ISSN:0975-9646

Axiomization of nw-MVD

V. Sangeetha[#1], V. Valli Kumari[*2]

[#]*Department of CSE, Andhra University*
Waltair, India

Abstract— **Nesting of relations is very much essential for many database representations in reality. Nesting of relations tends to reduce the data redundancy in the relations. It helps in dropping certain constraints in normalization which tend to omit certain sensitive data when tried to be normalized. Query processing or conversion of data repositories, as required, can be easily being tackled by the aid of this maintenance.**
Addressing of certain conditions about the commuting properties between different relations is dealt. It has been proved that the maintenance of databases with respect to multivalued dependency is enough for nesting that relation, as required. Multivalued dependency, though omits the rule of being in accordance to normalization, is very much useful to preserve sensitive data relations. More prominently weak multivalued dependencies are being maintained among the relations, where each relation is aimed
to be nested with other relations without loss of generality.
Certain inference system is being developed with different notations and implications posed on it, for studying multivalued dependent data, that is being nested. Complementation rule, which on being implied to the database, makes the relation to be easily decomposed into smaller relations. Axiomization is being tested on weak multivalued dependent relations, which have been nested.

Keywords— **Axiomization, Normalization, Multivalued dependency, MVD, w-MVD, nw-MVD.**

I. INTRODUCTION

Report generation is the motto for maintaining data in any of its form. There come many issues when there is a need for studying different data repository aspects before a particular report is being generated. There raises many steps where data is being processed by selecting data elements required, cleaning the data from dirt, if any, transforming data as per the level of requirement , combining data from different data slots etc. Major part of this maintenance runs around the concept of data cleaning.

Data cleaning is a typical process in which Relational database proves to be propounded for maintaining information for many commercial management systems. Relational databases are being maintained with both data elementary and the manner in which they are related. These relations are maintained in a flat, more promisingly with two-dimensional table that prevents relational structuring. Data is claimed to have been maintained, described and organized on par of database relational model, according to the design. Relational databases tend to be maintained significantly either in normalized or de-normalized manner.

Normalized form of relationships are maintained at a higher promising level but at the same level have more constraints being satisfied. As maintaining databases on par of normalization poses strict constraints, it will be more tedious when data bases are to be maintained at different hierarchies. Complex-data models have been proposed for overcoming several limitations of database design based on relational model. There has been models depicted where there can be more relations being linked to a single attribute.

Multivalued dependency is the major prospect that has been termed as full constraint between two sets of attributes in a relation. This is also termed as tuple-generating dependency. It can also be treated as special case of join dependency, more prominently termed as 2-ary join dependency. Databases satisfying multivalued dependencies tend to have duplicate values. MVD creates overheads in certain databases ending up with major types of anomalies that result in data loss. MVDs suffer with processing as the constraint is always tightly coupled with the attributes, i.e.; different anomalies comes into picture while processing the database that has been constructed strictly on the basis of MVD.

One solution for solving these types of situations is the use of weak multivalued dependencies. These w-MVDs claim to solve many of the dependency and redundancy removal. Process of nesting can be implemented on such wMVD in order to track down and eliminate data redundancies. The approach has been designed to work on basic synthetic data, which means that it is a real time values corresponding to any real time attribute. Though this process claims to be a processed in terms of de-normalization, as nesting removes the property of database to be in 1NF, in order to handle certain data loss in sophisticated data bases, this process of nesting can be treated as a prominent methodology for tracking the redundancy and data loss aspects. Multivalued dependent databases in its nested form tend to be checked for satisfying certain predominant level constraints. Each of the rules generally termed as inference rules are being checked through a premise and conclusion method.

II. LITERATURE SURVEY

Database technology is being tremendously used for maintaining relations. The countenance dealing with data relating to different levels of processing increases the thirst of finding new methods for maintaining data in terms of related relations. Statistical and machinery level techniques fail in processing this huge amount of data [1].

Many technological issues arise when this huge amount of data is to be processed for report generation. Many quality problems are to be handled when the same data is to be maintained at different hierarchies. Data scrubbing is being used for addressing such problems. Data cleaning tends to be very crucial when there comes requisition for integration of different non-homogenous data repositories.

Schema level transformations are needs more impact of data scrubbing. ETL; process corresponding to extraction, transformation and loading; relates to major part in the process of data cleaning [2] [3].

Special process hardware and software is generally used to work with database relations with more ease. Database machine designs like CASSM, RAP, DIRECT and DBC are generally proposed for working for such database processing. Data quality is aimed and is claimed to be achieved with successful research [4].

Many benchmarks have been posed at a very promising level in order to improve the ease of access, quality of data derived, reduction in time taken for computation, labour required. Scientific approaches have been derived from the era of unscientific approach, by increasing the querying ability. Multiuser benchmarks are also being designed for complex computations [5].

When dealing with data cleaning in the viewpoint of manipulating erroneous databases, generally termed as “garbage in, garbage out”, there will always be a formulation of certain tools of varying functionalities. Low-level programs that are hard to design and manipulate are initially designed. Data quality problems, formed from integration of single and multi-source data are generally being addressed to work at different hierarchies [6]. In general the analysis using monograph is derived which covers semantic, syntax, and computation aspects. Query evaluation, dependency enforcement and its updates are being tackled [7].

In general, data cleaning task includes record matching, de-duplication, and column segmentation which often claims processing beyond traditional query system. Inconsistency of data; which relates to acquiring unreliable information with conflicting versions of data; is the major aspect that is generally handled in the process of data cleaning. Several research efforts propose and investigate comprehensive and uniform treatment of data cleaning covering several transformation phases, specific operators and their implementations [8] [9].

Several extensions have been proposed to the pre-existing language computations like the formation of SQL/PSM as an extension to basic SQL, by primarily allowing a compact specification of each cleaning process. In general probabilistic, empirical and logic-based approaches have been quoted for this aspect of scrubbing [10].

To enhance the mechanism of refining the duplicate records, an external merge sort is being used to the primary sorting algorithms. The token based record match is taken to be sound in process of data scrubbing, which depends on the condition that “adequate tokens that bring potential duplicates together can be easily used in determining record match”. When compared to the basic field matching algorithm (BASIC), the tokenisation method is more likely to handle dirt [11].

The process of refinement is emphasized in large extent on basis of redundancy of relations. Refinement on basis of redundancy aspect, primarily aspects of anomalies will get into picture which degrades the relation status in a table or database on the whole [12]. On par of error correction; defined by detecting semantic error; generally formulated as integrity constraints, syntactical and semantic errors are the major types of errors that may occur. The term conjunction with respect to atoms and variables are the major aspects that are generally dealt. The major constraints correspond to Functional, Inclusion, Equality Generating Dependencies, Tuple generating dependencies and Full dependencies. Now-a-days constraints are becoming a part of data quality tools [13].

Despite of all attempts that are made to detect, clean and improvise dirty in data, the classical question that generally arises is about the methodology to identify how data is regarded as dirty or clean? Even though there have been several mechanisms that were designed for improving the quality of schema. Certain constraints are hard to be detected and to improve the quality of data [14].

Data dependencies play an important role in the design of database. Dependency in database in many databases are not explicitly defined, these are enforced in the transactions, which update the databases. Certain approaches have been designed to detect path patterns for implementing most commonly used methods that enforce data dependency [15]. Relational data model, a promising mechanism to store data elementary, over decades, has many approaches in dealing with the concept of dependencies, more effectively, the aspect of functional dependencies [16].

The traditional dependency consists of process where certain dependencies like inclusion dependencies are fairly used. Recovery of common data dependencies like functional dependencies, key constraints, inclusion dependencies, referential constraints and sum dependencies designed in a database. Conditional functional dependency is treated as an advancement made to the primitive FD by supporting certain patterns with constants which are semantically related. CFDs are aimed for capturing the consistency of data by incorporating bindings of semantically related values. In general, data cleaning task includes record matching, de-duplication, and column segmentation which often claims processing beyond traditional query system. Inconsistency of data; which relates to acquiring unreliable information with conflicting versions of data; is the major aspect that is generally handled in the process of data cleaning [17].

If A -> B, it is that we cannot have more than one value of A for different values of B, but the multivalued dependency claims to have more than one value for such relations. One solution for this is to work on the aspects of multivalued dependencies. Multivalued dependency is full constraint between two sets of attributes in a relation. This is more generally treated as tuple-generating dependency. Databases with multivalued dependencies exhibit redundancy. Thus, in database normalization fourth normal form requires that their every multivalued dependency X->>Y is trivial i.e.; for X->>Y, X should be the super key. Because of the same reason, this relation is considered to be full constraint relationship [18].

The unstructuredness of data degrades the performance of the query analysis. Certain managing aspects to map the unstructured data with relational database constraints and managing of such data between relational and NOSQL can be organized for effective query processing [19].

The major aspect that is needed to be satisfied when a relation is strict to normalize is that the relation maintained is on par of basic normalization constraint, 1NF- is the basic requirement that every tuple in the relation should be mapped to a single relation. One basic solution that is generally adapted is the process of storing each instance in a separate column. However this solution fails if the number of instances increase, the table size increases in the same pattern which ends-up with queries of multiple joins [20].

A non-1NF set of relations containing certain non-key attributes are maintained with the loss of atomicity. One of the major solutions for solving the aspects of maintaining relations without loss of generality is the use of weak multivalued dependencies. The approach has been designed to work on basic synthetic data, which means that it is a real time value corresponding to any real time attribute. It is trust worthy to note that first normal form condition of relational model is restrictive for some applications [21]. Relations are maintained on par of nesting of relations for computations to be performed with more accuracy, less effort and ease.

III. METHODOLOGY

We infer certain rules on the taken relation in its weak multivalued dependencies, where relations tend to be nested.

We define R to be a universal set which consists of the entire permitted relational attribute that we try to infer into the relation.

$$R = \{course, Lady, Gentleman\}$$

On inference to the relations above stated, all the attributes present are being resembled with certain attribute name for computations.

Let X define the course relation, V corresponds to the attribute with values relating to Lady and Y to values corresponding to Gentleman.

Relations tend to map with each other. As already being defined, as the relation is being maintained in multivalued dependency format or more crudely into nesting form that relations, being defined, there is no chance of maintaining the relation in accordance to normalization, as the relation defined does not refer to be in 1NF.

The process here is to produce inference to the certain rules corresponding to the relation maintained.

Let $\Sigma \cup \{\Psi\}$ denotes the set of n-wMVDs that are stated on the relation R, ϕ defines the set of inference rules that are to be proved on the relation, R.

In general, in order to prove the consequences of any rule stated, inference rules are termed to be in the form

$$\frac{premise}{conclusion}$$

Certain inference rules that occur without premise, i.e., which are directly concluded are generally termed to be axioms. Inference rules are selected where R-Complementation is the only rule that is dependent on R, the entire relational set. Rules are also been selected for natural R-Complementation, which is the only rule that is dependent on R. An inference rule is claimed to be R-Sound if the set of dependencies in the premise of the rule R-implies the dependency in the conclusion.

IV. RESULTS AND DISCUSSION

1. *Reflexivity Property*

Any relational attribute which infer itself is generally termed to be in reflexive property. As illustrated, as inference rule is generally termed in accordance to the premise and conclusion rule, the rule of this reflexivity is posed on different sets of attributes in the relation. It has been proven that his rule of Reflexivity is directly being taken to be proven, with a particular conclusion. Thus the relation on Reflexivity is termed as Axiom as here there has been a premise proven to be sound, without a particular conclusion derived. Thus for any of the relational attributes X, V, Y, Reflexivity is obeyed.

2. *Augmentation*

The premise of the relation to be in augmentation is given by

$$X \text{-->>}_w Y$$

From the relation taken, let X corresponds to the course and Y corresponds to Gentleman. Now the part to be proven is that Y is weak multivalued dependent on X being its nested form.

Table3. Relation from X -->>_{nw} Y

Course	Gentleman
Latin	Jose
Latin	Flavio
Latin	Ruy
Latin	Adam

It is clear from the above table3 that the above premise is true in case of course and gentleman. In order to prove this inference rule, it should be shown that the conclusion also holds well on this relational attribute.

The conclusion is being dealt by inferring that the nested form of Course with any of the other attribute of the relation R also holds with the multivalued dependency principle with Gentleman, taken as Y; in its nested form. Thus the relation Course is being nested with the other attribute, Lady, being taken as V and the same is being checked against the set of attributes corresponding to Gentleman, to hold the property of multivalued dependency.

From the table1, the nested from of the set corresponding to Course and Lady is as shown in the table4, below. The mapping derived from this set to Gentleman, Y, is also illustrated in the below table4.

Table4. Relation corresponding to XV->$_{nw}$Y

Course, Lady	Gentleman
{Latin, Racquel}	Jose
{Latin, Racquel}	Flavio
{Latin, Racquel}	Ruy
{Latin, Racquel}	Adam
{Latin, Lorena}	Jose
{Latin, Lorena}	Flavio
{Latin, Lorena}	Ruy
{Latin, Lorena}	Adam
{Latin, Flor}	Jose
{Latin, Flor}	Flavio
{Latin, Flor}	Ruy
{Latin, Flor}	Adam
{Latin, Eve}	Jose
{Latin, Eve}	Flavio
{Latin, Eve}	Ruy
{Latin, Eve}	Adam

3. *Transport 2 Rule, T2*

Premise: XV-->$_{nw}$ Y

The premise of this inference has already proven as a conclusion and is as shown in the table4. Now to infer this rule, it should be shown that the conclusion also holds well as per weak multivalued dependency rule, where the relations are termed to be in its nesting form.

The conclusion part is given by

XV-->>$_{nw}$ YV

Table4. Relation XV-->>$_{nw}$ YV

Course, Lady	Gentleman, Lady
{Latin, Racquel}	{Jose, Racquel}
{Latin, Racquel}	{Flavio, Racquel}
{Latin, Racquel}	{Ruy, Racquel}
{Latin, Racquel}	{Adam, Racquel}
{Latin, Lorena}	{Jose, Lorena}
{Latin, Lorena}	{Flavio, Lorena}
{Latin, Lorena}	{Ruy, Lorena}
{Latin, Lorena}	{Adam, Lorena}
{Latin, Flor}	{Jose, Flor}
{Latin, Flor}	{Flavio, Flor}
{Latin, Flor}	{Ruy, Flor}
{Latin, Flor}	{Adam, Flor}
{Latin, Eve}	{Jose, Eve}
{Latin, Eve}	{Flavio, Eve}
{Latin, Eve}	{Ruy, Eve}
{Latin, Eve}	{Adam, Eve}

4. *R-Complementation*

The premise of this inference rule is given by X-->>$_{nw}$ Y. The premise as stated is already depicted in table3 and has been already proven. Now, the conclusion part is being considered which is given as

X -->>$_{nw}$ R-XY

Here R corresponds to the complete relation given by

R= {Course, Lady, Gentleman},

Where X => Course, V=>Lady and Y=> Gentleman

The conclusion part has an augment R-XY which states the rule when the complete relation with all attributes is being set differenced with the inclusion set of {{course}, {Gentleman}}

R-XY= {Course, Lady, Gentleman} – {{Course} {Gentleman} }

⇨ {Lady}

Now, it is to be shown that X -->>V

Table5. Relation X-->> V

Course	Lady
Latin	Racquel
Latin	Lorena
Latin	Flor
Latin	Eve

5. *Natural R-Complementation*

R = {Course, Lady, Gentleman}

Where X=> Course, V=> Lady, and Y=> Gentleman

The premise of this inference rule is given by X-->>$_{nw}$ Y. The premise as stated is already depicted in table3 and has been already proven. Now, the conclusion part is being considered which is given as

X -->>$_{nw}$ R-Y

Here R corresponds to the entire relation, given by

R= {Course, Lady, Gentleman}

Where X=> Course, V=> Lady and Y=> Gentleman

The conclusion part has an augment R-Y which states that the rule when the complete relation with all attributes is being set differenced with attribute corresponding to {Gentleman}

R-Y= {Course, Lady, Gentleman} – {Gentleman}

⇨ {Course, Lady}

Now, it is to be shown that X -->>XV

Table6. Relation X-->>XV

Course	{Course, Lady}
Latin	{Latin, Racquel}
Latin	{Latin, Lorena}
Latin	{Latin, Flor}
Latin	{Latin, Eve}

V. CONCLUSION

The inference system relating to weak multivalued dependency in its nested form in two different notations are being proposed. First axiomization in fixed universe are being handled which is generally R-Complementary for all schemas from the relation, R. The system was given with axiomization of nw-MVDs and wMVDS in fixed universes. The system is included with R-Complementation rule for MVDs and nw-MVDs. the system represents a challenge for identifying an inference system in which R-Complementation rules are the basic inference rules that hails on basic universes. It is being derived that every inference of MVD being derived, there will be same property being posed in which R-Complementation rule is applied at every last step of the inference but the same application when given with nw-MVDs is not needed to be satisfied. Investigations are been made for axiomization of certain inference rules stated. As multivalued dependencies have no proper ground axiomization, so-called conflict-free embedded multivalued dependencies is finitely axiomatisable. It would be interesting for studying the classes of relational dependencies from the perspective of undetermined universes.

REFERENCES

1. R. Agrawal, T. Imielinski and A. Swami, "Database Mining: A Performance Perspective," IEEE Trans. Knowledge and Data Eng., vol. 5, no. 6, pp. 914-925, Dec. 1993.
2. Abiteboul, S.; Clue, S.; Milo, T.; Mogilevsky, P.; Simeon, J.: Tools for Data Translation and Integration. In [26]:3-8, 1999.
3. Heikomuller, johann-cristoph Freytag, S. Abiteboul, S. Cluet , T. Milo, tools for data transformation and integration, bulletin of the technical committee on data engineering , march 1999, vol.22, No. 1,3-8.
4. Jiawei Han, MichelineKamber and Jian Pei Data Mining: Concepts and Techniques, 3rded. The Morgan Kaufmann Series in Data Management Systems Morgan Kaufmann Publishers, July 2011. ISBN 978-0123814791.
5. Nakata, M. Dept. of Inf. Sci., Chiba-Keizai CollComputer Engineering and Technology (ICCET), 2010 2nd International Conference on (Volume:2), A benchmark for the primary data quality measure, Apr-1992.
6. Christie I. Ezeife, Timothy E. Ohanekwu: Use of Smart Tokens inCleaning Integrated Warehouse Data, International Journal of Data Warehousing & Mining, April- June (2005).
7. G. Grahne. The Problem of Incomplete Information in Relational Databases. Springer, 1991.
8. Israr Ahmed, QAbdul Aziz, dynamic approach for data scrubbing process, International journal of computer science and engineering, vol.2, No.02,2010.
9. Fan, W., Kementsietsidis, A.: Conditional Functional Dependencies for Capturing Data Inconsistencies. ACM Transactions on Data Base Systems (TODS) 33(2) (June 2008).
10. Declarative data cleaning: language, model and algorithms, Helena Galhardas, Daniela Florescu, Dennis shasha, Eric Simon, Cristian-Augustin satia, 27th VLDB conference, Roma, Italy,2001.
11. Shaoxu Song ; Hong Kong Univ. of Sci. & Technol., Hong Kong, China ; Lei Chen ; Yu, P.Ssequential data cleaning approach for a data warehouse, Dec-2009.
12. E. Rahm and H. H. Do. Data cleaning: Problems and current approaches. IEEE Data Engineering Bulletin, 23(4), 2000.
13. N. Koudas, A. Saha, D. Srivastava, and S. Venkatasubramanian, "Metric functional dependencies," in ICDE, 2009, pp. 1275-1278.
14. J. Wijsen, "Database Repairing Using Updates," ACM Trans. Database Systems (TODS), vol. 30, no. 3, pp. 722-768, 2005.
15. C.White. Consolidating, Accessing, and Analyzing Unstructured Data, December2005. Business Intelligence Network article.
16. A. Kementsietsidis, M. Arenas, and R. J. Miller. Data mapping in P2P systems: Semantics and algorithmic issues. In SIGMOD, 2003.
17. Wenfei Fan Inf. Forum, Univ. of Edinburgh, Edinburgh, UK Geerts, F. ; Jianzhong Li ; Ming Xiong, Discovering Conditional Functional Dependencies, Knowledge and Data Engineering, IEEE Transactions on (Volume:23 , Issue: 5), 2011.
18. N. Leavitt, "Will NoSQL Databases Live Up to Their Promis," Published by the iEEE Computer Society, 2010.
19. Yafooz, W.M.S. ; Fac. of Comput. & Math. Sci., Univ. Teknol. MARA, Shah Alam, Malaysia ; Abidin, S.Z.Z. ; Omar, N. ; Idrus, Z., Managing unstructured data in relational databases, Systems, Process & Control (ICSPC), 2013 IEEE Conference, 2013.
20. W. Fan, F. Geerts, X. Jia and A. Kementsietsidis, "Conditional Functional Dependencies for Capturing Data Inconsistencies," ACM Trans. Database Systems (TODS), vol. 33, no. 2, pp. 1-48, 2008.
21. P. Bohannon, W. Fan, F. Geerts, X. Jia, and A. Kementsietsidis, "Conditional functional dependencies for data cleaning," in ICDE, 2007.