PROCEEDINGS
ISBN: 978-602-98756-9

TEST OF DATA QUALITY OF THE DATABASE OF MANAGEMENT INFORMATION SYSTEM OF COKROAMINOTO PALOPO UNIVERSITY

Nirsal[1], Nahrun Hartono[2]

Department of Computer Engineering, University of Cokroaminoto Palopo[1,2]
(nirsal_ftkom@uncp.ac.id[1]), (nahrunhartono@gmail.com[2])

Abstract

Information system can not be separated with database. What is shown by system is the result of data processing in database. The needs of quick and flexible information make humanfind new things of data processing. They have been minimalizing paper use and change it with digital data saved in one computer as data centre. Now, most of system information already use Relational Database Management System (RDBMS). Compared with office application such as Microsoft Office Excel, RDBMS has spesial advantage, but itneeds maximum effort to implement it. The information generated by information system relies on existing data in the database, so that the quality of the data in the database is very influential.This research was conducted to test the quality of the data from the database Management Information System of Cokroaminoto Palopo University is the data of students and lecturers. This study uses data of students and lecturers taken from the student and lecturer tables inside database Management Information System of Cokroaminoto Palopo University. The test results indicate that the data quality of the students and lecturers data does not reach 100%.
Keywords: *Data Quality, Database.*

Background

Information System is one of the basic necessities for an institution particularly educational institution, information system is developed to help fulfill information necessity for the institution. Nowadays almost all of information system use *Relational Database Management System* (RDBMS) or it is often called database, generally RDBMS uses oracle, MsSQL Server, PostgreSQL, Db2, FirebirdSQL or MySQL(Raharjo, 2012). The necessity of quick and flexible data access is very important thing in this modern era where all people's need are digitized,the people have minimalized the use of paper and it is replaced by digital data saved in computer as data centre(Utami & Raharjo, 2006). Saving by using database system or RDBMS certainly has its own advantage, but in its implementation it absolutely needs more effort, it is different from the storing of data by using office affair application such as microsoft excel. Information system in educational institution is intended to help the process of service particularly academic service. This system is designed in such manner in order to be able to give information needed by its user. The role of database in academic information system is certainly as mediation to save information needed by the information system, so the database in academic information system should be designed seriously as good as possible, because the database saves important data such as data of students, lecturers, and other academic data.

Databaseis data collection which commonly describes activity and its user in an organization, for example academic database contains: students, lecturers, lectures, etc. By using database, so the information is saved in long time period, it enables to conduct backup regularly, it enables to store data in large quantity, it can be accessed together and it is easy to integrate with other application (Katsurai, Ohmukai, & Takeda, 2016). Database in its designing has its own rule and it uses its own programming language in operating. SEQUEL (*Structured English Query Language*) or often called SQL is

programming language which has special aim and it is designed to manage data in *Relational Database Management System* (RDBMS), or to manage in the flow of data in *Relational Database Management System*(RDBMS). SQL has three main parts, they are: defining data (*Data Definition Language*-DDL), manipulating and accsessing data (*Data Manipulation Language*-DML) andpart which is used to control the user (*Data Control Language*)(Darmanto, 2015). Table, column and*record*are the important elements in database,the three elements can not be separated one another, andit is possible there are many tables in a database, thre are many columns in table andthe columns contain *record* (data)(Hartono, Utami, & Amborowati, 2016).

There are purposes of using database, namely primary and secondary purpose.Primary purpose of database covers the use of database by many users, keep the intellectual investment, press the fee, omit the proliferation, performance, clarity, the ease of using, flexibility, the need of data which is not anticipated can be fulfilled quickly,easy changing, accuracy, as well as availability. While the secondary purpose of database covers the logical data independency, the fast access, the rapidness of searching, the standardization of data, the availability of data dictionary,inter program front in high level, language of end user, integrity, fast recovery from failure, the ease of arrangement (*tunning*), designing and controlling of tools, as well as reorganization of data can be conducted automatically (Martin, 1975). Beside that, it is cecessary to give attention to the quality of data. The quality of data is the measurement of data accuracy which fulfills the business need and supports in taking decision, there are some aspects to measure the quality of data, namely aspect of accuracy, completeness, derivation integrity, validity, nonduplicate(Hegadi & Manjunath, 2013). Information is needed in taking decision in organization, informationcomes from the collection of data so the quality of data will certainly affect toward the improvement of organization quality.

Method

This research is experimental research. The database used was database of Management Information System of Cokroaminoto PalopoUniversity (SIMUNCP)as a basis to design some tables for testing.From the beginning observation conducted, it showed that the database had 105 tablesand used MySQL as its *Relational Database Management System*. The quality of data tested at the database was aspectof non duplicate. The quality of data ofnon duplicateaspect is measurement ton ensure that there is *one to one*relationship between record and the real object. Datatested its accuracy level were student data saved in student table and lecturer data saved in lecturer table. The testing was conducted by executing query, query executed produced the number of data.The number of data resulted from the query execution was taken as a basis to measure the quality of data in database of Management Information System of Cokroaminoto Palopo University (SIMUNCP).

Results and Discussion

Management Information System of Cokroaminoto Palopo University (SIMUNCP) is web base application created with HTML (*Hyper Text Markup Language*) and PHP (*Pre Hypertext Processor*) programming language. The application is implemented at network of *Local Area Network* (LAN) which means that the application can be accessed only in Cokroaminoto Palopo University area. SIMUNCP uses MySQL as its database management system, in SIMUNCP database there are 105 tables, with 2104 columnsspread in several tables.

The quality of data in database is important thing because the data will produce information, the information will be a basis in taking decision in organization.In this research, the test of data quality is conducted only on non duplicateaspect of student data which is saved in student table and lecturer data which is saved in lecturer table, and it is

conducted by executing query SELECT without clause which will produce a number of data in table and column selected namely student and lecturer table shown at picture 1 and query SELECT with clause GROUP BY which will produce a number of real data shown at picture 2.

```
SELECT *
FROM
      Student
```

```
SELECT *
FROM
      Lecturer
```

Figure 1. Query SELECT without clause

```
SELECT *
FROM
      Student
      GROUP By NIM
```

```
SELECT *
FROM
      Lecturer
      GROUP By NIP
```

Picture 2. Query SELECT with clause

Picture 1 is *query*used to test the quality of data non duplicateaspect of student data where student data is saved in student table. In student table, it is conducted *query*test with clause GROUP BY id,nim, which group the same nim data, with assumption that student id should not be duplicated.While picture 2 is *query*used to test the quality of data non duplicate aspect of lecturer table where lecturer data is saved in lecturer table. In lecturer table, it is conducted *query*test with clause GROUP BY in id column,nip column which means that if there is data duplicated in lecturer table, so it will be grouped based onnip column. The selection of nip column with assumption that there is no the same NIP.

Eery query operated on database will produce a number of data, the result of query execution on both of tables can be seen at table 1

.

Table 1. The result of Query Execution on student and lecturer table

Name of Table	Result of Query	
	SELECT	SELECT GROUP BY
Student	20108	19360
Lecturer	516	511

The test of quality of datanon duplicateaspect is conducted by comparing the result of execution of query SELECT without clause with query SELECT with clause, the graphic of the comparison result can be seen at picture 3.

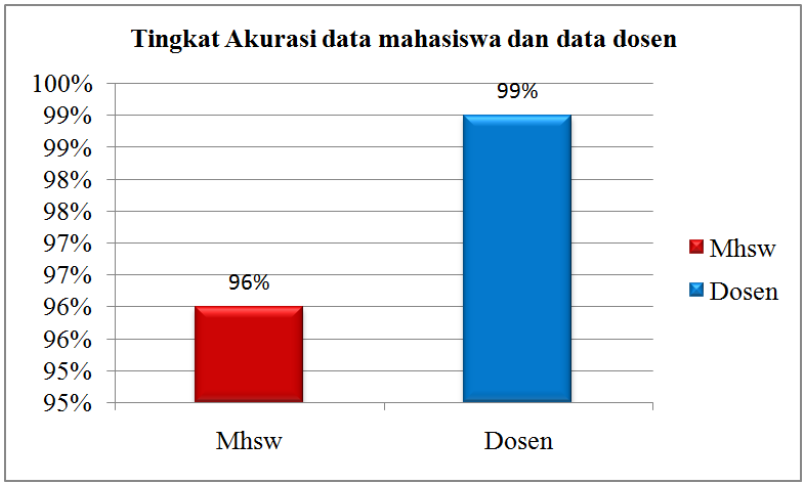

Figure 2. The level of accuracy of student and lecturer data

The graphic of picture 3 can be found out that the quality of data non duplicateaspect for student data which is saved in student table is about 0.96 or with percentage 96% which means that 4% data in the table is duplicate,and the lecturer data which is saved in lecturer table is 0.99 or with percentage about 99% which means that there is 1% from the total of data in the table duplicate.

Conclusion

From the result of the test, so it can be found out that about 4% of student data is duplicate andabout 1% of lecturer data is duplicate, the number is certainly not too significant but student and lecturer data in database have important role in database of SIMUNCP. Student and lecturer table saving student and lecturer data will relate to other tables, for example score table, schedule table, KRS table, etc.It means that if there is duplicate of student and lecturer table, it will certainly affect the quality of other table data so it will affect the information resulted by information system.

References

Darmanto, E. (2015). Analisa Optimalisasi Bahasa SQL Berdasarkan Relational Algebra pada Kasus Rekapitulasi Mahasiswa Layak Wisuda. *Jurnal SIMETRIS, 6* (2), 405-414.

Hartono, N., Utami, E., & Amborowati, A. (2016). Migrasi dan Optimalisasi Database Sistem Informasi Manajemen Universitas Cokroaminoto Palopo. *Jurnal Buana Informatika, 7* (4), 255-264.

Hegadi, R., & Manjunath, T. N. (2013). Data Quality Assessment Model for Data Migration Business Enterprise. *International Journal of Engineering and Technology, 5* (1), 101- 109.

Katsurai, M., Ohmukai, I., & Takeda, H. (2016). Topic Representation of Researchers' Interests in a Large-Scale Academic Database and Its Application to Author Disambiguation. *IEICE TRANSACTIONS on Information and Systems, 99* (4), 1010-1018.

Martin, J. (1975). *Computer Database Organizations, parth I.* New Jersey: Prentice-Hall, Inc.

Raharjo, S. (2012). Constraint Basis Data sebagai Fondasi yang kuat dalam pengembangan Sistem Informasi. *Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, (hal. 08-1 - 08-5). Yogyakarta.

Utami, E., & Raharjo, S. (2006). *RDBMS dengan PostgreSQL di GNU/Linux.* Yogyakarta: Andi.