International Research Journal of Mathematics, Engineering and IT
ISSN: (2349-0322)
Impact Factor- 5.489, Volume 5, Issue 6, June 2018
Website- www.aarf.asia, **Email** : editor@aarf.asia , editoraarf@gmail.com

DOES THE DATABASE FUNCTIONAL DEPENDENCY AND ITS NORMALIZATION MAKE UNIFORM DATABASEMANAGEMENT IN FUTURE?

Yagyanath Rimal
Lecturer, Computer Science/IT
Pokhara University

ABSTRACT

This review deals with the fundamental concept of database management, the functional dependence of data, and its normalization,toorganize daily recordsmanagement for information processing. The primary purpose is to explain the functional dependence its normalization using various suitable real world example which only meet present demand of database organization. Therefore, this paper presents the different database styles to meet the data management system for future analysis.

Keywords: Database, Management Systems, Relational, Key, Primary, Candidate, N, FBCNF, ER Model.

INTRODUCTION

Today database management systems have become an integral part of all types of work, whether in the management of business records or daily data organization. Data has not only become new thought but it is also the catalyst that is driving organizations to new heights of success (McKendrick, 2017). Similarly, blog atlassystem.com, argued that the database management is developing new trends that focus on security, mechanization, cloud and NoSQL were current trend of data design management. Similarly (Lohr, 2016)explained in New York survey claimed that 80% of the time and effort of a data scientist is dedicated to collecting, cleaning and preparing data for analysis as data sets in various sizes and of a different nature. Similarly recent survey of 16,000 data professionals concluded that the most

common challenges to data science included dirty data (36%), lack of data science talent (30%) and lack of management support (27%) are the essential component of database management. Most data professionals concludes thatthere were five categories of challenges grouped from 20 common changes(Hayes, 2018). Although, data mining is considered as an interdisciplinary sub domain of computer science, which basically apply intelligent methods in order to extract data in a fixed patterns that ultimately used for analysis for future planning is known as data mining. The data collected from different applications Knowledge Discovery in Databases (KDD), often called data mining, which aims to discover meaningful information from large collections of raw data (Sangeeta, 2016). Data analytics will be the closelyintegrated platforms are essential to identify patterns, optimize behavior and detect anomalies in internet of thing (IoT) without human intervention (Markarian, 2017).However the recent trendsof database management focus on join SQL / NoSQL. Therefore databases in the cloud / Platform as a service, automated management,which sets greater focus on security that always requires data stored in predefined patterns so that we could easily meets in the future hazard of data organization (Gates, 2017). (FRY, 1976) The data science has explained sharply increased in size in recent years since 1950 when McGee discusses the successful of generation. Human needs and requirements are increasing in the modern electronic age with the high use of electronic devices and information is stored for future requirements are the side of database management. If we analyze the electronic information storage curve, the use of information by the administration for future analyzes has been much broader per year; which finally makes human life easier than in recent years. The records of storage, organization and retrieval of information and its extraction procedures are the first secondary management for the data scientist. (Peng, 2017)He estimated that the number of mobile phone users in the world is expected to exceed the five billion in 2019. In 2016, it is estimated that 62.9 percent of the world's population already had a mobile phone. The penetration of mobile telephony is expected to continue growing, rounded up to 67 percent in 2019. However, most of the growth of the mobile phone market can be attributed to the growing popularity of smartphones. Around 38 percent of mobile devices were smartphone users.

Similarly database is an organized collection of data typically stored in electronic format that allows you to quickly insert, manage, organize and retrieve data .A present, data scientist are planning better algorithms to get and the future problem of the database. The data cleanup plays important role in data management in the right direction (Shakir Khan, 2012). Although, data scientists are investigating

ways for storing data through quantum physics, too little data connected to the spin of an electron (Pinola, 2017). The rapid advances in storage, communications and processing allow us to move all information to cyberspace (Gray, 2017).Mark Whitehorn, emeritus professor of analysis at university of Dundee, supports this multidimensional, business intelligence (MBI), data storage, hadoops helps bring big data at data warehousing, dimensional and factual tables, data warehouse performance management through R programming (Whitehorn, 2017).Gray also faced the most challenging problem is understanding data that there is no doubt that most of the data will be online soon, since it is economical to store data on computers and data server computers in cloud. Organizing these huge data files so that people can easily convert intoinformation is the real challenge which should be based on data patterns, trends, anomalies and relevant information from a large database is most interesting areas of data management (U.M. Fayyad, 1995).

Similarly, the survey conducted by Gil Press revealed that 57% of the data scientific believe that the cleaning and organization of data is the most tedious and least pleasant task in the data science process and 19% believe that the collection of sets of data is the least pleasant task (Press, 2018). Data becomes information once processed, which includes the identification of data analysis problems that offer the best opportunities for the correct organization of data sets and variables, the collection of large-scale structured and unstructured data from different sources, validate data to ensure accuracy, integrity and consistency, apply models and algorithms to exploit large data stores, analyze data to identify patterns and trends, interpret data to discover solutions and opportunities, communicate results to stakeholders through graph visualization. Therefore, there is a great deal of data that can be analyzed to understand the past or predict the future and support decision making. Although analyzing data by hand takes a long time. It is found that data has become much more prevalent, as business leaders now realizing that a data backed approach is the only way to the clients or investors(Ciupa, 2018).Data scientist's predictive analysis can be as good as the data collected for the stored values. The management of database systems in a systematic way is very necessary for database design to meet future demand.

The data in the table with rows and columns, is known as records designed in an organized way. The organization of the records in the column must be based on other records that must exactly match the organization's policies and the data structures. Suppose a relational table t1 [a] = t2 [a] and t1 [b] = t2 [b] are always uniform, here t1 and t2 are rows of tables and a and b are its column attributes.

Therefore, functional dependency tries to keep the primary keys so that each row of data has a unique relationship with other data from another table, while both dependencies must be of the same mode. Therefore, data management requires a secondary priority for normalization and hence functional dependency. Likewise, data in the spreadsheet is stored in rows and columns which indirectly known rows, and columns are known as single-argument attributes. Where each row in the table corresponds to individual records and contains different attributes that describe the row. If the records of a single table become insufficient, then other records must be stored in the database, so a relational database is required, which can be of different types and models to handle all records. Therefore, a relational database (RDBMS) is a collection of data tables that describe and organize mainly a relational model. Each table must identify a column or a group of columns called the primary key column. Then, only the results of the predictive analysis were assembled to the stored values. The database administrative system maintain of interrelated data and a set of programs to be stored and accessed. The management of technical data systems manager maintain necessary for the operation of data where information history is available in an organized formfor functional dependency should be the only data column that must be unique so that each line has the similar data value. Therefore, data management requires a priority to evaluate functional dependency of all attributes, which contribute relational database collection of tables called the primary key column.

Functional Dependency

Table 1: Student Database

Rollno	Name	Phone	State	Country	Age
1	Rupa	98467329	one	Nepal	21
2	Rupa	98467392	two	Nepal	23
3	Sunita	98467329	one	Nepal	31
4	Bimal	98467320	four	Nepal	22

Functional dependency is a restriction between two sets of attributes in a database relationship which is indicated by an arrow (→) among various data attributes. If an attribute A functionally determines B, then it is written as a → B.For example, employeeid → name means that employeeid functionally determines the employee's name. As another example in a student table database, {studentid, time} → {lectureroom}, the student's identification and time determine the lecture room where the student should be appears on time. The Rollno field has unique values, known as primary key fields, each student has a unique roll number in any classes. The fields have Rollno-> Name, Rollno-> Phone, Rollno-> State, Rollno-> Country, Rollno->Age are various relational properties. The State and Country field still have similar values, so they can have another relationship.

State->Country. The fields in the country also have a similar data value known as a set of functional dependencies. The nearest joint attributes can be determined by the nearest attributes. Therefore, it could be a table format of two relationships like (Rollno) + = {Rollno, Name, Phone, State, Country, Age} with another type relationship whereas (State +) = {State, Country} has another type. The attribute sets of all attributes have a super key here Rollno already assumes a unique value and each student table also has a unique name where, as if no subset of this relationship can functionally determine, all attributes will be candidate key here (Rollno +) is a candidate key whereasRollno and Name are super keys. Thus, the attributes that are part of the candidate key are known as main attributes while the others are non-core attributes. The canonical coverage occurs when any database is updated should verify the functional dependency when changes to the new database system should be restored with the help of the same functional dependency of the original set. Theextraneous attributes occur when functional dependence is extraneous if we can eliminate it without modifying the closure of the set of functional dependencies. Which is mainly based on the assumption that every left side of a functional dependency is unique.

Normalization

Database normalization is the process of organizing database attributes to reduce or eliminate data redundancy (with the same data but in different locations) that unnecessarily increases the size of the database as it grow. The same data are repeated in many places. Incoherency issues arise during insert, delete, and update operations because data is spread across multiple files and files can be in different formats; it is difficult to write new application programs to retrieve the appropriate data. Data redundancy eliminated from Table no 2, so data consistency and integrity are increased so that data is secure and flexible for database recovery. Standardization is the process of organizing data in the database that includes creating and establishing relationship between tables used to eliminate redundancy with repeating a group, not independent of the key, the independent multiple relationship. There are some forms of normalization that eliminate redundancy.

Table no 2:Un-Normalized Database

Roll	Name	Courses1	Course2	Course3
105	Sila	Java	Oracle	Oracle
106	Binod	C++	Java	Computer

First Normal Form

The first normal form indicates that the data is in entity form, which means that the following conditions have been met. If the relationship contains compound and multivalued attributes, it violates the first normal form. If all attributes are in the unique value attribute, which eliminates group repetition, the individual tables that eventually create a separate table for each related data set are configured with a unique primary key for the data set. While the design of the main principle of decomposing the main table into minor is that it sets multiple fields in a single table to store similar data, and each cell contains an atomic number, each column must have a single table header.

Table no 3: 1st Normalized

Roll	Name	Course
105	Sila	Java
105	Sila	Oracle
105	Sila	Oracle
106	Binod	C++
106	Binod	Java
106	Binod	Computer

Table no 4: Student database

Rollno	Name	Phone	State	Country	Age
1	Rupa	98467329	one	Nepal	21
		98456432			
2	Rupaa	98467392	two	Nepal	23
3	Sunita	98467329	one	Nepal	31
4	Bimal	98467320	four	Nepal	22
1st Normalization Form					
Rollno	Name	Phone	State	Country	Age
1	Rupa	98467329	one	Nepal	21
1	Rupa	98456432	one	Nepal	21
2	Rupaa	98467392	two	Nepal	23
3	Sunita	98467329	one	Nepal	31

In the same way table no 4 could be normalized with attributes of Rollno has two phones that can handle the repetition of all the attributes in a separate tuple in the table data. Similarly the table no 2 can be restructure as table no 3 with first normal form by applying the repetition of roll no and name course properties.

Second Normal Form

The records in the relational table must be first in 1st normal form and must not contain any partial dependency. The non-prime attributesother than candidate key create separate tables for sets of values that apply to multiple record. The records of new table haswith a foreign key and primary key. The records should not dependent on anything other than a table's primary key including compound key necessary. If a table of relation R(a,b,c,d) and ab->d and b-

Table no 5: Normalized from Table no 4

Roll	Name
105	Sila
105	Sila
105	Sila
106	Binod
107	Binod
108	Binod

Roll	Courses
105	Java
105	Oracle
105	Oracle
106	C++
106	Java
106	Computer

>c in table relation then ab is prime key so that all values can be determined by them and c and d fields are not prime fields. The table decomposition should eliminate the partial dependency. So the new table R(ab,d) and R(c,d) are in relation with primary key combination while design 2nd normal form Roll with Name and Roll with Course are designed to store all records.

Table no 6: Un- Normalized updated student database

Rollno	Name	Phone	State	Country	Age	Courseno	Course name
1	Rupa	98467329	one	Nepal	21	3	C#
1	Rupa	98456432	one	Nepal	21	4	Oracle
2	Rupaa	98467392	two	Nepal	23	5	PHP
3	Sunita	98467329	one	Nepal	31	2	Python
4	Bimal	98467320	four	Nepal	22	4	Java

Similarly when table no 5 is re-structure on the basis of 2nd normal form to separate two tables are redraw with similar with table no 7 with relational table studentinfo(Rollno,Name,Phone, State, Country, Age) and courseinfo(Rollno,Courseno,Course name).

Table no 7: Normalized with 2NF of Table no 5

Rollno	Name	Phone	State	Country	Age
1	Rupa	98467329	one	Nepal	21
1	Rupa	98456432	one	Nepal	21
2	Rupaa	98467392	two	Nepal	23
3	Sunita	98467329	one	Nepal	31
4	Bimal	98467320	four	Nepal	22

Studentinfo

Rollno	Courseno	Course name
1	3	C#
1	4	Oracle
2	5	PHP
3	2	Python
4	4	Java

courseinfo

Third Normal Form

Roll	Name
105	Sila
105	Sila
105	Sila
106	Binod
106	Binod
106	Binod

Student table

Table no 8 : Normalized Student

Roll	Course	Credit	Marks
105	Java		
105	Oracle		
105	Oracle		
106	C++		
106	Java		
106	Computer		

Registration table

The 3NF checks for transitive dependency (second level dependency of primary key) which eliminates field that's do not

Table no :9 Another database

Name	Faculty	Department
Sila		
Bindo		

Faculty table

depends on the key values that are not records key do not belong in the table. If the component of a group of field apply to move single records put these field in separate table. Suppose R(a,b,c,d) is relational table having ab->c and c->d relation then R(a,b,c) and R(c,d) which are non-triviali.e. a->d has transitive dependency of second level with c column.

The first table no 3 can be decomposed as three above tables when organization needs more attributes to be required. Likewise the above tableRollno->Name,Phone,State,Country,Age which violate Rollno->State,State->Country third Normal form so we may easily re-structure as Student(Rollno,Name,Phone,State,Age) and State (State, Country) tables in separate format for more attributes.

Table no 10: Normalized Student Database

Rollno	Name	Phone	State	Age
1	Rupa	98467329	one	21
2	Rupa	98456432	one	21
3	Rupaa	98467392	two	23
4	Sunita	98467329	one	31
5	Bimal	98467320	four	22

Rollno	State	Country
1	one	Nepal
2	one	Nepal
3	two	Nepal
4	one	Nepal
5	four	Nepal

Referential integrity is a database concept use to ensure the relationship between many tables in database remains synchronized during data modification. R1 can be used to ensure data is class may be helpful in optimizing your database environment and can assist in early detection of errors. The combination of primary key and foreign key constraints can be used to enforce referential integrity of any database. In addition foreign key referencing to primary key can also reference. A uniqueness of constrain help to maintain referential integrity which sets some special reference so that it can easily locate the particular value inside table.

If you want to get data from table 11, it is very difficult to get the appropriate value of c column because this table has no row references. Therefore, special keys are needed to access special data from table no 11. There are several types of keys that help relationships to find a special value in table no11. The field b and c are not the key from which itcannotsearch for special values from the table. Therefore, a->bc and bc-> a are the key, but if c-> a is not a key whose value cannot predict each of the attributes in the table. The super keys are those keys that can contain one or more attributes through which it can easily identify the row and column separately. Suppose that in table R (a,b,c,d) has some functional dependency a-> bcd and ab ->cd, abc-> d, bd-> abc-> to all the others except c-> ab we can easily find all the attributes become super key that could not find the

Table no 11:

a	b	c	d
1	a	x	
2	b	x	
3	b	x	
4	c	x	

value of b. The candidate key is the minimum key, only the super key. If you apply the concept of the appropriate subset of the super key must have minimum values it is known as a candidate key. The bd-> ab attribute is also a candidate key because it has been left in relation. The primary key is a candidate key in a single pass that is easily selectable by the database administrator. It is known as a primary key that must be only one primary key in any table.

Similarly, another relation R (A,B,C,D,E,F,G,H) table where AB-> C, A-> DE, B-> F, F-> GH clearly shows that AB does not have an inbound edge, therefore, AB are essential attributes, when designing tables A and B are always essential attributes through which we can easily find other attribute values in relational table.

The another example of R(A,B,C,D,E,F,G,H) relational table has AB->C, BD->EF,AD->G,A->H relation from relational inbound analysis ABD become candidate key through which it could find all values of attributes. In sometime it is necessary to find out the conditional relationship of attributes through which it could easily find out all attributes. Suppose R(A,B,C,D,E) having relationship AB->CD, D->A and BC->DA here B doesn't have incoming edge but could not be single candidate key therefore it is better to combine AB,BC,BD might be candidate key of tabular relationships.

Foreign key column or combination of columns that are used to establish a relationship between data in two tables. The column used to create the primary key in one table are also used to create the foreign key constraint can be used to reference data in the same table or in another table. The foreign key does not have reference but primary key has reference either the same table or another table. The foreign key table may accept null but if it contain null the reference process is skipped.

BoyceCodd Normalization Form

BCNF is the fourth level of database administration that should be the 3rd normal form. The second focus of the normalization in the partial dependence allows α-> β to be the fields of the relational table in which α is the main candidate key or determining field of β attributes, in this case α is prime and β is a non-prime attribute. The second normal form cannot allow non-primary attributes to be depend on a part of the candidate key α. Similarly to become the 3rd normal form, the first and second normal form must be there and α-> β in this case both αβ must be a non-prime attribute so that non-primary attributes can find other non-primary attributes known as transitive dependency. To become BCNF, both the partial condition and the transitive dependence could not match. Suppose that the first β, which eventually discovers the attributes, indirectly forms the BCNF forms. If R is the highest normal

form for every functional dependency, left hand side is non-trivial functional dependence is super-key. Suppose in relational table R (A,,B,C,D,E) then BC-> D, AC-> BE and B -> C are relation then AC become the candidate key in which A or B cannot be derived from any other attribute, so there is single candidate key (AC). The main attributes are part of the AC candidate key. Similarly, BDE are non-primary attributes, the relationship is first normal and does not allow multiple values or compound attributes. The BC-> D relation is in second normal form in which BC is not configured with the candidate key of AC and AC-> BE is also the candidate key B-> E is 2NFbut B is not the subset set of candidate key AC. The relation is not the third normal form because BC-> D and BC is super key or D is the main attribute. In B-> E, B is super key nor is the main attribute E. To become BCNF LHS, functional dependency must be super-key or RHS must be the main attribute. A relational schema R is in the normal form of Boyce-Codd, if and only if for each of its dependencies $X \rightarrow Y$, at least one of the following conditions is met:$X \rightarrow Y$ is a trivial functional dependency ($Y \subseteq X$) and X is a super key for the R schema.

FUTURE TRENDS IN THE DATABASES

Now is the urgent to look the future perspective of data science that present growing style of data management will meets future of databases design. Currently data scientist are starting to store in complex logic with separation of database itself; This feature allows a database to reside in more than one location and to be queried as a continuous unit which is called distributed or federated databases in data structure. While accessing data sets all intermediate servers run simultaneously in all locations. This is possible due to the increase in network speed. (Berg, 2013) predicts the following trends in the database world. Although data scientist are changes in architecture, such as cloud computing and the need to manage large amounts of data. An example is Google Map Reduce. Data collections, not databases, are increasingly important for acquiring and connecting knowledge. Cloud, mobile and virtual applications are money changers. Data storage becomes increasingly complex which requires the need for specialized functions. It is also necessary to store image data, scanned data and complex medical data, such as gene sequences. Medical record physical data and transform it into a digital format are in use and must be processed by computers. Processing this data to derive models, such as DNA sequences that cause a particular disease, also requires more storage and processing of data. However, (Guy Harrison, 2018) Dell said that the end of a database valid for all dimensions because the single architecture cannot satisfy the request for availability, data volume, and data transaction for

all types of electronic devices are used in the modern world. Web server, web DBMS, memory management (Spark) Hadoop, RDBMS with oracle, Operational RDBMS, SQL SHAS Analytics HANA, and ERP in-house CRM are available for data management application in modern world.

Data variety and speed of data management in the modern world of the revolution of the Bluetooth industry, 3G Wi-Fi, high-quality red, GPS, headphones, speakers, monitors, monitors, tools communication, silent alarm, podcast for the bedroom are modern needs silently maintain through in cloud.it has been used intelligently to access a commutative human being. For the most part, Big Data's personal data is Google, which is maintained by Google File System (GFS) and the server's back-end. In the DIR, Hadoop Map Reduce Managed Stack is the best Big Data technology that uses very important domains in the modern world. The third trend is NoSQL, which preserves the collection of data in several services, the data network and the read pattern only for slaves. Amazon uses constant hashing, uses the family of data in columns where ODBMS and XML use the document database. In the same way, the infinite record uses the basic data records. In summary, the database has several users, as a key value, based on tables, document base and chart base. The fifth trend is the end of the disk, in which we replaced disk storage with analytics, Volt DB, Spark Oracle 12 and cloud computing. Data Analysis is the process of systematically applying statistical techniques to describe and illustrate, condense and recap, and evaluate date. An essential component of data integrity is the accurate and appropriate analysis of research findings.

Therefore, data-driven science, is an interdisciplinary field about scientific methods, processes, and systems, or knowledge or insights from data in various forms, either structured or unstructured, similar to data mining is based on functional dependency and normalization of all tabular records so that future generation would easily processed.

CONCLUSION

However, future of databases appears bright with possibilities when data scientist could design uniform API which could handle and analyzing data huge amount of data system. Electronicdatabase and improving technology has made people all come to click or call away. Distributed transaction processing is becoming the norm for business planning in many countries. Data storing data base management shows variation in storage, size, and its architecture due to large organization needs. This can be made uniform in some extent from using functional dependency a d normalization techniques. This system is useful in data management context of data storage, large data in organized manner.

AUTHOR INFORMATION

Yagyanath Rimal is an assistant professor of Computer Science and Information Systems at Pokhara University,Nepal. He received his in Master of Science in Information Technology from Sikkim Manipal University Health Sciences with Information Systems Specialization he has wide area of interest of design web application, data base management and writing books.

REFERENCES

Berg, K. S. (2013). History of Database. International Journal of Management & Information Systems – First Quarter 2013, 3.

Ciupa, M. (2018). How the History of Data Gathering Lead to the Age of Big Data.

FRY, J. (1976). Evolution of Database Management System. 2.

Gate, R. (2013). International Journal of Management & Information Systems Volume 17, Number 1. Kristi Berg,, 2.

Gates, P. (2017, 3). https://www.datavail.com/blog/4-top-trends-db-management/. Retrieved from https://www.datavail.com/blog/4-top-trends-db-management/: https://www.datavail.com/blog/4-top-trends-db-management/

Gray, J. (2017). Data Management: Past, Present, and Future . Microsoft Research IEEE Computer 29(10): 38-46 .

Guy Harrison. (2018). Dell Software Group. Retrieved from Dell and Quest – a brief history.

Hayes, B. (2018). Analytics, Data Science, Machine Learning.

Lohr, S. (2016). New York Times.

Markarian, J. (2017). The Data Advantage. Retrieved from https://www.infoworld.com/article/3245644/technology-business/7-technology-trends-to-watch-in-2018.html

McKendrick, J. (2017). Hot Data Technologies and Trends for 2018. Retrieved from Hot Data Technologies and Trends for 2018: http://www.dbta.com/Editorial/Trends-and-Applications/Hot-Data-Technologies-and-Trends-for-2018-122280.aspx

Peng, D. (2017). Retrieved from https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/

Pinola, M. (2017). Data Storage Technologies of the Future. Retrieved from backblaze.com: https://www.backblaze.com/blog/data-storage-technologies-of-the-future/

Press, G. (2018, 3 7). Retrieved from https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/2/#5d5ee48f1492

Sangeeta, G. a. (2016). DATA MINING TREND IN PAST, CURRENT AND FUTURE.

Shakir Khan, D. A. (2012). Data Mining For Securiy Purpose. (I. 2277, Ed.) INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH , 2.

U.M. Fayyad, G. P.-S. (1995). Advances in Data Mining and Knowledge Discovery,. Cambridge: MIT Press.

Whitehorn, M. (2017, 3). Get six expert tips for improving data warehouse performance. Learn how database engines, SSDs and MOLAP cubes can affect your data warehouse performance. Retrieved from http://searchdatamanagement.techtarget.com/answer/Six-tips-for-improving-data-warehouse-performance: http://searchdatamanagement.techtarget.com/answer/Six-tips-for-improving-data-warehouse-performance