Identify Data Dependency in Relational Database: A Recent Survey

Varade Sagar Balu
M. Tech Student
Department of Computer Science & Engineering
Lord Krishna College of technology Indore

Vijay Kumar Verma
Assistant Professor
Department of Computer Science & Engineering
Lord Krishna College of technology Indore

Abstract

Identify Data dependency between the attributes of a database play very important role in the design of normalization. Data dependency identifies value of attributes which are uniquely determine form some other attributes. Identify data dependency is useful for handling large relational database efficiently. Data dependency also helps for a database designer to split a large relational table into several meaning full relations. This process helps to manage the large relational data efficiently and effectively without any redundancy. Identification of these data dependencies form database is difficult task. Several researchers have been proposed various methods to indentify data dependencies form relational data base. In this paper we proposed a general overview and comparison of some of the method.

Keywords: Dependency, Normalization, Relational, Attributes, Database

I. INTRODUCTION

Designing a database efficient is one of the most important tasks in the project development. Physical design of the data base includes data types, indexing, and other parameters related to the database management system. Conceptual schema and logical designs correctness and integrity of the database model[1]. Database designers are aware of specifying keys attributes in the tables and also determining relationships between attribute. Data normalization support database designers to make correct design of the database. With the help of normalization we break unstructured relation into separate relations[2]. The objective of the separation is to remove redundancy and reduce data inconsistency. There are different levels of normalization database designer used as per the requirement of the project. Most the database applications are designed up to be either in the third or the Boyce-Codd normal forms. Figure 1 shows the various level of normalization [3,4].

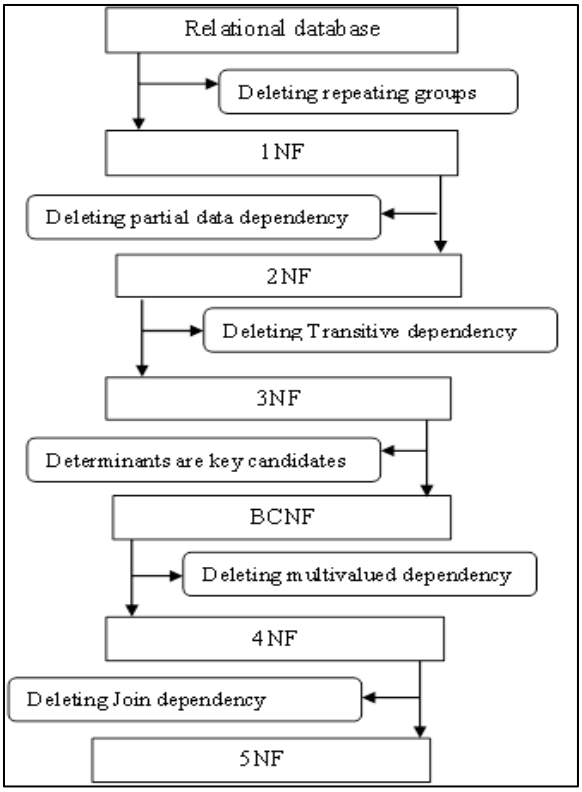

Fig. 1: Steps data dependency

II. BASIC TERMINOLOGY

There are three important axioms are used when to defined data dependency.

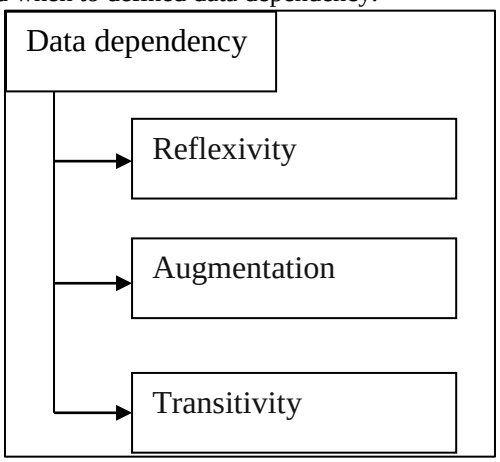

Fig. 2: data dependency axioms

Consider three set X, Y, Z.

1) Reflexivity: - Can be defined for two set X and Y.
   If $Y \subseteq X$ then $X \rightarrow Y$
2) Augmentation: - Can be defined for three set X, Y and Z.
   If $X \rightarrow Y$, then $XZ \rightarrow YZ$.
3) (Transitivity):- Can be defined for three set X, Y and Z.
   If $X \rightarrow Y$ and $Y \rightarrow Z$, then $X \rightarrow Z$.

Above these axioms can be used to defined two new two inference

1) Union: - If $X \rightarrow Y$ and $X \rightarrow Z$, then $X \rightarrow YZ$.
2) Decomposition: - If $X \rightarrow YZ$, then $X \rightarrow Y$ and $X \rightarrow Z$.

III. FUNDAMENTAL APPROACH

There are two basic approaches which are commonly used for mining data dependency in relational data base.

1) Top-down methods
2) Bottom-up methods

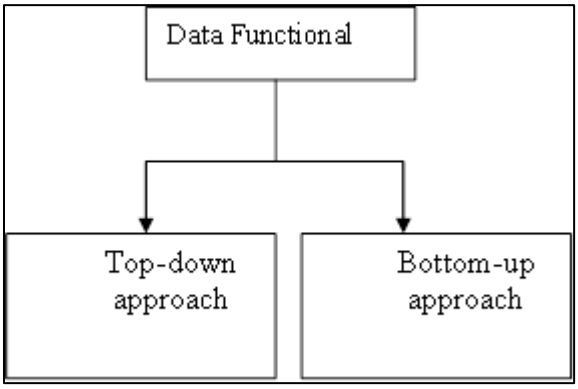

Fig. 3: basic approaches data dependency

IV. LITERATURE REVIEW

We have study some the current research paper related with our topics

In 2010 Y. V. Sreevani, T. Venkat Narayana Rao proposed "Identification and Evaluation of Functional Dependency Analysis using Rough sets for Knowledge Discovery". They explore inconsistencies in the existing databases by finding the functional dependencies extracting the required information or knowledge based on rough sets. They also discuss attribute reduction through core which helps in avoiding superfluous data. Suggested method used to solve problem of data inconsistency based medical domain [4].

In 2011 Nittaya Kerdprasop & Kittisak Kerdprasop proposed "Functional Dependency Discovery via Bayes Net". They proposed a novel technique to discover functional dependencies from the database table. The proposed approach helps the

database designers covering up inefficiencies inherent in their design. Proposed technique is based on the structure analysis of Bayesian network or Bayes net. Most data mining techniques applied to the problem of functional dependency discovery are rule learning and association mining [5].

In 2012 Thierno Diallo and Noel Novelli proposed "Discovering (frequent) constant conditional functional dependencies". They introduced CFD inference. They focused on two types of techniques inherited from FD inference: the first one extends the notion of agree sets and the second one extend the notion of non-redundant sets, closure and quasi-closure. They implemented these technique showed both the feasibility and the scalability of proposition [6].

In 2013 Challa Neehar and T. V. Sai Krishna proposed "Inconsistent Relational Data Cleaning By Detecting Conditional Functional Dependencies" They introduced three methods CFDMINER, CTANE and FASTCFD. CFDMINER is used for discovering constant CFDs; it employs item set mining on both free and closed item sets for constant CFD discovery. CTANE is used for discovering general CFDs. It uses breadth first approach or level wise approach for discovering general CFDs it works well when the data base size is large. FASTCFD is used for discovering general CFDs uses depth first approach for discovering general CFD.FASTCFD works more efficiently than CTANE when the arity of the relation is large [7].

In 2014 Ziawasch Abedjan Patrick Schulze proposed "DFD: Efficient Functional Dependency Discovery". They present a new algorithm DFD for discovering all functional dependencies in a dataset following a depth- first traversal strategy of the attribute lattice that combines aggressive pruning and efficient verification. Proposed approach is able to scale far beyond existing algorithms for up to thousands of tuples, and is up to three orders of magnitude faster than existing approaches on smaller datasets [8].

In 2015 P. Andrew, J. Anish Kumar proposed "Investigations on Methods Developed for Effective Discovery of Functional Dependencies". They give the details about various methods to discover functional dependencies from data. They also give details about discovery of conditional functional dependencies. Proposed works would promote a lot of research in the area of mining functional dependencies from data [9].

In 2015 Thorsten Papenbrock and Jens Ehrlich proposed "Functional Dependency Discovery:

An Experimental Evaluation of Seven Algorithms". They describe, evaluate, and compare the seven most cited and most important algorithms, all solving this same problem. Their descriptions provide additional details. They show that all functional dependency algorithms optimize for certain data characteristics and provide hints on when to choose which algorithm [10].

V. COMPARISON BETWEEN BASIC METHODS

We are comparing three important methods Tane Dep_Miner, and FUN based on concept used for mining dependency

Table – 1
comparison based concept used

Method Name	*Concept used*
Tane	*Proper subset*
Dep_Miner	*Agree set*
FUN	*Free sets*

REFERENCES

[1] Hong Yao, Howard J. Hamilton FD_Mine: Discovering Functional Dependencies in a Database Using Equivalences 2002 Hong Yao, Howard J. Hamilton and Cory J. Butz Department of Computer Science University of Regina Saskatchewan Canada.
[2] Jixue Liu Chengfei Liu "Discover Dependencies from Data - A Review". School of Computer and Info. Sci., University of South Australia.
[3] St_ephane Lopes, Lot Lakhal "Efficient Discovery of Functional Dependencies and Armstrong Relations". Zaniolo et al. EDBT 2000, LNCS 1777. Springer-Verlag Berlin Heidelberg 2000.
[4] Y. V. Sreevani, T. Venkat Narayana Rao "Identification and Evaluation of Functional Dependency Analysis using Rough sets for Knowledge Discovery "International Journal of Advanced Computer Science and Applications, Vol. 1, No. 5, November 2010.
[5] Nittaya Kerdprasop and Kittisak Kerdprasop "Functional Dependency Discovery via Bayes Net Analysis" Recent Researches in Computational Techniques, Non-Linear Systems and Control ISBN: 978-1-61804-011-4.
[6] Thierno Diallo Noel Novelli "Discovering (frequent) Constant Conditional Functional Dependencies" J. Data Mining, Modeling and Management, Vol. 4, No. 3, 2012.
[7] Challa Neehar and T. V. Sai Krishna "Inconsistent Relational Data Cleaning By Detecting Conditional Functional Dependencies" International Journal of Computer Science and Information Technology & Security ISSN: 2249-9555 Vol. 3, No.1, February 2013.
[8] Ziawasch Abedjan and Patrick Schulze "DFD: Efficient Functional Dependency Discovery" CIKM'14, November 3–7, 2014, Shanghai, China. Copyright 2014 ACM 978-1-4503-2598-1/14/11.
[9] P. Andrew, J. Anish kumar "A Survey on Strategies Developed for Mining Functional Dependencies International Journal of Innovative Research in Science, Engineering and Technology Vol. 4, Issue 2, February 2015 ISSN(Online) : 2319 – 8753.
[10] Thorsten Papen brock and Jens Ehrlich "Functional Dependency Discovery: An Experimental Evaluation of Seven Algorithms" 31st September 4th 2015, Kohala Coast, Hawaii. Proceedings of the VLDB Endowment, Vol. 8, No. 10 Copyright 2015 VLDB Endowment 21508097/15/06