

Relational Database Design: A Review

Kunal Kumar

University Department of Statistics and Computer Applications
T. M. Bhagalpur University, Bhagalpur-812007,
India

Sachindra Kumar Azad, PhD

University Department of Statistics and Computer Applications
T. M. Bhagalpur University, Bhagalpur-812007,
India

ABSTRACT

Relational Database Designing and Normalization approach is an essential work for the designing of relation up to higher normal form in the field of Database designer. Database Relation stored information into two dimensional matrix form i.e. rows and column. The task of Normalization is to remove data redundancy, data inconstancy and maintain atomicity within the database relation. Key is an attribute of a relation that identifies the tuple within a database relation, uniquely. This key attribute is generated by applying the closure operations on a given set of functional dependency. Functional dependency of a relation shows the various relations between the attributes and entity relationship diagram shows the prototype of the relation from which designer derived the functional dependency. Relational Database Designing has been spaciouly studied and reviewed by many researchers previously; still the works is going on in this area. This paper review the various researches works done earlier in this field and also explain the various applications of Database normalization.

Keywords

Entity Relationship, functional dependency, multivalued dependency, Full functional dependency, partial dependency, Normal forms, 1 NF 2NF 3NF and BCNF

1. INTRODUCTION

The relational database design methodology is divided the different categories and for each category there is a different tools and technique are mentioned. Database design methodology discussed in this paper is proposed by Connolly and Begg in year 2005. According to them, Relational database design methodology is mainly divided into three phases. The first phase is known as conceptual phase the second is known as logical phase and the last one is known as physical phases design. The conceptual level of the database designing phase deals with the conceptual data model which all is independent form the software and hardware details this level also explains how the relation of the database are related to each other. The second phase of the database design is logical design this phase generally binds of the data of conceptual level on to the logical model for some specific purpose data model, this level of database design is also independent from software and external consideration. The lowest level of database design is focused on hardware and

Deals with how data are actually stored on the physical level [1].

These three phases of database designing gives the systematic way for the development of a good database. The main goal of the conceptual database design is to be to generate a conceptual representation of the given database. The important phases that is involved in this phase is entity relationship modeling phases, which deals properties of entity,

relationship among the entity ,characteristics of the attributes defined on the entity set and key attribute. The second level of database design is logical database design, which describes the information stored in the table in many details, without knowing the physical implementation in the database. This level always deals with the entity set and relationship between the entities set with the mapping cardinality of the relationship. The concept of candidate key, primary key and other keys are also specified in this level [2]. The problem arises with multivalued attribute and many to many relationship are also resolve in this level. The most important section in covered by logical database design is Normalization. Is a technique used to normalize the relation of the database?. The lowest level of the database design is known as Physical database design. Is defines that how the information are actually stored in the database. A physical database model deals with the structure of the table, degree of the table, data type used for each attributes, all the constraints on the tables are applied, default value and various relation between the tables. Converting the entity relationship diagram to table, Converting the entity relationship diagram to foreign key, creating separate table for the multivalued attributes defined in entity relationship diagram, Converting all the attributes into column are also specified in the physical database design [3].

In the year of 1976, Peter Chen developed the Entity Relationship Diagram for the relational database management system. After the Charles Bachman and James Martin have modified and made changes on the original basic Entity Relationship Diagram principles. The visualization of the data and how they are related of each other in term of diagram is known as Entity Relationship diagram. Entity Relationship diagram have mainly five major components. **Entities:** Rectangles is used to represent the entities. An entity is a way defined object or concept about which designer has to collect information and stored it into the database table. **Relationship:** Diamond shapes are used to represent, what are the various possible relationship can be established between the two entities [4]. **Attributes** Ellipse are used to represent the attributes. It generally defines the characteristics of the entity. There are various types of attributes an entity can have such as key attribute, multivalued attribute, simple attribute, Composite attribute etc. **Connecting lines:** The arrow type solid lines are used to show the connection between entities, attributes. **Cardinality:** It generally specifies the relationships and the type of the relationship. It also specifies the minimum and maximum number of relationships and absolute minimum number of relationship. When the column of the table uniquely identify the another column of the same table of the same database, this behavior says that the column of the table is dependent on each other [5]. If a column of the table X uniquely identify the column of the attribute Y of the same table, then it can be says Y is functionally dependent on X .

It can be represented by $X \rightarrow Y$. Functional Dependency always play important role while determining the Key of the given table. The value of key attribute column is always unique value but there are many possibilities that the other dependent column attribute values are not same. The other attribute values are functionally dependent on the key attribute. The key attribute column may be set of attribute or a single attribute i.e. aadhar number. The super key of the table is considered as universal key, it contains a set of all possible valid set of keys. Candidate of a relation considered as minimal super key of the relation. In other words, the set of all possible key which are nominated to become a primary key is called Candidate key. The designer of the database choose any key form a list of candidate key and set as primary key in a table during deigning the database. The rest of the after selecting primary key form the list of candidate key is known as Alternate Key. After selecting a primary key form the list of candidate key it is possible that it may be single attribute or set of attributes. The primary key which is made up of more than one column of attributes is concern as concatenated primary key. The operations that can be performed on the functional dependency to find the set of all possible candidates key are closure operation. The Closure of an attribute is a set of attribute that can be defined as a set of all others possible attributes that can be functionally determined by Closure attribute. Normalization is step by step procedure to achieve highest normal form. The functional dependencies, key of a relation and applying closure property operation are playing major role for the determination highest normal form. Earlier database designer are using file processing system for storing the information. This system does not allow more than one user to access the information at same time, the information contained in these files are redundant, inconsistent and does not have easy access. The database always store information as centrally [6]. The database is a collection of in related information in the form of table. This table where the data is stored is to be normalized. It means that the table should not contain multi valued attribute, redundant data and information stored in the table are concurrent access. The Normalization always comes with stages i.e. 1NF, 2NF, 3NF and BCNF. The first normal says data should be atomic in the table. The second normal says that the non key attribute should be dependent part of the key; it should be fully dependent on the key. The third normal form says that there should not be the transitive dependent on the table. Boyce–Codd normal form is considered as highest level of normal form in the series of Normalization. If there is no non trivial functional dependencies of column on anything other than a superset of a key. Boyce–Codd normal form is strict that 3NF [7].

2. REVIEW OF PREVIOUS APPROACHES

Massimo Villari, Antonio Celesti, Maurizio Giacobbe, and Maria Fazio in 2016, discussed the new technologies in the field of relational database management system moving toward to solve the big data issues with the help of NoSQL and other language. The aim of this paper is to construct a new data model which the extension of previous entity relationship model. The authors of this paper are mainly extending the property and notations of entity relationship model to identify for the big data management. It also opens new way of designing for the hybrid SQL-like/NoSQL database system which is easily manage the data and big data issues [8].

Moussa Demba in 2013. In this paper author have proposed and algorithms for relational database normalization into 2NF and 3NF using their basic definitions in a step by step fashion. Before applying the proposed algorithm which is the first step towards the database normalization, author would like to pre processing on the set of functional dependencies to remove redundant dependencies. The proposed algorithm is used to normalize the database relation up to third normal form by taking into account all candidate keys, including the primary key and other keys. They may solve real world example and can be evaluated on the basis on the proposed algorithms [9].

Henning Koehler, Sebastian Link in 2010. The authors notices that the key of the relational database and functional dependencies defined on that relation are interact over the bags of row than they do over sets of rows. Storage of the bags of row is permitted on some commercial database system. The authors extend the some set of properties of functional dependencies defined on that database relation to bags. The authors also established the axiom of keys and functional dependency on some set of finite ground and, proved that their implication is in strong correspondence with that of goal and definite clauses in Boolean propositional logic. At the end, the authors do not allow any bags that contain any data redundancy or duplication on normal form condition that characterizes that schema. The output start to bridge the gap between the theories based knowledge of dependencies and the reality of database instances in which redundant tuples commonly occur [10].

In 2003, Tauqeer Hussain, Shafay Shamail et al., discuss in his paper that how the process of normalization can be removed from the finite number of steps of database designing. The authors also discuss the various new transformations to entity relation diagram to relation schema for the normalization process. This paper also discuss major problem, that how un normalized relation can be transformed from entity relation diagram. A set of methods is presented which guarantee that it generate a relation schema which

satisfy the highest level of normal form can be followed at the second phase of database designing i.e. conceptual design. The author also discussed that the violation of normal form due to insufficient information and their representation of real world in the corresponding entity relationship model. This main aim of the authors behind this paper is to save the database designer values time and effort [11].

Ashish Kamra and Elisa Bertino in 2011, in this paper, the author of the paper has discussed the response component of intrusion detection system for a DBMS. For issuing a suitable response to an anomalous user request anomalous user request is a responsibility of response component. The authors of the paper also intent of database response policies for specifying appropriate response actions. They also presented a language that is interactive Event-Condition-Action type response policy that makes it very easy for the database security administrator to some specify and accurate response actions for all different circumstances depending upon the anomalous request. The two major issues that are discussed by the authors in this paper are addressed in the context of such response policies are policy matching, and policy administration. For the policy matching procedure, they described algorithms to efficiently search the policy database for policies matching an anomalous request assessment [12].

Wenxiu Duan in 2009, the concept of the key is very important in the designing of the database. The key are used to avoid the duplicate entry in the database and maintain the

atomicity. Without generation of the key knows one can reach at the highest level of normal form. In this paper the author mention the A solution algorithm solving candidate keys of single attribute functional dependency sets based on Boolean matrix. The N Class attributes, sub graph and circuit are not considered in this method. The method is very easy in program realizations. The method can obtain all possible candidate key in the appointed problem field. This method also gives a good structure for finding the candidate key form the complex attributes of functional dependency [13].

Hongbo Li and Li Zhou in 2008 MAC algorithms is considered as fastest algorithm for the computing the closure of as individual set of attributes is improved to determining all possible candidate key need not compute the closure set of the total subset of the database relation. Computing closures of the subsets of any database relation which are closely relevant to candidate keys, namely, which are candidate keys or one part of candidate keys, this simplifies computation. MAC algorithms does keep not only minimum cover for the given set of functional dependency of database relation but also closure s of left set of minimum cover, and facilitate to compute closures of all attributes sets [14].

Amir Hassan Bahmani, Mahmoud Naghibzadeh and Behnam Bahmani in 2008. The authors proposed a method for the automatic normalization of the database relation. Normalization is a designing technique for the relation database relation. There is variety of methods are available for the normalizing the database relation. Firstly, the proposed method produces the dependency matrix and the directed graph matrix. After the generation of the dependency matrix and the directed graph matrix it proceeds towards the 1NF, 2NF, 3NF and BCNF. This method mentioned in this paper also helpful for the determination of the all possible candidate key. This is graph based method form reaching the highest level of normal form [15].

Zhang Yi-Shun In 2008, Determining all possible candidate key from the given set of functional dependency is a basic and complex problem for the database designer. In This paper authors proposed Kargaugh map method to finding the all possible key form the functional dependency. This K-Map method may work as same as K-map in Boolean logic algebra. The key point of the method is to compute the prime implicates for a given logic function. Later settle it by using the property of FD map [16].

The ITS is a tool for the database designer. The main of this tool to make simple the concept of the normalization for the students and readers. The database designer can use this system to understand the concept of the database normalization theory. The generations of all possible candidate key form the set of functional dependency, determining the highest level of normal form. The method proposed in this paper can be expanded the concept of dynamic problem generation and normalize the database table to the higher normal forms like 3NF, BCNF, 4NF and 5NF. The authors are decided to develop an intelligent tutor for the database normalization, which will help us designers of the database to understand the steps involves for the normalization and help them in normalizing the databases [17].

Savita B. Chavan, Dr. B. B. Meshram in 2012. This paper presents the various studies of the database normalization and functional dependencies in very simple and easier form. The first steps towards the normalization are to be to find out list of candidate key form the various dependencies of the table

attribute. The authors of the paper also mentioned the factors that are affecting the database normalization are; the single value is referred as the attribute that cannot be sub divided into sub part. The attribute having set of values is referred as multivalued attribute. The key of a table is an attribute where entire attributes are dependent and many independent and dependent factors are mentioned in this for the normalization of the table. In this paper authors mainly have discuss various types of norm normal form and the fundamental steps for the normalizations [18].

Entity Relationship is a pictorial representation of table, in other hand it can be considered as first step towards the database design. The requirement of the database designer is conceptualized with the help of Entity Relationship to design a effective database table. If the Entity Relationship is converted into database table than it ia easy for the designer to normalize the table. There are various rules and regulations are while converting the Entity Relationship to diagram. If there is a multivalued attribute in a Entity Relationship diagram the create separate table for this attribute and relation with parent table with the help of foreign key. In authors of this paper mainly focus on the conversion of Entity Relationship with binary relationships, ternary relationships, decomposing ternary relationships into its binary relation and then convert it into table for the normalization process. While constructing the entity relationship diagram if there is set of ternary relationship then convert the ternary relationship or represent it using binary relationship. Also find out the degree of the table, cardinality of the table and ratio of the relationship for normalizing the database relation. The work proposed by the authors is applicable for weak entity relation. A relation is said to be weak entity if the set of or attribute are unable to form primary for the given relation [19].

3. MAIN STEPS IN DATABASE DESIGNING PROCESS

Designing a database is mainly involves six steps. In this paper, It's going to discuss step by step approaches for the designing and implement the database. At the end of this section, the readers of this will known about database, what are the advantages of database system over the file processing system? The major steps of a database designing process, relation and the relationship with their various relations, data integrity and many more [20].

3.1 REQUIREMENT ANALYSIS

This step is considered as first step towards the database designing and also collecting data for the designing of the database. In this section the designer of database interviews with the client groups to get the requirement of the client and accordingly the database designer design the database as the client mentions their database function during interview. This step also identifies the application area and the client group. The Reporting part of database, the front end design, the documentation part of designing and many more section are discussed in the requirement analysis phase. The flow of information between various sources through different network, the transaction submitted form various clients at the same time and the frequency of the transaction are also discussed in requirement analysis phase. It also explains the business rules, creating various user view, the information or data flow diagram etc.

3.2 CONCEPTUAL DATABASE DESIGN

In the first step, which is known as Requirement analysis phase all the necessary data are collected from the client and

analyzed. The next step towards the database design is conceptual schema for the database applying the highest level of data modeling. The main aim of this phase is to convert the ideas share by the client to the database designer into entity relationship diagram or class diagram. This level is considered as high level of database design. So, it describes the characteristics of the entity, type of entity whether it is strong entity or weak entity. The various form of the relationship is also described by the conceptual schema.

3.3 LOGICAL DATABASE DESIGN

The logical database design phase is also known as data modeling mapping phase for the set of database relation. The Conceptual database design is seems as input for this phases. The entity relationship models that designer defined in the Conceptual database design are transformed into relation using various computation techniques. The aim of designing the database is to normalize the table and table can be form after the conversation of entity relationship diagram into table. The strong entity is converted into table and key attribute on the strong entity become a primary key of the table. If the entity relationship diagram having multivalued attribute the while converting into table make separate table. If more than one entity in a relationship the design more than one table and apply natural join operation for fetching records from various table.

3.4 NORMALIZATION

The last part of the logical design phase is database Normalization. The main goal is to keep the normalized tables in the database is to stored atomic tuples in relation. When database tables are not normalized the table contains redundant tuple, composite attributes and lots of attribute that not dependent on the key attribute of the table. The term Normalization in the database comes with various form named as first normal form, second normal form, third normal form, BCNF and fourth fifth normal forms. All the forms of the normalization are linked with each. It start with first normal form and the step by step it reaches to higher normal form i.e. BCNF. The functional dependency from table that are derived from various dependencies between the attributes. After applying the closure operation on these derived functional dependencies, It generate the key of the relation. The key of the relation play the major role for determining the highest formal form. A relation is said to be in the first normal form if it does not have composite attribute, multivalued attribute or repeating group. The tuple in first normal form always associated with atomic value. If relation is said to be in the second form, first it must be in the first normal form and there should not be the partial dependency on key attribute. A relation must be in 1NF and 2NF then it can further process for third normal form and there should not be transitive dependency on the key attribute. Boyce–Codd normal form is stricter than 3NF. It is considered as highest level of the normal form

3.5 PHYSICAL DATABASE DESIGN

The level of the database design is seems as lowest level of the designing phase and it defined that how data are actually stored in the database at lowest level. In physical database design level it must be clear that which database management system should be used for the database. Today, Database Designer have variety of database are available in the market and all have different data types and storage capacity and the structured query language for creating and developing the database. The index, the authorization to the user and various permissions are managed by physical database design level.

At the user insert data for the testing of the database, when applying the SQL query it should give correct tuple from the database relation.

3.6 IMPLEMENTATION

This phase mainly describes the purpose to create the database. The application program is ready to access the database from networks. The Data Definition Language are compiled are ready to execute on database server. The application programs are written and execute from remote site. The DDL and DML languages are embedded with the front end application program are ready to execute on server machine where database is installed.

4. CONCLUSION

Relational database design is one of the important design pattern towards the normalizing the database relation. The demand of database design is increasing day by day with the new concept of fast computers. These days Big Data is using for storing and analysis for the huge amount of data. But still, database design is concerned as preliminary stage and a lot of research is needed to handle the complexity and issues in Relational database system. This paper mainly presented a short review of currently existing approaches of Relational database design. Database Designer has investigated Relational database design for various environments. Some database designer Researcher like Tauqueer Hussain et. al. has been discussed the problem that are arises at paramilitary stage of database design i.e. transformation of entity relationship diagram to relation. Due to insufficient information from the real world while designing the entity relationship diagram, the database designer could not reach highest level of normal form. The authors of the paper also present set of rule which must be followed during conceptual database design to avoid un-normalized database design [11].

5. REFERENCES

- [1] Churcher, Clare, “Beginning database design”, Beginning Database Design, page no 1-240, 2007
- [2] Birney, Ewan Clamp et al., “Biological database design and implementation.”, Briefings in bioinformatics, Vol-5, Issue-1, page no 31-38, 2004
- [3] Chilson, David W. Kudlac, Michael E., “Database design”, ACM SIGMIS Database, Vol-15, Issue-1, page 11-19, 1983
- [4] Lightstone, Sam Teorey et al., “Physical Database Design”, Physical Database Design, 10.1016/B978-0-12-369389-1.X5000- 1, 9780123693891, 1860949X, 2007
- [5] Garcia-Molina, Hector Ullman et al., “Database Systems: A Practical Approach to Design, Implementation, and Management”, International Journal of Computer Applications, Vol-49, Issue -4, page 90-107, 2010
- [6] Batini, C.Ceri, S. Navathe, Samkant B, “Conceptual Database Design: An Entity-Relationship Approach”, Database, 9780805302448, 0805302441, page 470, 1992
- [7] Storey, Veda C. Thompson et al., “Understanding database design expertise”, Data and Knowledge Engineering, Vol-16, Issue -2, page 97-124, 1995
- [8] Massimo Villari, Antonio Celesti et al., “Enriched E-R Model to Design Hybrid Database for Big Data Solutions”, Management of Cloud and Smart City Systems 2016.

- [9] Moussa Demba, "Algorithm for relational database Normalization up to 3NF", *International Journal of Database Management Systems (IJDMS)* Vol.5, No.3, June 2013
- [10] Henning Koehler, Sebastian Link, "Armstrong axioms and Boyce–Codd–Heath Normal Form under bag semantics", *Information Processing Letters* 110 (2010) 717–724.
- [11] Tauqeer Hussain, Shafay Shamail et al. , "Eliminating Process Of Normalization In Relational database Design", *IEEE, INMIC*, 2003
- [12] Ashish Kamra and Elisa Bertino, "Design and Implementation of an Intrusion Response System for Relational Databases", *IEEE Transactions On Knowledge And Data Engineering*, VOL. 23, NO. 6, June 2011
- [13] Wenxiu Duan, "Candidate keys Solving Method Based on Boolean Matrix for Single Attribute Functional Dependency Sets", *World Congress on Software Engineering*, 978-0-7695-3570-8/0, 2009 IEEE
- [14] Hongbo Li, Li Zhou, "Use closure of relevant sets of attributes to efficiently find candidate keys", 2008 *International Conference on Computer Science and Software Engineering*
- [15] H. Bahmani, M. Naghibzadeh et al., "Automatic Database Normalization And Primary Key Generation", *CCECE/CCGEI* May 5-7 2008, IEEE
- [16] Zhang Yi-Shun, "Determining All Candidate Keys Based on Karnaugh Map", 2009 *International Conference on Information Management, Innovation Management and Industrial Engineering*, 978-0-7695-3876-1/09, 2009 IEEE.
- [17] N. Mendjoge, Abhijit R. Joshi et al., "Intelligent Tutoring System for Database Normalization", *ICCUBEA*, page 1-6, 2016
- [18] Savita B. Chava, Dr. B. B. Meshram, " Study of Various Normal forms and Functional Dependency", *International Journal of Emerging Technology and Advanced Engineering*, ISSN 2250-2459, Volume 2, Issue 7, July 2012
- [19] S Vimala1, H Khanna Nehemiah et al., "Design Methodology For Relational databases: Issues Related To Ternary Relationships In Entity-Relationship Model And Higher Normal Forms", *International Journal of Database Management Systems (IJDMS)* Vol.5, No.3, June 2013.
- [20] Elmasri, Ramez Navathe, Shamkant B, "Fundamentals of Database Systems", Vol- 28, page 1029, 10.1016/S0026-2692(97)80960-3, 0321122267, 19406029, 2003