QUALITY METRICS VALIDATION FOR THE COST EFFICIENCY MODELS IN DATA WAREHOUSE

Vidya Mohanty
Department Of Computer Science & Engineering
Aryan Institute Of Engineering & Technology, Bhubaneswar
Rashmita Panigrahi
Department Of Computer Science & Engineering
Nm Institute Of Engineering & Technology, Bhubaneswar
Arabinda Dash
Department Of Computer Science & Engineering
Capital Engineering College, Bhubaneswar
Nikunja Bihari Kar
Department Of Computer Science & Engineering
Raajdhani Engineering College, Bhubaneswar

Abstract- Data warehouses are huge repositories designed to enable the knowledge workers to make better and faster decisions. Due to its significance in strategic decision making, there is a need to assure data warehouse quality in the presence of evolution events which may be generated as result of change in schema / software or data warehouse requirements. One of the factors affecting the data warehouse quality is view maintenance models quality. Although there are some useful guidelines for designing good view maintenance models, but objective indicators, i.e., metrics are needed to help designers to develop quality view maintenance models. In our previous work, a quality metric for View maintenance models of data warehouse is proposed [25] However, the proposal lacks theoretical and empirical validation of the metric proposed. Hence, the metric practical utility could not be established. This paper validates the metrics both theoretically and empirically. The theoretical validation is performed using Briand's property based framework [63] while empirical validation is carried out using MVPP (Multiple View Processing Plan) to explore the relationship between the proposed metrics and cost efficiency of View maintenance models. The results show that all the four metrics NBR, NVM, NAMV and NFMV have significant impact on the cost efficiency of View maintenance models.

Keywords-Data Warehouse; Data Warehouse Evolution; View maintenance models; Multiple View Processing Plan; Business Intelligence;

I. INTRODUCTION

Having accurate and up-to-date data warehouses is essential for Business Intelligence and Decision Support. A data warehouse (DW) design, apart from performance guarantees, should also provide correctness guarantees. Every time an evolution event occurs anywhere in the warehouse environment (e.g., a design change at the operational sources) it should be smoothly absorbed without causing any further inconvenience. For achieving this, the warehouse and its counterparts should be easily maintainable and the process of populating it should not be destructed by evolution events. The Evolution of a DW constitutes the backbone of typical data warehouse architecture.

Most of the research for improving DW evolution designs has focused solely on improving performance. However, based on practical experience, maintenance makes up for up to 60 % of the resources spent in a warehouse project [34], and therefore, maintainability is an important factor for the determination of the quality of a design [19,36]. Although practitioners are well aware of this problem, still, we miss a formal and concrete answer to deals with the quality perspective for View Maintenance models of DW. George et.al. [49] Proposed a set of metrics for the evaluation of the vulnerability warehouse modules to future changes and for the assessment of the quality of alternative designs of the warehouse. These proposed metrics are based on graph-theoretic properties of the warehouse graph to assess the sensitivity of the graph to a set of possible events.

Most of the researches consider only structural properties of the DW evolution or constructs internal to the underlying databases. But the employed approach neither accounts for the constructs surrounding the database into their models, nor the fact that a software construct evolves over time. In practice, the problem is hard since changes in the schema of database-centric systems affect not only its internals but also the surrounding deployed applications. Hence, the minimal interdependence of these software modules results in higher tolerance to subsequent changes and should be measured with a principled theory. Related work for evolution of data-intensive applications [9], view redefinition [10,17,26], and data warehouse evolution [4,5,11,14] has provided rewriting techniques and theoretical cost models. Related work [30] includes an approach to impact analysis and management of schema evolution, which represents the structural properties of the data warehouse schema, along with any views and queries defined over this schema, as a graph. Graph-based model [49] captures all the parts (or, modules) of an environment, i.e., relations, views, and queries (which are practically the parts of ETL scripts that work the underlying data, or the elementary activities of a GUI based scenario that are involved in the ETL process). In [65] authors focused on a set of graph-theoretic metrics for the prediction of evolution impact and their fitness into real-world ETL scenarios. However, these research works are dealing at design level only without considering the maintenance and quality perspective for View Maintenance models (VMM) of DW.

Following this consideration, we have defined a set of metrics for the View Maintenance (VM) models of DW [25]. In this paper, we built upon the aforementioned approaches with the goal of theoretically and empirically validating the proposed metrics. We formally present these metrics and show that such metrics typically act as predictors for the vulnerability of a VM module of a data warehouse (e.g., a dimension table, or external, an aggregated measure etc.) to future changes to the structure of the warehouse. Secondly, they facilitate the assessment of the quality of alternative designs of the data warehouse VM models with a particular viewpoint on the evolution of the data warehouse. Theoretical validation assures that the metric is correctly defined and the metric actually measures what it purports to measure [26]. The theoretical validation helps us to know when and how to apply the metrics. In this work, we will be characterising our metrics using Briand's property based framework [63]. Empirical validation involves carrying out case studies or controlled experiments etc. to prove practical utility of the metrics. In this work, empirical validation is performed using MVPP (Multiple View Processing Plan – A Cost Based approach) [11]. The proposed metrics have been defined for measuring the cost efficiency of data warehouse VM models. The remainder of this paper is structured as follows: Section 2 presents definition of proposed metrics. Section 3 summarizes the theoretical validation of the proposed metrics. Section 4 and 5 deals with the empirical validation of the proposed metrics and threats. Finally, Section 6 draws conclusions and sketches the immediate future works arising from the conclusions reached in this work.

II. PROPOSED METRIC

Metrics should be defined on the basis of clear measurement goals. So, we have defined a set of metrics for the VM models of DW. The proposed sets of metrics are dealing with two major characteristics. Firstly, they act as predictors for the vulnerability of a VM module of a data warehouse (e.g., a dimension table, or external, an aggregated measure etc.) to future changes to the structure of the warehouse. Secondly, they facilitate the assessment of the quality of alternative designs of the data warehouse VM models with a particular viewpoint on the evolution of the data warehouse. The proposed metrics have been defined for measuring the cost efficiency of data warehouse VM models [25].

1. Number of Base relations: These metric counts the number of base Relations in the view maintenance models for data warehouse.
2. Number of Views materialized: These metric counts the Number of Views Materialized in the view maintenance models for data warehouse.
3. Number of Attributes in materialized views: This metric records the total number of attributes considered for data warehouse VMM.
4. Number of Foreign Keys in materialized views: This metric counts the number of foreign keys in data warehouse VMM.

III. THEORETICAL VALIDATION OF THE METRICS USING PROPERTY FRAMEWORK

The main goal of theoretical validation is to assure that the metric is correctly defined and to assess whether the metric actually measures what it purports to measure [26], such as structural complexity. The theoretical validation helps us to know when and how to apply the metrics. In this work, we will be characterising our metrics using Briand's framework. The framework is briefly discussed in the following subsection:

A. Introduction to Briand's et al. framework [63]

The Briand et al. framework [63] provides a set of mathematical properties that characterise and formalise several important measurement concepts: size, length, complexity, cohesion and coupling. These factors contribute towards structural complexity and hence, have significant impact on quality. The properties defined by Briand's framework may be used to guide the search for new measures and also help to avoid future confusion, often encountered in the literature, about which properties product measure (metrics) should or should not have. Hence, studying metrics properties is important to provide discipline and rigour in the search for new product metrics. In this work, we examine the proposed metrics using Briand's framework to ensure that the metrics are satisfying the properties required for size, length, complexity measures and hence, are contributing towards structural complexity of multidimensional schemas. This framework is based on a graph-theoretic schema of a software artefact, which is seen as a set of elements linked by relationships. Basic concepts of this framework [63] are defined as follows:

Systems and Modules: A system S is represented as a pair ⟨E, R⟩, where E represents the set of elements of S, and R is a binary relation on E (R ⊆E × E) representing the relationships between S's elements. Given a system S as ⟨E, R⟩, a sub-system m = ⟨Em, Rm⟩ is a module of S if and only if Em⊆E, Rm⊆Em × Em and Rm⊆R.

The elements of a module are connected to the elements of the rest of the system by incoming and outgoing relationships.

The Input R(m) is set of of relationships from elements outside module m = ⟨Em, Rm⟩ to those elements present inside module m is defined as Input R(m) = ke1, _ e2l [R|e2 [Em and e1 [E – Em]

The set Output R(m) is a set of relationships from the elements of a module m = ⟨Em, Rm⟩ to those of the rest of the system is defined as Output R(m) = _ke1, e2l [R|e1 [Em and e2 [E – Em]

MS = ⟨E, R, M⟩ is a modular system if S = ⟨E, R⟩ is a system according to the definition given above and M is a collection of modules of S. The validation of metrics using Briand's framework is discussed as follows.

B. Theoretical validation of the proposed metrics

For the purpose and according to Briand's et al. framework, multidimensional schemas qualify as 'system'. Dimensions, facts, dimension hierarchies or levels in materialized views may be the 'modules' of the 'system'. Attributes in dimensions, facts and levels in dimensional hierarchies are the 'elements'. The relationships required to define the 'system' may be the relationships between the dimensions and facts; among attributes of dimensions; among attribute of facts; among attributes of the levels of hierarchies etc. The theoretical validation (using Briand's framework [63]) of each metric considered in this work is discussed in detail as follows:

1. Theoretical validation of NMV metric:

The NMV metric satisfies all the properties required to characterise a metric as complexity metric as shown below: The complexity of a system S (materialized views) is a function Complexity(S) can be characterized by the below following properties Complexity.1 - Complexity.5.

- Non-negativity: If there are any materialized views in the system, it must be characterised by some levels of complexity in the materialized views otherwise it can be zero. Hence, NMV can be either zero or greater than zero. However, it cannot be negative, indicating the validity of this property for NMV.
- Null value: If there are no attributes in dimensions in the multidimensional schema, there will not be any complexities and hence, NMV is null. So, this property holds true for NMV metric. So, The complexity of a system S=⇨<E,R> is non-negative. Complexity (S) ≥ 0 i.e., the complexity of a system S = <E,R> is null if R is empty; R = Ø Complexity(S) = 0
- Symmetry: The complexity of a system S =⇨E,R> does not depend on the convention chosen to represent the relationships between its elements: (S=<E,R> **and** S^{-1}=<E,R^{-1}>) Complexity(S) = Complexity(S^{-1}) So, suppose the a system S = ⟨E, R⟩ is equal to sum of the sizes of two of its modules m1 = ⟨Em1, Rm1⟩ and m2 = ⟨Em2, Rm2⟩ such that any element of S is an element of either m1 or m2.then the complexity of the S does not depend on the convention chosen to represent the relationships between its elements.
- Module Monotonicity: The complexity of a system S=⇨<E,R> is no less than the sum of the complexities of any two of its modules with no relationships in common (S = <E,R> **and** m1 = <Em1,Rm1> **and** m2 = <Em2,Rm2> **and** m1 U m2 ≤ S **and** Rm1 ∩ Rm2 = Ø) Complexity(S)≥ Complexity(m1)+Complexity(m2) For instance, the complexity of the system is not smaller than the sum of the complexities of m1 and m2.

- So, this property is true for NMV as if there are two dimensions in the multidimensional schemas which are not connected to each other, then adding a relationship between two attributes (elements) of different dimensions (modules) will not decrease the value of NMV.
- Disjoint Module Additively: The complexity of a system S⟹<E,R> composed of two disjoint modules m1, m2 is equal to the sum of the complexities of the two modules (S = <E,R> **and** S = m1 U m2 **and** m1 ∩ m2 = Ø) Complexity(S) = Complexity(m1) + Complexity(m2) For instance, the complexity of system S is the sum of the complexities of its modules m1, m2, and m3. As a consequence of the above properties Complexity.1 - Complexity.5, it can be shown that adding relationships between elements of a system does not decrease its complexity.(S⟹E,R'> **and** S" = <E,R"> **and** R' ≤ R") Complexity(S') ≤ Complexity(S") So, the length of a system S = ⟨E, R⟩ made of two disjoint modules m1, m2 is equal to the maximum of the lengths of m1 and m2. The metric NMV calculates the maximum number of views materialized in the system. Hence, if there are 'x' levels in a materialized views in one modules and 'y' levels in a materialized views in another module, then the value of NMV of the system will be x if x > y.

2. Theoretical validation of NAMV metric:

It is demonstrated below that NAMV satisfies all the properties for a size measure.

- Non-negativity: The size of a system S = ⟨E, R⟩ is non-negative, that is, size (S) ≥ 0. A multidimensional schema may or may not have attributes in materialized views. Hence, either NAMV = 0 or NAMV > 0, but it cannot be less than zero. Hence, NAMV satisfy this property of size measure.
- Null value: According to this property of size measure,' 'if there are no elements in the system, then size of the system will be null'. If there are no elements, that is, no attributes in the multidimensional schema, there will not be any materialized views, hence the NAMV, that is, NAMV = 0. Thus, this property is also satisfied by NAMV.
- Module additivity: The size of a system S = ⟨E, R⟩ is equal to sum of the sizes of two of its modules m1 = ⟨Em1, Rm1⟩ and m2 = ⟨Em2, Rm2⟩ such that any element of S is an element of either m1 or m2.A multidimensional schema is composed of modules like dimensions and if these modules have no multiple materialized views in common, then NAMV in the 'system' is equal to the sum of the NAMV present in the 'modules'. Hence, NAMV satisfy the module additivity property of a size measure. The above properties characterise a metric as size metric. It is shown that NAMV metric satisfies all the properties for size metric.

3. Theoretical validation of NBR metric:

It is demonstrated below that NBR satisfies all the properties for a size measure.

- Non-negativity: The size of a system S = ⟨E, R⟩ is non-negative, that is, size (S) ≥ 0. A multidimensional schema may or may not have materialized views. Hence, either NBR = 0 or NBR > 0, but it cannot be less than zero. Hence, NBR satisfy this property of size measure.
- Null value: According to this property of size measure,' 'if there are no elements in the system, then size of the system will be null'. If there are no elements, that is, no attributes in the multidimensional schema, there will not be any materialized views, hence the NBR, that is, NBR = 0. Thus, this property is also satisfied by NBR.
- Module additivity: The size of a system S = ⟨E, R⟩ is equal to sum of the sizes of two of its modules m1 = ⟨Em1, Rm1⟩ and m2 = ⟨Em2, Rm2⟩ such that any element of S is an element of either m1 or m2.A multidimensional schema is composed of modules like dimensions and if these modules have no multiple materialized views in common, then NBR in the 'system' is equal to the sum of the NBR present in the 'modules'. Hence, NBR satisfy the module additive property of a size measure. The above properties characterise a metric as size metric. It is shown that NBR metric satisfies all the properties for size metric.

4. Theoretical validation of NFMV metric:

The NFMV metric satisfies all the properties required to characterise a metric as length metric as shown below:

- Non-negativity: A system (multidimensional schema) may or may not have a dimension in materialized views. If there are any materialized views in the system, it must be characterised by some levels in the materialized views otherwise number of levels will be zero. Hence, NFMV can be either zero or greater than zero. However, it cannot be negative, indicating the validity of this property for NFMV.
- Null value: If there are no attributes in dimensions in the multidimensional schema, there will not be any hierarchies and hence no levels, that is, NMMV is null. So, this property holds true for NFMV metric.
- Non-increasing monotonicity of the connected components: Let S be a system and m be a module of S such that m is represented by a connected component of the graph representing S. Adding relationships between the elements of m definitely, does not increase the length of system S. This property is satisfied by NFMV metric as adding a relationship between elements of a module does not increase the value of NFMV metric, that is, number of levels in materialized views.
- Non-decreasing monotonicity for non-connected components: Let S be a system. m1 and m2 be two modules of S such that m1 and m2 are represented by two separate connected components of the graph representing S.

- Adding relationships from the elements of m1 upto the elements of m2 definitely, does not decrease the length of system S. This property is true for NFMV as if there are two dimensions in the multidimensional schemas which are not connected to each other, then adding a relationship between two attributes (elements) of different dimensions (modules) will not decrease the value of NFMV.
- Disjoint modules: The length of a system S = ⟨E, R⟩ made of two disjoint modules m1, m2 is equal to the maximum of the lengths of m1 and m2. The metric NFMV calculates the maximum number of levels in the hierarchies in the system. Hence, if there are 'x' levels in a materialized views in one modules and 'y' levels in a materialized views in another module, then the value of NFMV of the system will be x if x > y. Hence, NFMV metric satisfies all the properties of the length measure. The summary of characterization of the proposed metrics is summarized in Table 1.

TABLE 1: CHARACTERISATION OF METRICS USING BRIAND'S PROPERTIES [63]

Metrics Properties	NBR	NVM	NAMV	NFMV
Non negativity	Yes	Yes	Yes	Yes
Null value	Yes	Yes	Yes	Yes
Module Additivity	Yes	-	Yes	No
Non-Increasing Monotonicity	-	-	-	Yes
Non-Decreasing Monotonicity	-	-	-	Yes
Disjoint module	-	-	-	Yes
Symmetry	-	Yes	-	-
Module Monotonicity	-	Yes	-	-
Disjoint Module Additivity	-	Yes	-	-
Conclusion	**Size**	**Complexity**	**Size**	**Length**

We observed that the metrics considered in this paper satisfy all the properties which characterise the metric as either the size, length or complexity measure. NMV only accounts for complexity measure also. All the other metrics considered in this work are of size and length measures. The size, complexity and length represent interesting internal attributes of materialized views and significantly contribute towards its structural complexity. Hence, we have sufficient reasons to claim that the metrics are theoretically valid and contribute significantly towards structural complexity. It is widely accepted that size and length measures are significant measures which may be used for assessment, prediction and improvement purposes [15, 25, 26].

According to Briand's framework, the size properties, properties of Complexity measure and length properties hold when applying the admissible transformation of ratio scale. There is no contradiction between the concept of size/length and the definition of size/length measures on a ratio scale [23]. Also, there is no contradiction between concept of complexity and the definition of complexity measures on a ratio scale. Hence, we can interpret that the metrics proposed in this work is also at ratio scale. .

IV. EMPIRICAL VALIDATION

After proving theoretical validity of metrics defined in Section 2, in this section, we intend to corroborate that these metrics are really related to cost efficiency of View Maintenance Models of Data Warehouse, i.e. measure of internal consistency will be applied. This is done on the basis of empirical validation. We have conducted controlled experiment for this reason. The dependent variable in this study is cost efficiency and the independent variables are the View Maintenance metrics defined in Section 2. Twenty One multidimensional schemas were collected to carry out this controlled experiment. The domains of these selected models were general and well known in order to avoid the problems arising out of domain understanding. The metrics values for each schema were collected manually using MVPP approach [11]. There is no subjectivity in calculation of metrics values, as the metrics considered in this work are calculated by simply counting (e.g. NBR is calculated by counting the base relations present in the schemas). The collected data was shown in Table 2.

TABLE 2: METRICS VALUES

NBR	NVM	NAMV	NFMV	Cost (Query processing and Maintenance cost)
23	13	46	3	70
24	13	46	2	75
31	9	47	6	80
27	5	35	6	63
18	7	30	3	41
24	7	37	5	54

40	8	54	7	91
46	9	64	9	95
22	9	38	5	56
18	5	28	2	35
14	5	23	1	37
13	7	25	1	39
19	9	34	3	45
23	10	42	4	66
26	13	49	3	62
44	16	75	6	98
23	13	25	2	44
17	11	38	3	43
26	5	35	4	48
27	11	49	3	64
38	12	60	6	88

Cronbach’s alpha is used to measure internal consistency, i.e., how closely a set of variables are related. It is also considered to be a measure of so called scale reliability. A “high” value for alpha does not simply that the measure is unidimensional. Cronbach’s alpha may be written as a function of the number of test variables and the average inter-correlation among the variables. If the number of variables is increased, Cronbach’s alpha is also increased. Additionally, if the average of inter-item correlation is low, the alpha value will be low. So, as the average inter-item correlation increases, Cronbach’s alpha also increases as well.

TABLE 3: RELIABILITY STATISTICS

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.829	.908	5

TABLE 4:INTER-ITEM CORRELATION MATRIX

	NBR	NVM	NAMV	NFMV	Cost
NBR	1.000	.372	.894	.863	.921
NVM	.372	1.000	.607	.052	.505
NAMV	.894	.607	1.000	.707	.932
NFMV	.863	.052	.707	1.000	.777
Cost	.921	.505	.932	.777	1.000

TABLE 5: ITEM-TOTAL STATISTICS

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
NBR	116.90	1335.790	.921	.917	.732
NVM	133.38	1893.348	.504	.660	.856
NAMV	100.86	1032.829	.946	.913	.683
NFMV	138.76	1899.690	.766	.838	.853
Cost	81.14	659.129	.951	.910	.757

Cronbach's alpha is 0.829 in table 3, which shows a high level of internal consistency among variables. The Item-Total Statistics table 5 represents the "Cronbach's Alpha if Item deleted" in the final column. This column presents that none of the values are less than 0.50, i.e. none of the variables can be deleted from the scale. Cronbach's alpha provided with an overall high reliability coefficient for the set of variables used. This shows that all the measures are significant in assessing or predicting the cost efficiency of View Maintenance Models of Data Warehouse and are highly correlated.

V. THREATS TO VALIDITY OF RESULTS

Different threats exist regarding the validity of results in a performed experiment. In this section, we have discussed the problems that could affect the construct, internal, external and conclusion validity of the results of the experiment and how we have tried to solve them or at least tried to minimise them by different means.

A. Internal validity: The internal validity is the degree to which conclusions can be drawn about the causal effect of independent variables on the dependent variables. The following issues should be considered:

- Differences among schemas. The domains of the schemas were different and this could influence the results obtained in some way.
- Precision in the metric values.
- Learning effects.
- Fatigue effects. The average time for completing the wok was smaller than an hour. With this range of times we believe that fatigue effects hardly exist at all. Furthermore, the different order of the tests helped to avoid these fatigue effects.
- Persistence effects. In our case, persistence effects are not present because such manner of work never accomplished.
- Work motivation.
- Plagiarism

B. External validity: The external validity is the degree to which the results of the research can be generalised to the population under study and to other research settings. The greater the external validity, the more the results of an empirical study can be generalised to actual software engineering practice. Two threats to validity have been identified which limit the ability to apply such generalisation:

- Materials and tasks used. We tried to use schemas and operations representative of real world cases in the experiments, although more experiments with larger and more complex schemas could have been used.
- Approach Applied.

C. Conclusion validity: The conclusion validity defines the extent to which conclusions are valid. The only issue that could affect the validity of this study is the size of the sample data (21 schemas), which perhaps is not enough for both parametric and non-parametric tests. We will try to obtain bigger sample data through more experimentation.

D. Construct validity: The construct validity is the degree to which the independent and the dependent variables are accurately measured by the measurement instruments used in the study. The dependent variable we use understands cost, so we consider this variable constructively valid. The construct validity of the measures used for the independent variables is guaranteed by the Distance and Property framework used for their theoretical validation. Although, we know that several aspects threaten the validity of the results, we have tried to alleviate them by different means. In this section we have discuss the problems that could affect the results of the experiment and how we tried to solve them. We know that even though we put a lot of effort in alleviate the threats, some of them can affect the results and could lessen the strength of the results. As we have made a family of experiments and we have obtained the same results in all the experiments, we think that those threats have had a small impact on the results. We plan to make more experiments and case studies varying some empirical settings to get more conclusive results.

VI. CONCLUSION AND FUTURE WORK

In this work, we validated the metrics, to evaluate the cost efficiency of VM models for data warehouse. These metrics are theoretically validated using Briand's property framework and it is concluded that these metrics are theoretically sound and are characterised as either size or length measure. We have also conducted controlled experiment in order to provide the empirical validation of the proposed metrics. The empirical validation using controlled experiments suggests that the metrics have significant effect on the cost efficiency parameter. Hence, this study shows that these metrics are significantly contributing towards the cost efficiency of the VM models. Although, this is not a complete set of metrics to assess cost efficiency of VM models but these metrics along with other already proposed may act as objective indicators for the quality attribute. This set of metrics may not be comprehensive and other consecutive research could further complete this set by defining new metrics from other different perspectives. Replicated studies with more data need to be carried out to generalise the results. These metrics need to be empirically validated with the help of case studies or industrial data and by involving professionals to draw strong conclusion which can be applied in practice.

REFERENCES

1. Inmon, W.H., Building the Data Warehouse. John Wiley, 1992.
2. Bellahsene, Z.: Schema evolution in data warehouses. Knowl. and Inf. Syst. 4(2) (2002)
3. Kimball, R. The Data Warehouse Toolkit. John Wiley, 1996.
4. Blaschka, M., Sapia, C., Höfling, G.: On Schema Evolution in Multidimensional Databases. In: Mohania, M., Tjoa, A.M. (eds.) DaWaK 1999. LNCS, vol. 1676. Springer, Heidelberg (1999)
5. Fan, H., Poulovassilis, A.: Schema Evolution in Data Warehousing Environments – A Schema Transformation-Based Approach. In: Atzeni, P., Chu, W., Lu, H., Zhou, S., Ling, T.-W. (eds.) ER 2004. LNCS, vol. 3288. Springer, Heidelberg (2004)

6. Favre, C., Bentayeb, F., Boussaid, O.: Evolution of Data Warehouses' Optimization: A Workload Perspective. In: Song, I.-Y., Eder, J., Nguyen, T.M. (eds.) DaWaK 2007. LNCS, vol. 4654. Springer, Heidelberg (2007)
7. Zuse, H. A.: 'Framework of Software measurement', Walter de Gruyter, Berlin, 1998.
8. Gupta, A., Mumick, I.S., Rao, J., Ross, K.: Adapting materialized views after redefinitions: Techniques and a performance study. Information Systems (26) (2001)
9. Gupta, A., I.S. Mumick, "Maintenance of Materialized Views: Problems, Techniques, and Applications." Data Eng. Bulletin, Vol. 18, No. 2, June 1995.
10. Nica, A., Lee, A.J., Rundensteiner, E.A.: The CSV algorithm for view synchronization in evolvable large-scale information systems. In: Schek, H.-J., Saltor, F., Ramos, I., Alonso, G. (eds.) EDBT 1998. LNCS, vol. 1377. Springer, Heidelberg (1998)
11. Jian Yang, Kamalakar Karlapalem, Qing Li, "A Framework for Designing Materialized Views in Data Warehousing Environment "Proceedings of the 17th International Conference on Distributed Computing Systems (ICDCS '97), IEEE 1997
12. Yousry Taha, Arsany S. Sawiros, Noha Adly, "An efficient data warehousing framework" Computing - Technology and engineering, e-Publisher: CiteSeerX , 2009.
13. Miranda Chan, Hong Va Leong, Antonio Si, "Incremental Update to Aggregated Information for Data Warehouses over Internet" DOLAP 2000 ACM, ISBN 1-58113-323-5/00/0011
14. José A. Rodero, José A.Toval, Mario G. Piattini, "The audit of the Data Warehouse Framework" Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'99) Heidelberg, Germany, 14. - 15. 6. 1999
15. M. Golfarelli, S. Rizzi, "A Methodological Framework for Data Warehouse Design", Proceedings of First International Workshop on Data Warehousing and OLAP (DOLAP, in connection with CIKM'98), Washington, D.C., USA, November 1998.
16. Darja Solodovnikova and Laila Niedrite," Evolution-Oriented User-Centric Data Warehouse", Proceedings of the 19th International Conference on Information Systems Development by Springer, 2010
17. C. Quix. "Repository Support for Data Warehouse Evolution". In Proc. of the Intl. Workshop DMDW, Heidelberg, Germany (1999)
18. Anisoara Nica, Elke A. Rundensteiner," Using Containment Information for View Evolution in Dynamic Distributed Environments" DEXA '98 Proceedings of the 9th International Workshop on Database and Expert Systems Applications ,Page 212 , IEEE Computer Society Washington, DC, USA1998
19. Dragan Sahpaski, Goran Velinov, Boro Jakimovski, Margita Kon-Popovska," Dynamic Evolution and Improvement of Data Warehouse Design" Fourth Balkan Conference in Informatics, IEEE,2009
20. Claudine Bréant, Gérald Thurler, François Borst, Antoine Geissbuhler, "Design of a Multi Dimensional Database for the Archimed DataWarehouse" Connecting Medical Informatics and Bio-Informatics R. Engelbrecht et al. (Eds.) ENMI, 2005
21. George Papastefanatos, Panos Vassiliadis, Alkis Simitsis, Yannis Vassiliou," Design Metrics for Data Warehouse Evolution" ER 2008, LNCS 5231, pp. 440–454, 2008.
22. Matthias Jarke, Christoph Quix, Diego Calvanese, Maurizio Lenzerini, Enrico Franconi, Spyros Ligoudistianos, Panos Vassiliadis, Yannis Vassiliou, " Concept Based Design of Data Warehouses: The DWQ Demonstrators". SIGMOD Conference 2000: 591
23. David Botzer, Opher Etzion, "Optimization of Materialization Strategies for Derived Data Elements," IEEE Transactions on Knowledge and Data Engineering, vol. 8, no. 2, pp. 260-272, April, 1996.
24. Golfarelli, M., Lechtenbörger, J., Rizzi, S., Vossen, G.: Schema versioning in data warehouses: Enabling cross-version querying via schema augmentation. Data Knowl. Eng. 59(2), 435–459 (2006).
25. Gosain A., Sabharwal S., Gupta R., "Quality Metrics for View Maintenance Models of Data Warehouse", ERCICA 2014, Bangalore, India.
26. Gosain A., Sabharwal S., Gupta R., "An Efficient Feature Selection Approach for Materialized Views" IEEE, ICCCCM2013, Allahabad, India.
27. Zuse,H.: Properties of software measures, Software Quality Journal, 1992, 1, pp. 225- 260.
28. Darja Solodovnikova and Laila Niedrite," Evolution-Oriented User-Centric Data Warehouse", Proceedings of the 19th International Conference on Information Systems Development by Springer, 2010
29. Dimitri Theodoratos, Mokrane Bouzeghoub," A General Framework for the View Selection Problem for Data Warehouse Design and Evolution" DOLAP '00 Proceedings of the 3rd ACM international workshop on Data warehousing and OLAP, Pages 1 – 8, ACM New York, NY, USA ©2000
30. M. Blaschka. "FIESTA: A Framework for Schema Evolution in Multidimensional Information Systems". In 6thCAiSE Doctoral Consortium, Heidelberg, 1999.

31. C. Quix. "Repository Support for Data Warehouse Evolution". In Proc. of the Intl. Workshop DMDW, Heidelberg, Germany (1999)
32. Anisoara Nica, Elke A. Rundensteiner," Using Containment Information for View Evolution in Dynamic Distributed Environments" DEXA '98 Proceedings of the 9th International Workshop on Database and Expert Systems Applications ,Page 212 , IEEE Computer Society Washington, DC, USA1998
33. Mahesh B. Chaudhari, Suzanne W. Dietrich, "A Distributed Event Stream Processing Framework for Materialized Views over Heterogeneous Data Sources", VLDB 2010, Singapore.
34. Ericka-Janet Rechy-Ram´ırez , Edgard Ben´ıtez-Guerrero," A Model and Language for Bitemporal Schema Versioning in DataWarehouses" Proceedings of the 15th International Conference on Computing (CIC'06), IEEE2006.
35. S. Chen, X. Zhang, E.A. Rundensteiner. "A Compensation-based Approach for Materialized View Maintenance in Distributed Environments". In Computer Science Technical Report, Worcester Polytechnic Institute, Worcester, MA, USA (2004)
36. Chuan Zhang, Jian Yang," Materialized View Evolution Support in DataWarehouse Environment" Sixth International Conference on Database Systems for Advanced Applications (DASFAA'99), 1999.
37. Dragan Sahpaski, Goran Velinov, Boro Jakimovski, Margita Kon-Popovska," Dynamic Evolution and Improvement of Data Warehouse Design" Fourth Balkan Conference in Informatics, IEEE,2009
38. Amy J. Lee, Anisoara Nica, Elke A. Rundensteiner," The EVE Approach: View Synchronization in Dynamic Distributed Environments", IEEE transactions on knowledge and data engineering, vol. 14, no. 5, september/october 2002
39. Robert M. Bruckner, Tok Wang Ling, Oscar Mangisengi, A Min Tjoa," A Framework for a Multidimensional OLAP Model using Topic Maps" IEEE 2002.
40. Xin Zhang, Elke A. Rundensteiner," The SDCC Framework For Integrating Existing Algorithms for Diverse Data Warehouse Maintenance Tasks" Database Engineering and Applications, 1999. IDEAS '99. International Symposium Proceedings , Aug 1999 Page 206 - 214
41. PAN Ding, PAN Yunshan," Metadata Versioning for DW2.0 Architecture" Proceedings of the 29th Chinese Control Conference July 29-31, 2010, Beijing, China
42. Claudine Bréant, Gérald Thurler, François Borst, Antoine Geissbuhler, "Design of a Multi Dimensional Database for the Archimed DataWarehouse" Connecting Medical Informatics and Bio-Informatics R. Engelbrecht et al. (Eds.) ENMI, 2005
43. C´ecile Favre, Fadila Bentayeb, and Omar Boussaid, "Evolution of Data Warehouses' Optimization: A Workload Perspective" DaWaK 2007, LNCS 4654, pp. 13–22, 2007.
44. Bartosz Bebel, Zbyszko Królikowski, and Robert Wrembel, "Managing Evolution of Data Warehouses by Means of Nested Transactions", ADVIS 2006, LNCS 4243, pp. 119–128, 2006
45. J. A. Nasir, M. Khurram Shahzad, "Architecture for Virtualization in Data Warehouse" Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, 243–248. 2007 Springer.
46. M.K. Shahzad, J.A. Nasir, M.A. Pasha. "CEV-DW: Creation and Evolution of Versions in Data Warehouse". In Asian Journal of Information Technology, 4(10) (2005) 910-917
47. E. Ben´ıtez-Guerrero, C. Collet, and M. Adiba. "The WHES Approach to Data Warehouse Evolution". Digital Journal e-Gnosis [online], http://www.e-gnosis.udg.mx, ISSN No. 1665-5745, 2003.
48. Joseph M. Firestone," Architectural Evolution in DataWarehousing and Distributed Knowledge Management Architecture" White Paper No. Eleven July 1, 1998.
49. George Papastefanatos, Panos Vassiliadis, Alkis Simitsis, Yannis Vassiliou," Design Metrics for Data Warehouse Evolution" ER 2008, LNCS 5231, pp. 440–454, 2008.
50. George Papastefanatos, Panos Vassiliadis, Alkis Simitsis, Konstantinos Aggistalis, Fotini Pechlivani, Yannis Vassiliou, "Language extensions for the automation of database schema evolution" ICEIS (1) 2008: 74-81.
51. B. Ashadevi, Dr. P. Navaneetham," A Framework for the View Selection Problem in Data Warehousing Environment" International Journal on Computer Science and Engineering Vol. 02, No. 09, 2010, 2820-2826
52. B. Bebel, Z. Królikowski, R. Wrembel," Formal approach to modelling a multiversion data warehouse" Bulletin of the polish academy of sciences technical sciences Vol. 54, No. 1, 2006
53. Garima Thakur, Anjana Gosain," DWEVOLVE: A Requirement Based Framework for Data Warehouse Evolution" ACM SIGSOFT Software Engineering Notes, Page 1, November 2011 Volume 36, No.6.
54. Resmi Nair, Campbell Wilson, Bala Srinivasan," A Conceptual Query-Driven Design Framework for Data Warehouse" , World Academy of Science, Engineering and Technology 25 , 2007.
55. Jose-Norberto Maz´on, Juan Trujillo," A Hybrid Model Driven Development Framework for the Multidimensional Modeling of Data Warehouses" , SIGMOD Record, June 2009 (Vol. 38, No. 2)

56. Elena Ferrari, Elisa Bertino, Claudio Bettini, Claudio Bettini, "On Using Materialization Strategies for a Temporal Authorization Model", Proc. Post-SIGMOD Workshop Materialized Views: Techniques and Applications, 1996.
57. Ernest Teniente , Toni Urpí, "A Common Framework for Classifying and Specifying Deductive Database Updating Problems" Proc. Int. Conf. on Data Engineering (ICDE '95).
58. Richard Hull, Gang Zhou. "A Framework for Supporting Data Integration Using the Materialized and Virtual Approaches". In Proceedings of the ACM SIGMOD International Conference on Management of Data, pages 481-492, June 1996
59. Solodovņikova D. "The Formal Model for Multiversion Data Warehouse Evolution", Postconference proceedings of the 8th International Baltic Conference on Databases and Information Systems, Frontiers in Artificial Intelligence and Applications, IOS Press, 2008.
60. D. Agrawal, A. El Abbadi, A. Singh, T. Yurek., "Efficient View Maintenance in Data Warehouses". In Proceedings of the 1997 ACM International Conference on Management of Data, pages 417-427, May 1997.
61. Thilini Ariyachandra, Hugh Watson, "Key organizational factors in data warehouse architecture selection", 2010 Elsevier.
62. Briand LC, Wuest J, Ikonomovski S and Lounis H (1999) Investigation of quality factors in object oriented designs: An industrial case study. Proc. of 21st International Conf. on Software Engineering, Los Angeles, pp 345-354.
63. Briand, L.C., Morasca, S., Basili, V.R.: 'Property based software engineering measurement', IEEE Trans. Softw. Eng., 1996, 22, pp. 68–86.
64. H. Gupta, 'Selection of views to materialize in a data warehouse', ICDT'97, Springer 1997.
65. George Papastefanatos, Panos Vassiliadis, Alkis Simitsis, Yannis Vassiliou, 'Metrics for the Prediction of Evolution Impact in ETL Ecosystems: A Case Study' Springer-Verlag 2012.