

Analysis and Design Complex and Large Data Base using MySQL Workbench

Dedi Iskandar Inan¹ and Ratna Juita²

¹Department of Computer Engineering, Papua State University, Indonesia
¹dediiskandar04@gmail.com

²Department of Computer Engineering, Papua State University, Indonesia
²r4tna_04@yahoo.com

ABSTRACT

This study will be demonstrated on how to conduct analysis and design in building a large and complex data base. The data in this study originated from the Attachment to the Decree of the Minister of Manpower and Transmigration (Kepmenakertrans) No. 250/MEN/XII/2008, namely the attachment of AF, which consists of 800 tables of data. Due to the large data base which will be formed because so many tables involved then the process is required - the normalization process (functional dependence, Normal Normal First, Second, Third and Boyce-Codd) in forming a good database that can later be used for purposes further, for example, form a management information system and so forth. In the process DBMS MySQL and MySQL Workbench is a tool used to facilitate this process in shaping the Entity relational diagram. By utilizing these tools will help data base engineer in designing a good and normal data base. It is expected that the formulation of the data base will be used as a standard in designing applications of manpower information systems across the provinces in Indonesia.

KEYWORDS

DBMS, MySQL, MySQL Workbench, Large Database, Normalization.

1. Introduction

1.1. Definition

Before discussing anything further about the database, here are some excerpts of what the data base. According to Wikipedia [1], the database is a collection of information stored on the computer in a systematic way so that it can be checked using a computer program to obtain information from the database. Silberchatz et al [2] described the database as a set of data that contains information on an entity. And according to Powell [3], the database is an implementation or creation of physical database on a computer. While Sumathi [4] described the database as a collection of data that relate to and have a certain meaning, which is stored properly and can be accessed in various ways and sequences.

From descriptions above then in general it can be concluded that the database is a collection of data items that are interconnected with one another which is organized by a particular scheme or structure, stored on computer hardware and software to perform manipulation for a particular purpose, for example in the system that tree/multiple-tier client-server architecture. Therefore, the database is built in a certain way, namely the model database to facilitate the user in this case is a human being as an end user to use them easily.

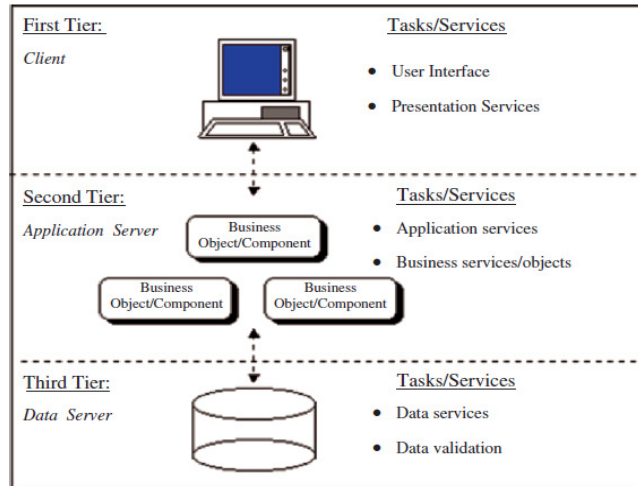


Figure 1. Client- server/tree-tier architecture [4].

Hence, it is required a management system to store and to access the data or information from the database, called DBMS (Data Base Management System).

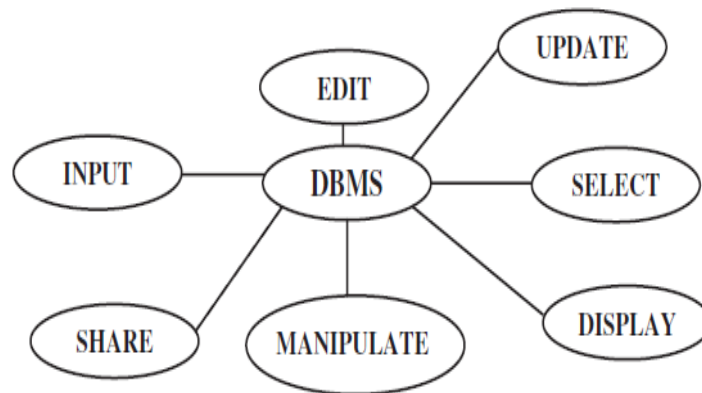


Figure 2. DBMS Capability [4].

1.2. The Important of Database

Currently almost all levels of human life have made use of data bases, for example in the fields of business, education, law, health, sports, culture and so forth. Field - the field with various subfields below have used the database to store all sorts of data, either directly or indirectly, related to their daily work [4], for example in the field:

1. Banking: for storing customer information, account, transaction, credit, and so forth;
2. University: for storing information of students, faculty, staff, value, rank, etc;
3. Finance: to store information on sales, purchases, companies, bonds, etc;
4. Human resources: to store information about employees, salaries, taxes, insurance, etc;

All of which are stored in a DBMS has been ease of man as the end user to use them. Some of the reasons for the importance of the database are as follows:

1. The database is one important component in the system information, because it is fundamental in providing information.

2. The database determines the quality of information: accurate, timely and relevant. Information can be said to be worth when the benefit is more effective than the cost to get it.
3. The database reduces duplication of data (data redundancy).
4. By applying database data relationships can be improved.
5. The database can reduce waste reservoirs outside.

Therefore, the database design is essential so that we can have a database that is effective in design, efficient in processing, easy to manipulate. It is very highly supported because of the cheapening of the price of storage media.

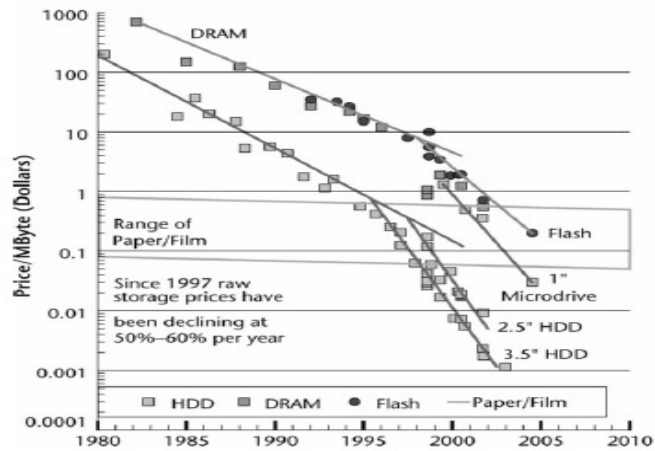


Figure 3. Price trends of data storage media [5].

2. Design Methods Database

2.1. Relational Database

The study of how to model databases in a graphical form which can be easily understood by humans was first proposed by Chen [6] who later called the Entity Relational Diagram (ERD) is the base model that describes the objectivity of the real world that consists of the objects base called entities and relations between objects.

Because ERD diagram is a model that should be built first in the process of database design to bridge between business objectivity of data available on the file system and data on the real world before it made its data base in the DBMS, it is necessary step - a step in the process, as following [3]:

1. Analysis controlling needs, namely the need to collect data, determine the type and size of the data and determine the metadata or data about data. The process can also be done by conducting interviews with users of such data.
2. Building the ERD of the data including the form tables including the attributes, relationships between tables and normalization process.
3. Design of logic, which is forming the basis of data on the physical stage by writing down the database using the syntax of SQL (Structured Query Language) on stage DDL (Data Definition Language), i.e. create, modify and delete tables.
4. The design of the physical, i.e. write and store data bases that have been processed in stages - stages prior to be stored in a DBMS.
5. Optimization of the data base, which form the index is good and right, change the table engine if necessary and so the aim is to establish a data base can be used effectively and efficiently.

Figure 4. Data on real world examples.

2.2. Normalization

After a needs assessment process and obtain information then the next step is to build the ERD. But there are stages in the process of normalization to be followed. This is necessary because a good data base, which was formed later, must meet the following criteria:

1. Database structure (tables - tables and the relationships among the table) that is more proportional.
2. The structure of each tables need to be made systematically and efficiently.
3. Needs a more efficient storage space. Because of the smaller size of the table it will be more quickly process the data base that will be done.
4. In a relational database, data redundancy is unavoidable but should be kept to a minimum. This will improve data integrity.
5. There is no ambiguity of data in all tables in the database.

However, the normalization process is not entirely a best solution that can be used to form a good data base, for example in the data warehouse [3].

2.2.1. Fuctional Dependency (FD)

In forming the ERD then the designer must have been getting as much information as possible. Examples of information obtained for use in the process of database design is a receipt, tables data on the file system and so forth. To shape it into a table for use in the process of making the rules FD ERD should be used [2] [3].

For example, given a table T that contains at least 2 (two) attributes A and B, can be described as follows:

Table 1. Tabel T

Atribut_A	Atribut_B

Then the attributes that can be formed into:

$$A \Rightarrow B \tag{1}$$

International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 5, Oct 2011 which means that the functional A determines B, or B is determined by A. For that purpose, in the table T there is, at least, 2 (two) values are the same if A then B will also be the same value [2] [3].

2.2.2. Boyce-Codd Normal Form (BCNF)

To get tables in the design process can then apply the rules of BCNF. From equation (1) above can, using the BCNF, be said that the attribute A must be super key. If not then the table must be composed again by following simple rules of BCNF [2] [3] [4] [5].

2.2.3. First Normal Form (1NF)

Form of normalization phase I met if a table is no longer has many attributes or attribute value must be atomic. That means that these attributes must be the smallest element of an entity that cannot be broken again [2] [3] [4] [5]. The following table will explain it:

Table 2. Schedule

Mata Kuliah	Waktu	Tempat	Dosen
Basis data II	Senin, 08.00-12.00	Elearning 1	Ratna Juita, ST, MT
	Selasa, 13.00-17.00	LabKom	
Pemrograman Web I	Senin, 08.00-12.00	Elearning 1	Dedi Iskandar Inan, ST, MT
	Rabu, 13.00-17.00	LabKom	
Bahasa Inggris III	Kamis, 13.00-17.00	Aula FMIPA	Denny Unggul, ST

In order for the table to meet the 1NF form the table should be decomposed to follow the rules KF, into 2 (two) tables as follows:

Table 3. Lecture

kode kul	nama kul	sks	semester	nama dos
MKK123	Basis Data II	3	3	Ratna Juita, ST, MT.
MKD234	Pemrograman Web I	3	3	Dedi Iskandar Inan, ST, MT.
MKK345	Sistem Operasi	3	2	Tedy Aripianto, ST.
MKU432	Bahasa Inggris III	2	3	Denny Unggul, ST.

and

Table 4. Schedule

kode kul	Waktu	Tempat
MKK123	Senin, 08.00-12.00	Elearning 1
MKK123	Selasa, 13.00-17.00	LabKom
MKD234	Senin, 08.00-12.00	Elearning 1
MKD234	Rabu, 13.00-17.00	LabKom
MKU432	Kamis, 13.00-17.00	Aula FMIPA

2.2.4. Second Normal Form (2NF)

A table is said to have met the normalized form of the second phase if it has to meet 1NF and all non-key attributes of primary (non-primary key) has the KF on the primary key attribute (primary key)[2][3][4][5].

2.2.5. Third Normal Form (3NF)

The third form of normalization has the following stages: (a). Meets the second phase of normalization (2NF) and (b). Every non-key attributes must not be dependent on other non-key attributes [2] [3] [4] [5].

2.2.6. Degrees of Cardinality

The degree of cardinality is a level that describes a number of entities that are interconnected through an associative relationship.

There are 4 (four) types of relationship cardinality:

- (a). One to one, states that one entity in A is associated with only one entity in B and vice versa.
- (b). One to many, states that one entity in A is associated with more than one entity in B but not vice versa.
- (c). Many to one, the same as in (c) but reversed from B to A.
- (d). Many to many, states that one entity in A is associated with at least one entity in B and vice versa.

In the diagram, can be described as follows:

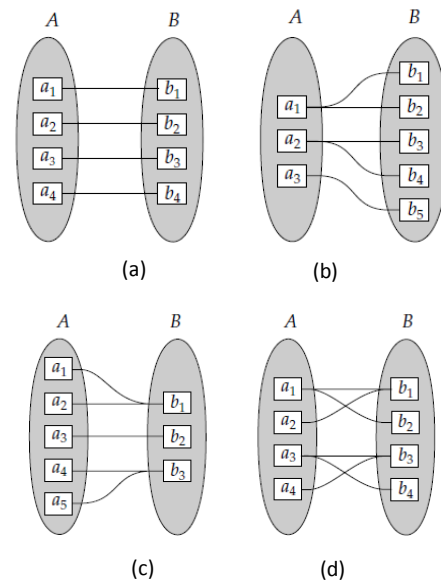


Figure 5. Degrees of cardinality

3. Research Methodology

Implementation of this process is done by using a database (DBMS) version 5.0.51b MySQL, MySQL Workbench Version 5.2. CE and use the operating system Windows 7 on an Intel Core™ 2 Solo 1.4 GHz, 4 GB DDR3 RAM, 800 MHz FSB.

Selection of MySQL because the MySQL DBMS has several advantages [7]:

1. Speed; Benchmarking it can be seen <http://dev.mysql.com/techresources/benchmarks/>, the performance comparisons of several existing DBMS.
2. Easy to use.
3. Support standard SQL Query Capability.

4. Level of connectivity and high security Portability.
5. Small Sized.
6. Pricing and Availability (Total Cost of Ownership is very small).
7. Code - the code is open source and distribution.

To facilitate the data base designer in establishing the ERD is an assistive MySQL Workbench software used in this study. Therefore the applications were also built by the same developer and are dedicated to working with the MySQL DBMS compatibility and excellent interoperability. Both applications can be freely downloaded at the MySQL official site [7].

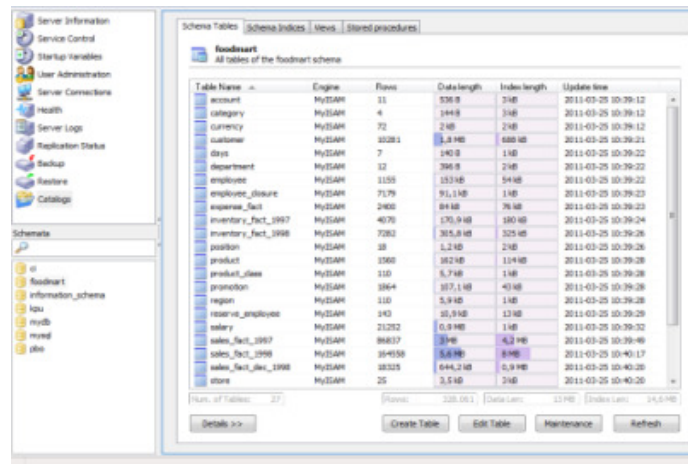


Figure 6. MySQL GUI Administrator



Figure 7. MySQL Workbench

3.1. Sources of data

Real data on the process of this study were obtained from Appendix A-F of the Decree of the Minister of Manpower and Transmigration Republic Of Indonesia (KEPMENAKERTRANS) No.250/MEN/XII/2008 about classification and characteristics data. This appendix shape table consisting of more than 800 tables, which should be used by the Department of Manpower and Transmigration in local and central government to get information about employment in Indonesia.

3.2. Design Process

Functional dependency

At this stage, tables that already exist can be combined or separated by seeing his FD. For example, given a table of appendix A part of education, as follows:

Table 5. Education by sex

Pendidikan	Jenis Kelamin		Jumlah
	Laki-laki	Perempuan	
≤SD			
SMTp			
SMTA Umum			
SMTA Kejuruan			
DIPLOMA I/II/III/AKADEMI			
UNIV			
Jumlah			

Sumber :

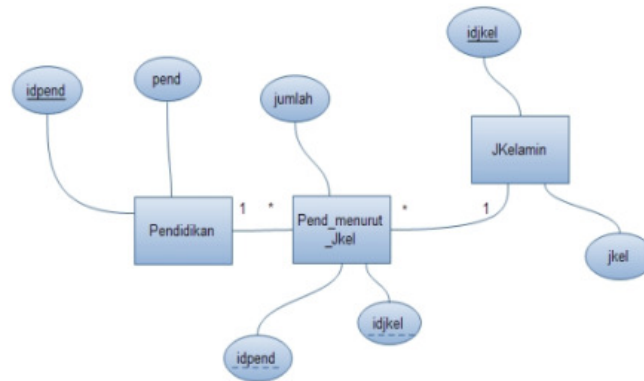


Figure 8. ERD of table 5.

From the table shows that education and gender attributes for which data will be a string into a separate table that consists of:

- Education = { idpendidikan, education }
- Gender = { idkelamin, sex }
- Pendidikan_menurut_jeniskelamin = { education, gender, number }

First Normal Form (1NF)

Of the 3 (three) the result tables decomposition of the main tables above shows 1NF table has been filled for Education and Gender, but not for Pendidikan_menurut_Jeniskelamin table. Therefore there still exists redundancy data in that table then be decomposed again, becomes:

- Education = { idpendidikan, education }
- Gender = { idkelamin, sex }
- Pendidikan_menurut_jeniskelamin = { education, gender, number }

Second Normal Form (2NF)

From the tables above it can be determined from the primary key of each table as follows:

- Education = {idpendidikan, education}
- Gender = {idkelamin, sex}
- Pendidikan_menurut_jeniskelamin = {education, gender, number}, is a relation between the tables of Education and Gender.

From Figure 8 shows that Pendidikan_menurut_Jeniskelamin table is a table of the relationships among the tables of Education and Jeniskelamin. The new table is formed because the degree of relation cardinality is many to many.

Third Normal Form(3NF)

From the tables that formed in 2NF, it has been seen that tables also meets 3NF so that 3NF stage have been met.

3.3. MySQL Workbench

To apply the ERD that form on the physical stage, an application tool used to facilitate this, the MySQL workbench.

With the help of this application then the process of forming the ERD can be directly applied to the physical design stage. This course will facilitate the database designer in designing the database and bring the results of the draft on the physical stage.

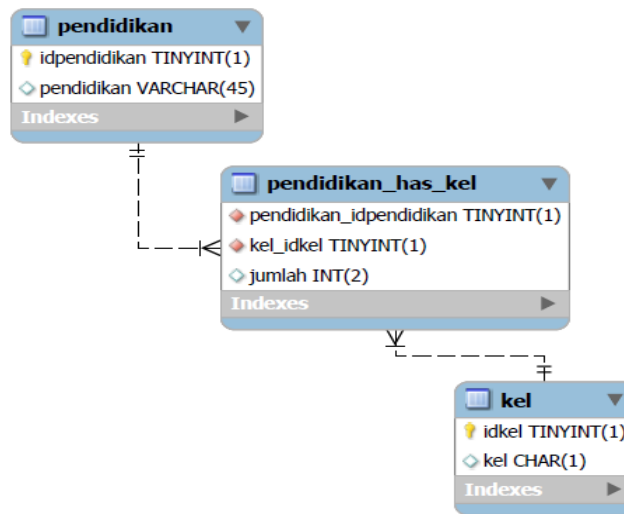


Figure 9. The design of the MySQL workbench

Table `pendidikan_has_kel` is many to many table relationship between tables education and tables kel and therefore the relationship is forming a new table.

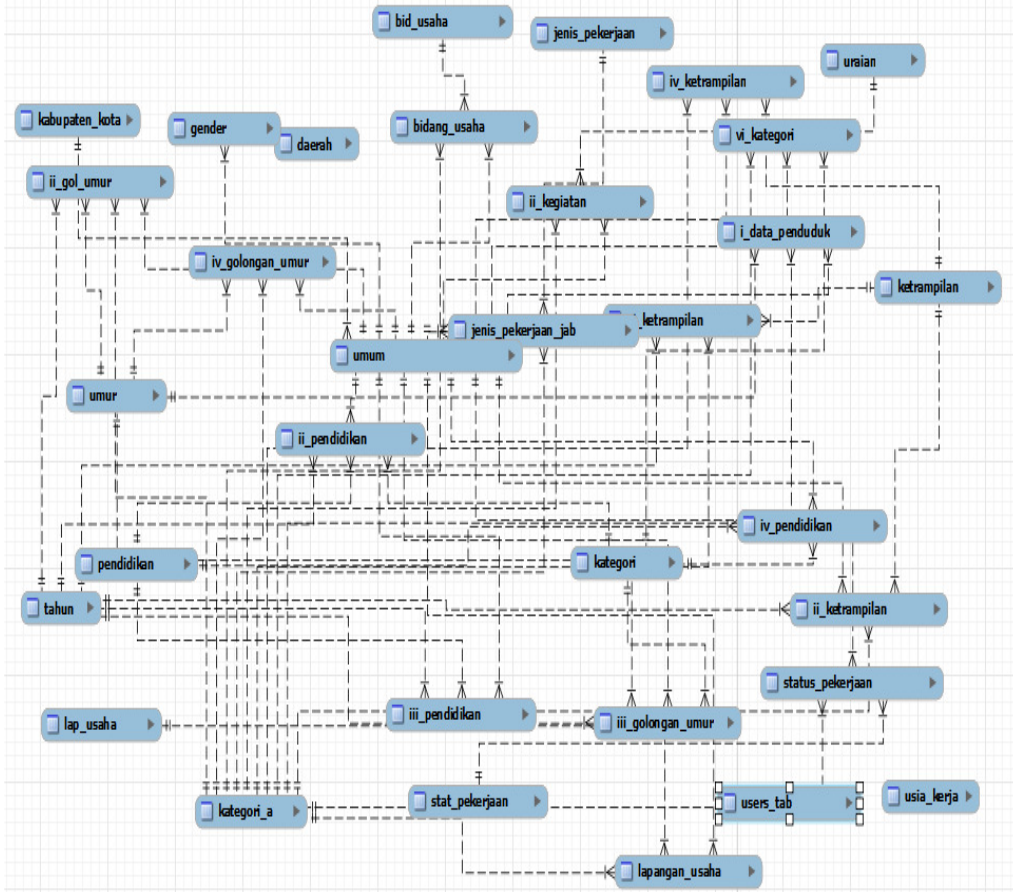


Figure 10. Normalized database based on Kepmenakertrans No. 250/MEN/XII/2008

From Figure 10 shows that more than 800 tables in appendix A to F Kepmenakertrans No. 250/MEN/XII/2008, have been normalized so there are only 33 tables form as a result of normalization process.

4. Conclusion

From this study it can be concluded that the normalization process is very important in the design process that must be considered by the designer of the database. This process must be done before the database was set up in the physical design stage. It is intended that can be later formed the atomic table, in other words, tuples/ records data that formed the atomic data.

Use of design tools, in this case is the MySQL workbench, helps designers build good, efficient and fast database.

References

- [1] Silberschatz-Korth-Sudarshan., (2001) *“Database System Concepts”*, Fourth Edition, The McGraw-Hill Inc.
- [2] Powell, Gavin., (2006) *“Beginning Database Design”*, Wiley Publishing, Inc. Crosspoint Boulevard.

- [3] Sumathi, S. and Esakkirajan, S., (2007) “*Fundamentals of Relational Database Management Systems*”, Publisher: Springer-Verlag Berlin Heidelberg.
- [4] Lightstone, Sam., Teorey, Toby., And Nadeau, Tom., (2007) “*Physical Database Design: The Database Professional's Guide to Exploiting Indexes, Views, Storage, and More*”, Publisher: Morgan Kaufmann.
- [5] Chen, P.P., (1976) “*The Entity-Relationship Model: Toward a Unified View of Data*”, ACM Transactions on Database Systems. 1976.
- [6] <http://www.mysql.com> (Verify access: March 29, 2011)
- [7] http://id.wikipedia.org/wiki/Basis_data (Verify access: March 25, 2011)

Author

Dedi Iskandar Inan obtained his master degree from University of Gadjah Mada in 2006 with Cum Laude Graduation. His research interests include Data Base, Data Ware house, Data Mining, Information System, Web Technology and Networking. He is highly involved in the development of data base and information system, especially web based, in Papua and Papua Provinice in Indonesia. Education, Heath and also Manpower web based information system in Papua and West Papua office Provinces are few among web based information system under his supervisons. He also has publised many journal and proceeding in web based technology topics.



Ratna Juita obtained his master degree from University of Gadjah Mada in 2007. Her research interests include Data Base, Data Ware house, Data Mining, Information System. She also has experiences in publishing journal in this area. She got bachelor degree from Diponegoro University, Semarang, Indonesia 2003,

