Journal of Information Technology and Computer Science
Volume 2, Number 2, 2017, pp. 66-75
Journal Homepage: www.jitecs.ub.ac.id

Scheduling Optimization For Extract, Transform, Load (Etl) Process On Data Warehouse Using Round Robin Method (Case Study: University Of Xyz)

Agung Yudha Berliantara[1], Satrio Agung Wicaksono[2], Aryo Pinandito[3]

Dept. Informatic/Computer Sicene, Faculty of Computer Science, University of Brawijaya
Veteran Road. 8, 65145 Malang, Indonesia

[1]agung.berliantara@gmail.com, [2]satrio.agung.w@ub.ac.id, [3]aryo@ub.ac.id

Received 09 July 2017; accepted 04 December 2017

Abstract. ETL scheduling is a challenging and exciting issue to solve. The ETL scheduling problem has many facets, one of which is the cost of time. If it is not handled correctly, it may take a very long time to execute and inconsistent data in very large data. In this study using Round-robin algorithm method that proved able to produce efficient results and in accordance with conventional methods. After doing the research, the difference between these two methods is about execution time. Through this experiment, the Round-robin scheduling method gives a more efficient execution time of up to 61% depending on the amount of data and the number of partitions used.

1. Introduction

In today's rapid technological development, various forms of technology coming into various areas of life. Especially in database technology. There are many reputable universities in Indonesia, and University of XYZ implements data warehouse technology. University of XYZ prioritizes many things in its business processes. So that one of the main needs for University XYZ are student academic data analysis covering the distribution of GPA, faculty income in every semester, and student questionnaire results in every semester. Based on the official website of XYZ University until January 2015, XYZ University has 64,037 active students from levels spread across 15 faculties [1]. We can conclude that University of XYZ has large number of data. It needed very large resources to process and do the analysist.

So that the huge amount of data can be processed and eventually provide useful information, it is necessary to gather all the data from various sources so that the results can be used to support the decision at the managerial level. One of promising alternative solution is the implementation of data warehouse.

Utilization of data warehousing in some educational institutions has provided many benefits in providing and providing accurate information as supporters in decision level at managerial level. The data in the data warehouse can be retrieved from many data sources through the Extract, Transform and Load (ETL) process. Data warehouse uses a multidimensional model (schema) to store data. Multidimensional models have basic components, namely dimensional tables and fact tables. The contents of the fact table are numerical measurements, such as numbers of data from summary. While the content

of the dimension table is the dimension of corporate entity related to the fact table.

There is a problem that arises during the process of scheduling for the ETL process as an example, while performing the ETL process for processing very large data will require enormous time costs. As in the journal paper written by Revathy Sreekumar on his research entitled "ETL Scheduling in Real-Time Data Warehousing" discusses the data warehouse architecture, the costs, and resolves scheduling. Thus, in this case a scheduling is required for an efficient ETL process as a solution to the problem.

In addition to research written by Revathy Sreekumar, this study is referring to other research written by Anastasion Karagiannis in his research entitled "Scheduling Strategies for Efficient ETL Execution". The discussion of this paper journal is not much different from Sreekumar's research, but there is a comparison between scheduling algorithms with the memory allocation parameters used and the execution time.

The scheduling process in ETL data warehouse is a process that must be passed for data warehouse formation [2]. If it is not handled properly, there will have inconsistency data when the data is already stored in the data warehouse. According to Anastasios Karagiannis research, the scheduling process on ETL is highly relevant for response time optimization or memory consumption allocation in the flow of processing operations (consumption) of data on tuples, more efficient use of ETL process time, and maintaining data consistency. In addition, the ETL scheduling process is also to minimize execution time or memory consumption without losing data, as well as avoiding deadlock [3].

The Round-robin method is chosen because, according to Karagiannis, the ETL scheduling process is used mostly to evaluate more sophisticated algorithms, and to improve the algorithm to be developed and the round-robin algorithm takes care of all activities without specific priorities. The advantage of using Round-robin scheduling is that every process that runs will get a fair share of CPU, easy to implement, and can be used to find out the worst time for a response time in a process [4].

Based on above problems it can provide an overview for the development of University XYZ multidimensional data structures against the needs of the correct data warehouse. Then, it is expected that Round-robin method can be implemented into ETL scheduling process so it can give the benefit and impact in quality and quantity of the data to be presented and the execution time saving.

2. Related Research

Revathy Sreekumar on his research entitled "ETL Scheduling in Real-Time Data Warehousing" discusses the data warehouse architecture, the importance of scheduling, and an explanation of some scheduling algorithms that are often applied to the ETL process [5]. The output of this journal paper is the result of a comparison of each algorithm that aims to increase the execution time and reduce memory consumption.

Anastasion Karagiannis in his research entitled "Scheduling Strategies for Efficient ETL Execuiton". The discussion of this paper is not much different from the journal written by Revathy Sreekumar on his research entitled "ETL Scheduling in Real-Time Data Warehousing", but there is a comparison between scheduling algorithms with memory allocation parameters used and execution time [2].

Based on the above studies, it can be concluded that the implementation of scheduling on ETL. That can help an organization to minimize the execution time when processing the ETL.

3. Round-robin Scheduling

3.1. Round-robin Scheduling in ETL

Sreekumar (2014) in his research mentions the scheduling algorithm is very importantfor a real-time system. Many scheduling algorithms are available. The choice of algorithm is very important in any real-time system and is strongly influenced by the type of algorithmic system to be executed. A scheduler provides policies for working on various processes in a real-time system. The scheduler is to ensures that all processes are executed according to the set of priorities. An online scheduler makes scheduling decisions based on the scheduling algorithm and the current state of the system [5].

The Round-robin scheduling algorithm can be applied to the ETL scheduling data warehouse process, along with a brief description [5]; Round-robin scheduling can be implemented in the ETL process. A list of the input tables where data has been updated in the staging database can be obtained by using the system tables. For ETL scheduling, it is assumed that push technology will be used to retrieve data from the data source into the staging database. In Round-robin scheduling, all the input tables will be given time to push data into the ETL process.

In an ETL process that implements the Round-robin algorithm, applying partitioning techniques to data that will be loaded into the data warehouse. In a classical environment, the ETL process runs on a unique machine called "ETL server" where all the data is processed by one ETL case. To distribute the ETL process to a cluster on a computer and run it on a parallel path, the data should also be distributed. Thus "data partitioning" is an issue in a parallel / distributed environment [6]. In the data partitioning technique performed in the ETL process there is also an equation formula that will divide the data into the data partition.

□□□□(□□□□□□□□□□) = □□□□(□)/□□_□□□□ (E 1)

As seen in the equation E 1, a simple partition is assigned a data volume v, a simple technique of generating a partition equal to the equation E 1 where nb_part is the number of data partitions [6]. Fig 1 illustrates how in Equation E 1 is implemented into the ETL process.

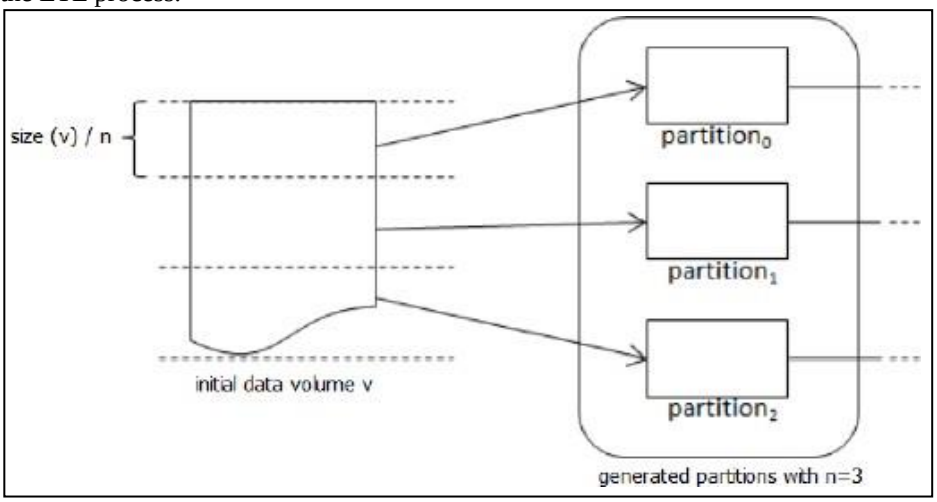

Fig 1. Round-robin partition

As mentioned before, in Figure 1 is an overview of the equation E 1. The equation divides the data on the left into the right partitioning session dividing the data into three partitions. For large amounts of data volumes, different parallel policies must be designed: dividing (partitioning) from the dataset into smaller sets. The idea is to use a different instance of the ETL process to handle any partition data. At the end of the process, the data partition will be merged and loaded to the target recordset. To divide (partitioning) the data, many implementations have been proposed with the aim of providing equal-sized partitions to facilitate data load to a single target, one of which is

Round-robin partitioning. In the Round-robin partitioning method, records are distributed between different processing nodes in a round-robin manner: the first record is sent to the first node, the second record of the second node, and so on. This method is suitable for resizing partitions that have data sets [7]. It can be concluded that the partition using the Round-robin method is to divide the amount of data flat into each of the previously created partitions.

The purpose of dividing data into partitions is to balance the amount of data that is sent to the data warehouse when in the ETL process. In the Round-robin Adaptive Server partition does not use the partition criteria. Adaptive Server provides Round-robin rows for each partition so that each partition contains the same or more equal amount of data from the row, so the load balancing on the processed data can be achieved. Since there are no criteria on the partition, the row is distributed randomly across partitions [8]. Additionally, just because of the partition, the calculation of the final query results is partially very fast and has a less meaningful impact in the overall query execution time. But the execution time also varies depending on the complexity of the query and most importantly the number of rows that are the result of partial of data [9].

To implement the Round-robin partition method in the ETL process, we need support tools to share and distribute the data into multiple copies in one step. For the example is in Fig 2.

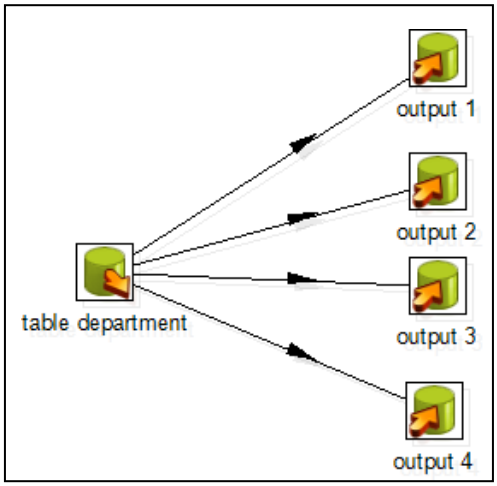

Fig 2. Example to applying the Round-robin scheduling from the ETL tools by dividing the data into 4 partitions

In Fig 2 the table source "table department", the application of the Round-robin method lies in the distribution of the target table division named output 1, output 2, output 3, and output 4. In this partitioning method, it is dividing data in the table source named "table department" into 4 (four) parts evenly, i.e. spreading the data on average the amount of data into each target table.

4. Methodology

4.1. Literature Review

The literature review is a search and reference collection stage that can be obtained from journal papers, books, e-books, research manuscripts, and the internet to get information and explanation of the theories used to support the basis of this research. From the literature collection, there are several literatures that support the research, among others:

1. Paper research from journals that support the related research that is about data warehouse and implementation Round-robin scheduling on ETL data warehouse,
2. E-book about database especially data warehouse, scheduling ETL data warehouse,

And relevant websites.

4.2. Interview

In the interview phase, this phase is done to get information about environment (environment) to manage the database of the University. The purpose of this interview is as reference material for the development of an environment database that will be used to support this research, so that the implementation process which will be run at least resembles the real environment.

Interview phase is done to determine the state of the environment system used to process the database includes OLTP and data warehouse itself. This stage is done by digging the information through interviews conducted with the ICT XYZ University.

Based on the interview result, it is known that to manage the academic data of XYZ University students there are several faculties that manage data by using their own database. The faculties are: Faculty of Law, Faculty of Administration, Faculty of Animal Husbandry, Faculty of Agriculture, Faculty of Medicine, Faculty of Mathematics and Natural Sciences, and Faculty of Civil Sciences. While some other faculty use one database to be used together in managing academic data. The faculties include: Faculty of Computer Science, Faculty of Agricultural Technology, and Faculty of Economics and Business. So it can be concluded that the amount needed to manage academic data faculty there are 8 different databases in one server. In addition to managing academic data, this database is also used to accommodate the results of student questionnaires for lecturers, the questionnaire is done at the end of each semester.

In addition, to manage the financial data of each faculty found in University of XYZ requires 1 database. The database is to accommodate the financial income of each faculty every semester.

4.3. Analysist and Designing of Database

The process of analysis is used to draw conclusions from the results of interviews that have been done before. In the previous interview process that produces an overview of the database environment on the actual situation, the implementation of the database environment will be carried out in this research. So this research is done based on the database environment in the real situation. Then also the analysis for the design of database structures, such as making observations to find out what tables in the database to be used to store data to be used next. Then after the analysis process, then the next stage is to design the database schema which is then applied to the database with SQL DDL to create the tables.

In the phase of analysis and design of this database will be design Physical Data Modeling (PDM) and physical architecture design that will be applied to OLTP database and data warehouse database. In accordance with the results of interviews that have been implemented and through the process of verifying the scheme design on OLTP and data warehouse, there are 2 (two) major schemes in the academic database and financial database. Therefore, the schema of the database tables are designed to resemble the actual situation at XYZ University.

4.4. Design Implementation

Implementation of the design at this phase is to implement the design of the database environment that has been done in the analysis phase. Then also done the implementation of the schema database table design by writing SQL syntax in SQL tools to generate the tables in the database.

Implementation of the database to be explained include the implementation or

application of OLTP schemes on academic schemes and OLTP schemes of the financial section. To implement the data warehouse implementation as multidimensional modeling, the implementation uses Structured Query Language (SQL) language is using Data Definition Language (DDL).

4.5. Build the Data

At this phase, the process of data development. The construction of the data in question is to build the data to be inserted into the tables in the database schema, which then these data will be used for processing and as ETL process material.

In the design of data is to fill data on OLTP tables, used dummy data. To build dummy data is required software processor numbers (spreadsheet). In creating a single table data, it takes at least one sheet on one spreadsheet file.

For data such as student name, student id, lecturer's name and lecturer id can be generated by sorting and filling of the automatic data provided by the software to generate numerical data and the ordered name. Then, the data is saved as CSV. For the construction of student and college data, it is necessary data of the number of students of each study program and course data first. The data obtained from the interview and data request to the ICT of the university. For data development of lecturers, the data is obtained from PDDIKTI. Then the ready data is inserted into the database table.

After built the data on each table in the OLTP scheme is completed, the next step is to import data into the tables in the OLTP scheme. This process is done by using SQL processing software.

4.6. ETL Structure Design

ETL structure design is a process used to design the flow of data to be processed in the ETL process. This stage is used in order to know from anywhere the data is processed so that it can be done mapping process on ETL and inserted into data warehouse scheme. After designing the system, which includes designing the PDM on OLTP and on the data warehouse scheme and then done to design and construction of the data, then next is to design the ETL process. This design is useful for mapping the attributes in the OLTP schema table to the data warehouse scheme for the ETL process.

4.7. Implementation of ELT Structure

In the implementation phase of the ETL structure, an implementation of ETL structure design was previously designed. This implementation will produce the ETL structure to be used, so in the next process can be analyzed to this ETL process.

In the implementation phase of this ETL structure, a pre-designed ETL structure was implemented. In addition, in this section will also be carried out the implementation of ETL process testing using round-robin method and which do not use the round-robin method. After that will also be shown the results of this implementation process. For round-robin method it will use four group of partitions, i.e. 2 partitions, 4 partitions, 5 partitions and 6 partitions.

4.8. Testing

The purpose of the test is to assess the results of a pre-designed ETL implementation. The test result is a performance test from ETL. The ETL performance tested in this study is viewed in terms of execution time. This ETL performance testing phase is by comparing ETL data flow that applies Round-robin scheduling method with conventional method. So it will produce a graph of the difference between the ETL process using the Round-robin scheduling method and which does not use the Round-robin method.

In the ETL performance testing phase, we will analyze the data transfer rate received

by the dimension and fact tables. Analysis of the speed is shown while the step that processed in the output table used as a target.

In analyzing the results of this process, ETL execution (run) is executed 5 times. But before the run process, the data on the target table, i.e. dimensions or facts, will be emptied first, so there is no data update process or purely for insert data only.

In the analysis process is executed 5 times and will be observed from the execution time on each execution process. For ETL process using round-robin method will be taken average execution time on each output table or partition.

4.8.1. Performance Testing for ETL for Dimension Table DIM_DOSEN

Table 1. Testing Result ETL execution of dimension table DIM_DOSEN.

Testing seq	Testing results (seconds)				
	Conventional	RR 2 parts	RR 4 parts	RR 5 parts	RR 6 parts
1	2.9	1.9	1.3	1.8	1.1
2	2.8	1.9	1.3	1.8	1.1
3	2.7	1.9	1.35	1.8	1.1
4	2.7	1.9	1.325	1.8	1.1
5	2.7	1.9	1.3	1.8	1.1

In Table 1 the result of ETL execution analysis for DIM_DOSEN data with 1577 rows of data. In the conventional method to execute the ETL takes about 2.76 seconds in average. While the analysis using the Round-robin method only takes an average of about 1.1 to 1.9 seconds.

4.8.2. Performance Testing for ETL for Dimension Table DIM_PRODI

Table 2. Testing Result ETL execution of dimension table DIM_DOSEN.

Testing seq	Testing results (seconds)				
	Conventional	RR 2 parts	RR 4 parts	RR 5 parts	RR 6 parts
1	0,7	0,7	0,7	0,7	0,7
2	0,7	0,7	0,7	0,7	0,7
3	0,7	0,7	0,7	0,7	0,7
4	0,7	0,7	0,7	0,7	0,7
5	0,7	0,7	0,7	0,7	0,7

In Table 2 is the result of ETL execution analysis for data DIM_PRODI of 77 rows of data. In the conventional method to execute the ETL takes an average time of 0.7 seconds and from the analysis using the Round-robin method shows the same result.

4.8.3. Performance Testing for ETL for Dimension Table DIM_STATUS_KULIAH

In Table 3 is the result of ETL execution analysis for data DIM_STATUS_KULIAH with 8 rows of data. In the conventional method to execute the ETL takes an average of 0.3 seconds and from the analysis using the Round-robin method shows the same result.

Table 3. Testing Result ETL execution of dimension table DIM_STATUS_KULIAH.

Testing seq	Testing Results (seconds)				
	Conventional	RR 2 parts	RR 4 parts	RR 5 parts	RR 6 parts
1	0.3	0.3	0.3	0.3	0.3
2	0.3	0.3	0.3	0.3	0.3

3	0.3	0.3	0.3	0.3	0.3
4	0.3	0.3	0.3	0.3	0.3
5	0.3	0.3	0.3	0.3	0.3

4.8.4. Performance Testing for Dimension ETL table DIM_MATKUL

Table 4. Testing Result ETL execution of dimension table DIM_MATKUL.

Testing seq	Testing Results (ms)				
	Conventional	RR 2 parts	RR 4 parts	RR 5 parts	RR 6 parts
1	3.4	3.2	3.1	2.7	2.6
2	3.4	3.2	3.1	2.7	2.6
3	3.4	3.2	3.1	2.7	2.6
4	3.4	3.2	3.1	2.7	2.6
5	3.4	3.2	3.1	2.7	2.6

Table 4 is the result of ETL execution analysis for DIM_MATKUL data of 5100 rows of data. In the conventional method to execute the ETL takes an average of 3.4 seconds. While the analysis using the Round-robin method only takes an average of about 2.6 to 3.2 seconds.

4.8.5. Performance Testing for ETL for Dimension Table DIM_PERTANYAAN_KUESIONER

In Table 5 is the result of ETL execution analysis for DIM_STATUS_PERTANYAAN_KUESIONER data of 28 rows of data. The data has been through the process of extract and transform before so that the results of the analysis there are differences in execution between conventional ETL performance testing and ETL performance testing using Round-robin method. In the conventional method to execute the ETL takes an average time of 0.3 seconds and from the analysis using the Round-robin method shows the same result.

Table 5. Testing Result ETL execution of dimension table DIM_PERTANYAAN_KUESIONER.

Testing seq	Testing Results (seconds)				
	Conventional	RR 2 parts	RR 4 parts	RR 5 parts	RR 6 parts
1	0.3	0.3	0.3	0.3	0.3
2	0.3	0.3	0.3	0.3	0.3
3	0.3	0.3	0.3	0.3	0.3
4	0.3	0.3	0.3	0.3	0.3
5	0.3	0.3	0.3	0.3	0.3

4.8.6. Performance Testing for ETL for Dimension Table DIM_KELAS

In Table 6 is the result of ETL execution analysis for DIM_KELAS data as many as 4 rows of data. In the conventional method to execute the ETL takes an average of 0.2 seconds and from the analysis using the Round-robin method shows the same result.

Table 6. Testing Result ETL execution of dimension table DIM_KELAS.

Testing seq	Testing Results (seconds)				
	Conventional	RR 2 parts	RR 4 parts	RR 5 parts	RR 6 parts
1	0.2	0.2	0.2	0.2	0.2
2	0.2	0.2	0.2	0.2	0.2
3	0.2	0.2	0.2	0.2	0.2
4	0.2	0.2	0.2	0.2	0.2
5	0.2	0.2	0.2	0.2	0.2

4.8.7. Performance Testing for Dimension ETL table DIM_SELEKSI

In Table 7 is the result of an ETL execution analysis for DIM_KELAS data of 7 rows of data. In the conventional method to execute the ETL takes an average time of 0.3 seconds and from the analysis using the Round-robin method shows the same result.

Table 7. Testing Result ETL execution of dimension table DIM_SELEKSI.

Testing seq	Testing Results (seconds)				
	Conventional	RR 2 parts	RR 4 parts	RR 5 parts	RR 6 parts
1	0.3	0.3	0.3	0.3	0.3
2	0.3	0.3	0.3	0.3	0.3
3	0.3	0.3	0.3	0.3	0.3
4	0.3	0.3	0.3	0.3	0.3
5	0.3	0.3	0.3	0.3	0.3

4.8.8. Performance Testing for ETL for Fact Table FACT_MAHASISWA

In Table 8 is the result of ETL execution analysis for FACT_MAHASISWA data of 39542 rows of data. On the use of conventional means to perform ETL it takes an average time of 16.2 seconds. While the analysis using the Round-robin method only takes an average of about 7.1 to 9.45 seconds.

Table 8. Testing Result ETL execution of dimension table FACT_MAHASISWA.

Testing seq	Testing Results (seconds)				
	Conventional	RR 2 parts	RR 4 parts	RR 5 parts	RR 6 parts
1	16.2	9.45	7.2	7.12	7.1
2	16.1	9.45	7.2	7.12	7.1
3	16.1	9.45	7.2	7.12	7.1
4	16.2	9.45	7.2	7.12	7.1
5	16.2	9.45	7.2	7.12	7.1

4.8.9. Performance Testing for ETL for Fact Table FACT_KUESIONER

Table 9. Testing Result ETL execution of dimension table FACT_KUESIONER.

Testing seq	Testing Results (minute)				
	Conventional	RR 2 parts	RR 4 parts	RR 5 parts	RR 6 parts
1	35	30.4	27.51	25.1	25
2	35	30.4	27.51	25.1	25
3	35	30.4	27.51	25.1	25
4	35	30.4	27.51	25.1	25
5	35	30.4	27.51	25.1	25

In Table 9 is the result of ETL execution analysis for FACT_KUESIONER data with 6.735.228 rows of data. In the conventional method of doing the ETL takes an average of 35 minutes. While the analysis using the Round-robin method only takes an average of about 7.1 to 9.45 minutes.

5. Results

In this paper, implementation of ETL scheduling optimization has been done by comparing the execution time from ETL execution by using Round-robin method and conventional method. As shown in Table 1, the results of 9 execution times show different results, and some of the others show similar results despite using the Round-robin method.

In the execution results using the Round-robin method the results tend to be superior or more efficient. Because the data is divided into parts that have been provided previously. The data previously processed in the extract and transform process will be divided into several groups when entering the loading process. In the loading process the data will be distributed evenly into the part of groups. For example there are 1000 rows of data with 3 parts for Round-robin, the data on row 1 will go into parts 1, the second row data will go into the 2nd part, the third row data will go into the 3rd part, the fourth row data entry To parts 1, the 5th row data goes into the 2nd part, and so on. Once the data is loaded and then together the data stored in the loading step was put in the target table. The Round-robin method becomes superior because several rows of data are entered together to shorten the execution time. While the conventional method, the data will be distributed one by one (queued) into the target table.

6. Conclusion

The scheme of multidimensional data warehouse scheme in accordance with the requirement resulted 2 fact tables added with 7 dimension tables and has got verification of eligibility from ICT side University of XYZ. For ETL structures for ETL processes in the XYZ University data warehouse are divided into 2 (two) types of structures. For ETL processes in conventional method, the data sent to dimension or fact tables is not divided into several groups, so the incoming data will be queued. As for ETL structure using Round-robin method, the data will be divided into the parts that have been provided. So we will need less time used to process data into the target table.

Then the most important of this research is ETL scheduling processed with conventional method proved less efficient when compared with ETL scheduling process using Round-robin method. Where when using Round-robin scheduling method can be a time savings up to 61.1% depending on the amount of data and the number of parts. Then by adding the number of parts, then the time required to run the ETL process will be less or in other words, the more parts provided the less time required for the ETL process.

References

[1] U. Brawijaya, “Universitas Brawijaya,” Universitas Brawijaya, [Online]. Available: http://www.ub.ac.id/. [Diakses 8 November 2016].

[2] R. Kimball dan J. Caserta, The Data Warehouse ETL Toolkit : Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data, Wiley, 2004.

[3] A. Karagiannis, P. Vassiliadis dan AlkisSimitsis, “Scheduling strategies for efficient ETL execution,” *Elsevier*, 2013.

[4] P. Krzyzanowski, "Paul Krzyzanowski's Site," 2015. [Online]. Available: https://www.cs.rutgers.edu/~pxk/416/notes/07-scheduling.html. [Diakses 18 September 2016].

[5] R. Sreekumar dan S. Balaji, "ETL Scheduling in Real-Time Data Warehousing," *International Journal of Computer Science & Engineering Technology (IJCSET),* vol. 5, 2014.

[6] M. Bala, O. Boussaid dan Z. Alimazighi, "P-ETL: Parallel-ETL based on the MapReduce Paradigm," *IEEE,* 2014.

[7] S. Kozielski dan R. Wrembel, New Trends in Data Warehousing and Data Analysis, 3rd penyunt., Springer Science & Business Media, 2008.

[8] Sybase Inc., "SyBooks Online," 2009. [Online]. Available: http://infocenter.sybase.com/help/index.jsp?topic=/com.sybase.infocenter.dc32300.1550/html/sqlug/sqlug448.htm. [Diakses 10 November 2016].

[9] J. Bernardino dan H. Madeira, "Experimental Evaluation of a New Distributed Partitioning Technique for Data Warehouses," *IEEE,* pp. 318-319, 2001.