-Verlag, 1998, pp. 153–167.

* J.-N. Mazón, J. Lechtenbörger, J. Trujillo, A survey on summarizability issues in multi-dimensional modeling, Data & Knowledge Engineering 68 (12) (2009)
1452–1469.
* J. Akoka, I. Comyn-Wattiau, N. Prat, Dimension hierarchies design from UML generalizations and aggregations, Proc. of the 20th International Conference on
Conceptual Modeling (ER 2001), Yokohama, Japan, November 2001, LNCS 2224, Springer-Verlag, 2001, pp. 442–455.
* Object Management Group, “Production Rule Representation” (PRR) version 1.0 specification, document number formal/2009–12–01, http://www.omg.org/
spec/PRR/1.0/PDF/ (visited July 22, 2010), 2009.

* B. Hüsemann, J. Lechtenbörger, G. Vossen, Conceptual data warehouse modeling, Proc. of the Second International Workshop on Design and Management of Data Warehouses (DMDW 2000), Stockholm, Sweden, June 2000 .

* E. Malinowski, E. Zimanyi, Hierarchies in a multi-dimensional model: from conceptual modeling to logical representation, Data & Knowledge Engineering 59
(2) (2006) 348–377.

* D. DeHaan, D. Toman, G. Weddell, Rewriting aggregate queries using description logic, Proc. of the International Workshop on Description Logics (DL 2003),Rome, Italy, September 2003 http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-81/.

* Object Management Group, Common Warehouse Metamodel (CWM) specification, document number formal/2003–03–02, http://www.omg.org/spec/
CWM/1.1/PDF/ (visited January 6, 2010), 2003.

* Object Management Group, Unified Modeling Language (UML) version 2.2. Superstructure specification, document number formal/2009–02–02, http://
www.omg.org/spec/UML/2.2/Superstructure/PDF (visited January 6, 2010), 2009.

