Design of Data Warehouse System to Support the Quality Management of Information Technology Based School

Didik Nugroho¹, Sri Siswanti², Tri Irawati³, Kustanto⁴

¹ STMIK Sinar Nusantara Surakarta, Jawa Tengah, Indonesia
² STMIK Sinar Nusantara Surakarta, Jawa Tengah, Indonesia
³ STMIK Sinar Nusantara Surakarta, Jawa Tengah, Indonesia
⁴ STMIK Sinar Nusantara Surakarta, Jawa Tengah, Indonesia

Abstract
The quality of each school in Indonesia is appraised by the government based on an Accreditation Standard established by the Board of School Accreditation (Badan Akreditasi Sekolah). The standard of accreditation system consists of eight standards. There are more than a hundred chapters among those eight standards that must be answered and proved by related documents, and data records. In addition, there are several chapters of criteria that need data measurement of the result of the recording. At present, there have been many schools applying OLTP to record the activity of data recording. However, there have not been many schools that apply the Information Technology for processing the school accreditation. In this research funded by the Board of High Education, the researchers would like to make a management system of data warehouse so that the data from OLTP can be loaded into the data warehouse and later can be analyzed by OLAP tools. By storing data, it will support the accreditation system which makes easier for the school or the accreditation reviewer easier to conduct the accreditation process. In addition, it also becomes one of the forms of Total Quality Management (TQM) implementation based on Information Technology at school.

Keywords: quality, data warehouse, accreditation, TQM, OLTP, OLAP

1. Introduction
The quality of each school in Indonesia is appraised by the government based on an Accreditation Standard established by the Board of School Accreditation. The standard of accreditation system consists of eight standards. There are more than a hundred chapters among those eight standards that must be answered and proved by related documents, and data records. In addition, there are several chapters of criteria that need data measurement of the result of the recording. Frequently, those data need to be prepared in the five years before the accreditation is conducted.

At present, there have been many high schools applying daily data transaction recording (OLAP) by using School Administration Program (PAS) or JBAS program issued by related department. These applications have not been able to give data recording for the need of the accreditation process. This, according to our observation, is due to the making of the PAS and JBAS applications was not based on the School Accreditation Standard. Since OLTP does not support the accreditation, the process of accreditation faces many problems in preparing the data since five years before the accreditation process is carried out.

Schools that have not applied OLTP in recording their data will face a much worse problem in preparing the data for the accreditation process. As a result, they must dig many paper documents. Surely, it will waste much time and disturb the other activities.

Based on the problem of preparing the data for the school accreditation above, we are interested in making a research on making a design of datawarehouse house based on the School Accreditation Standards. With this, we will design how to input activities of data transaction at school into the datawarehouse house. By using Information technology, it is hoped that the process of accreditation will be much more helped and run even more smoothly.

2. Background of design
In the literature to write our paper:
2.1 Quality Standard in Education

The quality control in education is the process which is integrated one to each other to collect, analyze, and report data of an education program or activity to reach the goaled quality of education. The process of quality control is started by identifying the aspect of achievement and the priority excalation and supplying data as the base of planning and decision making and helping to build the culture of continuously quality control. The achievement of the education quality for basic and intermediate education is examined based on eight national education standards from the Board of national Education Standard [1][2][3]. The direct and continuous quality control surely contributes the raising of the education quality.

The Indonesian Government Act [3] no.19, 2005 on the national education standard can be viewed as an important base to reach the national standardized education. It is said, in that act, that the National Education Standard (Standar nasional Pendidikan/SNP) is the minimum criteria on education system in all area of Indonesian republic. Those eight standards are (1) content standard,(2) the process standard, (3) the graduate competence standard, (4) the educators and the education labor standard, (6) tools and equipment standard, (7) the management standard, and (8) the education appraising standard

In order to meet those criteria, the education ministry has made a book of the national standardized education fulfillment. This book explains the components that need to be fulfilled in order to reach the quality standard and the indicators that must be achieved by a certain school. The indicators must be fulfilled by showing the document or detailed data record in the form of either ordinary record or information technology based record. There are certain indicators that need resume process of all those data records.

The process of the quality fulfillment is the process of achieving the quality assurance. Quality assurance is a way of producing products that are free from deformity and fault. The next step of the quality assurance is the Total Quality Management (TQM) that tries to create a culture of quality by pushing all members to satisfy the customers continously.

2.2 Data Warehouse

A data warehouse is a repository of an organization's electronically stored data. Data warehouses are designed to facilitate reporting and analysis[10]. A data warehouse is a subject-oriented, integrated, time-varying, non-volatile collection of data in support of the management’s decision making process[12][27]. Data Warehouses are mainly databases that responsible for collection and storage of historical and current business data[13][17][25]. Data warehouses can server the requirement of extraction of information from more than one subject area. Data warehose is a single database is dedicated to stores a copy of several transactional databases within the company and is optimized for query and analysis[26].

Data warehouse is not a product but an environment in which a user is able to find strategic information [5]. Data warehouse [5][6] is a computer system to archive and analyze the historical data of an organization such as data of sales, salary, and other information of daily operation. Generally, an organization copies the information of its operational system (such as sales and human resources) to the data warehouse based on regular schedule like every night or every weekend. After that, the management can carry out several complex quieries and analyze (like data mining) on the information without burdening the operational system.

Data warehouse is built to separate the historical data from the transactional data, is more static, and the data is collected for business analysis. Data warehouse is a system consolidating data from the available systems into a dimensional storage periodically. Generally, data warehouse stores the historical data for several years and queries them for intelligence business necessity or activates the other analysis. Data warehouse is rejuvenated batchedly not spontaneously when a transaction takes place on the source systems [8]. The characteristics of data warehouse is Subject-Oriented or being oriented on the subject, Time-variant, means that the data changing is traced and recorded so that the report can be made by showing the time of the changing, Non Volatile, means that the recorded data cannot be changed, and the last characteristic is integrated in which the warehouse includes all operational data of an organization which is stored consistently.

2.3 Structure of data warehouse

Data warehouse consists of parts called ETL, stage, database, metadata and analyzing tool (OLAP).

2.3.1 ETL

Development of data warehouse includes development of systems to extract data from operational systems. The data from these sources are converted into a form suitable for data warehouse. This process is called Extraction, Transformation and Loading[15][18][20]. ETL
management module performs data conversion tasks stored
in metadata which extracts tasks and resolve them to the
rules of data cleansing, transformation and loading, and
cleans them according cleaning rules of the relevant
data[13]. The warehousing system has the
responsibility to provide proper data cleaning technique to
clean the dirty data which occurs in the applications. Also,
the cleaned data has to be transformed and to be loaded properly[23].

2.3.2 Stage

The data from OLTP database will be copied to the
database of data warehouse through a stage area. Stage
area is a temporary area to select data being copied from
OLTP database. The data which is copied by data totally
has good quality. Stage is the former place to evaluate the
data source from the data warehouse[22].

2.3.3 dimension and Fact table

A dimension is a collection of members or units of
the same type of views. In a diagram, a dimension
is usually represented by an axis [11]. The
dimensions are used to charac-terize measures, which in
turn are stored in fact tables [16].

Fact table is a table that is going to be related to the
different dimension tables. The fact table is the having
measured fact [11][16]. A fact table consists of the
relationship among keys of multi-table attribute dimension
table and attributes that have numeric value [7]. The
numeric value is the measured value or something that
shows a measurement of a certain aggregate value [7].

A fact is a collection of related data items, consisting of
measures and context data. Each fact typically represents a
business item, a business transaction, or an event that can
be used in analyzing the business or business processes.
In a data warehouse, facts are implemented in the core tables
in which all of the numeric data is stored [11][16].

The fact will consist of many attribute keys so that they will
compose multidimensional database known as Cube [6].
Cube is a form of database in which data is stored in the
form of Cell, and the cell’s position is determined by
several variables called as Dimension. Theoretically,
the number of the dimension can be unlimited and
unquantified to form a three dimensional in the form of
cube. The terms Cube and the imaging of the cube form (3
dimension) are meant to make our visualization about
multidimensional characteristics easier.

2.3.4 OLAP

OLAP [9] is a category of software technology that enables ana-
lysts, managers and executives to gain insight into data
to reflect the real dimensionality of the enterprise as understood b y the user. On-line Analytical Processing (OLAP) is theanalytical capabilities provided
by the data warehouse[10][22]. On-line analytical processing (OLAP) is a term used to describe the analysis of complex data from the data warehouse [12][24].

2.3.5 Metadata

Metadata is a data used to explain data. Metadata is a data
used to explain data in a database [28]. Metadata represents[20] the information about data to be stored in Data Warehouses. Meta data repository which stores each and every characteristic of the data items loaded in the Data Warehouse[22]. Metadata helps in data integration,lineage, data quality and populating transformed data into data warehouse. Metadata is database , This data base contains data about data-description of source data, target data and how the source
data has been modified into target data[15].

3. Desain Datawarehouse and archive data

3.1 Data Analysis

The school accreditation process is carried out by the Board
of School Accreditation or Badan Akreditasi Sekolah. A
reviewer will ask the school to show him the data in the
form of the school management document, the school
activity document, and its results which are based on the
standard determined by the Board of the School
Accreditation. The followings are the data of the Middle
School Accreditation standards:
1. The standard of content
2. The standard of process
3. The standard of graduates and cooperation
4. The standard of educators and the labor of education
5. The standard of tools and equipment
6. The standard of management
7. The standard of financing
8. The standard of appraising

The data resulted from the school activities will be used by
the reviewer to appraise the level of the school’s
achievement toward the eight standards determined by the
board of the school accreditation. The needed documents
and data are sometimes those that record a certain period
of time. Therefore it is necessary to carry out data and document archiving. From there the application of the information technology on the school quality management is used as a media to store data for a certain period of time to be able to be analyzed. The approach that can be used is by making data warehouse. The data warehouse will be used to store data form OLTP that support the accreditation. The data that can be stored in the data warehouse is categorized into 3 forms, they are:

1. Data in the form of document
2. The data record
3. The data with value that can be used to measure

Those three data are based on the reviewer’s need during the process of accreditation

The data source needed for the system of school quality management is stemmed in the OLTP database.

a. Student data
b. Finance data
c. Academic data
d. Human resource data
e. Alumna and cooperation
f. Tools and equipment

The source of those data is from the OLTP transaction database. Generally, the information available in data warehouse and the data in the OLTP application database are the same.

3.2 Stages of System Design

Data warehouse is defined as a collection of data that can be used to support the management decision making that is oriented on topic (subject), therefore it is called data warehouse approach. The quality management system is a system to fulfill the need of data reviewer, hence the process of making that we did is started from the analysis of the need of data on the quality standard that is determined by The Board of the School Accreditation. After that the structure design of data warehouse database that can store the need of data and document to meet those standards is determined. Next, the stage design will be used for temporary data warehouse process. Stage is the database that we make based on the structure of data warehouse database and the structure of data from the data source from the OLTP. After the structure design of the database is made, then we make tools to input the transaction database to the stage (ETL). The tools move the data from stage to the database of the data warehouse and the OLAP tools display the data from the database of the data warehouse to the reviewer.

3.3 The Architecture of data warehouse System

Data warehouse Serves the need of the school accreditation data which is stemmed on the OLTP that processes the student data, the human resource data, the syllabus, the academic/education data, asset, and finance. Those data are exported in the form of excel. We choose Excel because most schools still use excel in their data processing.

The excel file from OLTP and ETL then is loaded into the stage table. In order to make it easier in making the process of moving the data from stage to the database of the data warehouse, we make the structure of the stage database matched with the structure of dimension tables from the database of the data warehouse.

The data which is in the stage database is moved to the database of the data warehouse to its dimension tables. The data in the dimension tables then are copied into the fact tables on the database of data warehouse.

![Figure 1: Architecture of the datawarehouse system](image)

The data needed for the system of the school quality management are:

1. Student data
2. Finance
3. Academic dept
4. Human resource
5. Alumna and cooperation
6. Tool and equipment

The source of the data is from the database of OLTP transaction. Generally, the information available in the data warehouse tables is in the form of the recording of the data in the OLTP database in a long period of time.

3.3.1 Stage

Stage is a temporary database saving data from the data source in which later they can be transferred into the dimension table in the database of the data warehouse. The structure of the stage database is made similarly with the
structure of the dimension table. Therefore, the process of transferring data will not suffer resistance. The process of transferring is by using copy command after that the data available on the stage must be deflated or emptied.

3.3.2 ETL on Stage and Cleaning

ETL is a process to transfer data from OLTP to the stage. In this case, we make ETL in the form of Php Script. The data needed from the transaction database is made in excel, after that is loaded with Php script. In Php script, along with the transferring of data from excel to the stage database, the process of cleaning is also carried out by checking every line of the excel whether it has been matched with the requirements put in the script or not. If it has been matched then the lines in the excel is copied to the stage. The structure of the stage database is made as the structure of the dimension table in the database of the data warehouse.

Dynamic ETL, In the application of the quality management, the quality standard is always increased. This increasing will influence the need of data to measure the standard. As a result, it will influence the database of the data warehouse that will also make the stage table change. So, to accommodate the changing of the quality standard, the process of ETL is made to be dynamic.

The process that we make is by making name of ETL script that is stored in a table, for example, command_etl table. Therefore, when a user is doing an ETL process, he or she will only need to read all names of the script file and runs it in a process. The process of running the ETL command is managed by a procedure in a certain period of time depending on the organization’s need in carrying out the quality management.

3.3.3 The database structure of data warehouse

The database in the data warehouse consists of several types:

a. Dimension table type

Dimension table is a table that has one key attribute. Based on the previous references, we conclude that Dimension is a group of data consisting of individual and is not overlapped with the data element. Dimension is a group of information references about the measured event. The following is the table of employee training standard. We provide three tables: employee table, skill table, and time table (figure 3).

![Figure 3. Dimension Table time, Skill and Employee](image)

The goal of data warehouse is to make the data analysis easier to run, meanwhile the process of accreditation needs several proves in the form data recording, the data recording in the transaction table is also unique by showing a primary key. With this primary key, the transaction data is not overlapped. Therefore, we use the tables of the data transaction at school that have been normalized as a dimension table. The transaction table sometimes does not have measurement but only data record. Hence, we make this transaction table into a dimension table. The following is the example of dimension table that records the time of employee training.
b. Fact table

Grain is a data from a future fact that can be analyzed. Choosing a grain means determining what is being represented by record in a fact table. In this case, the choosing of grain is based on the standard need that must be fulfilled. The fact table in this system is used to display the data of the achievement measurement\[11]\[16] of a number of human resource training on a certain field that is asked by a quality standard. The fact table is made/filled from the data of the table dimension. Here is a structure of fact table of number_of_training on a certain skill. The data of the fact table is resulted from the process of ETL from stage to the database of the data warehouse. ETL is a query command that runs the aggregate command or resume. In the training table, the command of sum is runned based on the id_skill group and year.

Table 1  Fact_resum skill

<table>
<thead>
<tr>
<th>Id_k</th>
<th>Year</th>
<th>Count_skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1</td>
<td>2010</td>
<td>0</td>
</tr>
<tr>
<td>K2</td>
<td>2011</td>
<td>0</td>
</tr>
<tr>
<td>K3</td>
<td>2012</td>
<td>1</td>
</tr>
<tr>
<td>K4</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

c. Archive table

Archive table is a type of table which is used to archive data in the form of document. The document is standardly made in Pdf format. Archive tables will save document information and its pdf file name (the document file). The physical file will be stored in a server of a data warehouse. In this data warehouse, the archive data is used to display the need of physical documents saved in pdf format.

d. Metadata table

Metadata table is a table containing information to give explanation on other data. Metadata represents the information about data to be stored in Data Warehouses [19].

3.4. Schema

In a data warehouse, a database will have a scheme. Most previous writers mentioned that scheme consists of STAR or snowflake [9][10][16]. In the database usage for this accreditation, we use dimensional tables in which there are several ones that are normalized. Therefore, it will make the scheme form is not a star or a snowflake anymore. Otherwise, we name it as a network scheme. We are also assuming that the normalization of the three areas is a fact table. Examples for the design data were assess the amount of employee training scheme show figure 5.

![Schema resum skill](image)

Fig. 5. Scheme resum skill

3.5 ETL stage to Data warehouse

This process is in which a script that is being used to transfer the stage database to the database of the data warehouse. Data from the stage database will be copied to the dimension tables. Next, the content of the stage database tables will be emptied.

From the dimension tables, the ETL script will run a process of transformation to fill the data in the fact table. Therefore, the data on the fact table will always be updated. The process of transferring data is done batchedly and periodically based on the manual procedure made by the organization. With this continuous process, the need of data during accreditation is met.
3.6 Analysis and Report Tools

As many writers wrote, the data warehouse is used to help making decision. The providing of information is resulted by using tools called OLAP [Swanhart, 2010]. OLAP is a query command accessing database of the data warehouse. In this case, the data warehouse is used to give information/data needed by the accreditation reviewer. The analysis and report tools are scripts used to display the data needed by the accreditation reviewer that are matched with the quality standard. If the data needed during accreditation is the data of the result of the transaction of events/data record, the physical document data, then the OLAP script contains query commands that read the dimension tables.

The providing of data is in the form of recapitulation or measurement scores on the achievement on OLAP quality standard in the form of script that runs query to access fact tables.

4. Conclusions

From the analysis on the usage of information technology in managing quality management, it can be concluded that:

1. The database of the data warehouse is influenced by the quality standard, therefore the making of the database structure is based on the need of the accreditation standard data.
2. The data warehouse supports the quality management that consists of data table in the form of dimension table, fact table, and archive table. The archive table saves the document’s file names. The document data is saved in an archive folder.
3. In the process of accreditation, the need of data is real or detail by accessing the dimension table. The process shows information of the measurement result using access to the fact table.

Acknowledgments

We are deeply grateful for Dirjen Dikti that has facilitated and funded this research so that we can develop it.

References

[1] Government Regulation No. 19 Year 2005 on National Education Standards (State Gazette of the Republic of Indonesia Year 2005 Number 41, Supplement to State Gazette of the Republic of Indonesia Number 4496)
[2] Law Number 20 Year 2003 on National Education System (State Gazette of the Republic of Indonesia Year 2003 Number 78, Supplement to State Gazette of the Republic of Indonesia Number 4301)


Nouha Arfaoui, Jalel Akaichi, Data Warehouse: Conceptual And Logical Schema-Survey, International Journal of Enterprise Computing and Business Systems ISSN (Online) : 2230-8849


Didik Nugroho is a lecturer of Informatics technique department in STMIK Sinar Nusantara, Surakarta, Indonesia. He graduated from Master Degree program in the field of computer at Dian Nuswantoro University, Semarang, Indonesia.

Sri Siswanti is a lecturer of Informatics technique department in STMIK Sinar Nusantara, Surakarta, Indonesia. She graduated from Master Degree program in the field of computer at Dian Nuswantoro University, Semarang, Indonesia.

Tri Irawati is a lecturer of Informatics technique department in STMIK Sinar Nusantara, Surakarta, Indonesia. She graduated from Master Degree program in the field of accounting at Sebelas Maret University, Surakarta, Indonesia.

Kustanto is a lecturer of Informatics technique department in STMIK Sinar Nusantara, Surakarta, Indonesia. He graduated from Master Degree program in the field of computer at Gajah Mada University, Yogyakarta, Indonesia.