Development of Student Data Mart using Normalized Data Store Architecture

by Slamet Riyadi

Submission date: 21-Mar-2018 01:14PM (UTC+0700)
Submission ID: 933705752
File name: Main_Article_B2.pdf (252.43K)
Word count: 3084
Character count: 16579

Advanced Science Letters
Vol. 21, 3226–3230, 2015

Development of Student Data Mart using Normalized Data Store Architecture

Luthfia Rahman, Slamet Riyadi, Eko Prasetyo
Muhammadiyah University of Yogyakarta, Indonesia

As one of the private university in Indonesia, Muhammadiyah University of Yogyakarta make assessing the performance and feasibility of the institution. The data required in the accreditation process includes various aspects, such as student, alumnus, employee, facilities and others. However, to obtain that data, the institution is still having trouble because the data is not yet integrated. Integrated data obtained by building a data warehouse. This research uses data warehouse with Normalized Data Store (NDS) architecture. There are some noise was found that need to be cleaned, such as null data and duplication data. Student data mart has been built and it can meet the requirements of the institution that are displayed in a report

Keywords: Accreditation, Data Mart, Data Warehouse, Student.

1. INTRODUCTION

A. Background

Every level and education units require accreditation to determine the program eligibility and that educational unit. The government will accredit the institutional and the program level periodically. The accreditation process requires some data that must be completed. The data is based on the accreditation forms that have been determined by Badan Akreditasi Nasional (BAN).

Muhammadiyah University of Yogyakarta (UMY) is one of educational institution in Indonesia. UMY will not be separated from feasibility program and educational units assessment, it means the accreditation. The data required in the accreditation process includes various aspects, such as student, alumnus, employee, facilities and others.

Data management that are not integrated will cause the possibility that the difference the data output. And institution would be difficult to get the information quickly and accurately because they who have been applying database technology will always save the data, so they have data with a very large number. To get information from the data with large size not been easy, because data stored in a database is detailed and not integrated.

A solution that can be applied is data warehouse. As one of private university in Indonesia, UMY will do that assesment, so data warehouse with Normalized Data Store (NDS) architecture be used as an instrument for supporting accrediting institution.

B. Formulation of the problem

According to the background which has been described above, so the formulation problems that arises is how to build student data mart on data warehouse of Muhammadiyah University of Yogyakarta using Normalized Data Store (NDS) architecture.

C. Goal and Benefit

The purpose of this research is to build student data mart on data warehouse of Muhammadiyah University of Yogyakarta using NDS architecture. The benefits expected from this research is providing facilities for the institution in obtaining information when the process of accrediting institution.

2. THEORETICAL

A. Literature Review

Research related to data warehouse has been done several times. Some references for this research are:

- Research titled Desain dan Implementasi Data Warehouse Studi Kasus Pemetaan Daerah Rawan Bencana Provinsi Sumatera Barat by Novianto Budi Kurniawan from Sekolah Teknik Elektro dan Informasi (STEI), Bandung Institute of Technology. In his research, Novianto assessing about the design of the data

warehouse to be implemented to produce a report mapping information the area vulnerable to disasters the province of West Sumatra.[3]

- A system information journal titled Perancangan dan Pembuatan Data Warehouse untuk Kebutuhan Sistem Pendukung Keputusan di Bidang Akademik pada Jurusan Sistem Informasi, ITS, Surabaya from Restia Rezalini P.S., Wiwik Anggraeni, Radityo Presetianto Wibowo. They built a data warehouse for support the academic decision making in Department of Information System. The result is the data warehouse can give the output some data that used to evaluate and analyze the student activity. But the data warehouse is not integrated yet and the application dashboard didn't ready to show the data [4].
- Journal from Windarto also discuss about data warehouse. It titled Pemanfaatan Data Warehouse sebagai Sarana Penunjang Penyusunan Borang Akreditasi Standar 3 pada Fakultas Teknologi Informasi Universitas Budi Luhur use star schema model. With Business Intelligent, information from this data warehouse can give in tabular or graphic appropriate with the stakeholder. And it can support to fill in Borang Standard 3 document for accreditation [5].

B. *Data Warehouse*

Data warehouse is a system that retrieves and consolidates data periodically from the source systems into a dimensional or normalized data store. It usually keeps years of history and is queried for business intelligence or other analytical activities. It is typically updated in batches, not every time a transaction happens in the source system [1].

A data warehouse is not a product, but a computing environment where users can find strategic information, an environment where users are put directly in touch with data they need to make better decisions. It is a user-centric environment [2].

C. *Characteristics of Data Warehouse*

The characteristics of data warehouse are:

1) *Subject-oriented*

TABEL I
DIFFERENCE OF OPERATIONAL DATABASE AND DATA WAREHOUSE

	Operational Database	Data Warehouse
Data Content	Current values	Archieve, derived, summarized
Data Structure	Optimized for transactions	Optimized for complex queries
Access Frequency	High	Medium to Low
Access Type	Read, update, delete	Read
Usage	Predictable, repetitive	Ad hoc, random, heuristic
Users	Large number	Relatively small number

2) *Integrated data:* The source data for data warehouse not only come from an operational data but also come from external source. Data inconsistencies are removed it means data from diverse operational applications is integrated.
3) *Non-volatile:* The data in an operational database will periodically move into data warehouse according to the schedule that had been determined, daily, weekly, monthly and the others.
4) *Time-variant:* A data warehouse has to contain historical data, not just current values. Data is stored as snapshots over past and current periods. Every data structure in the data warehouse contains the time element. You will find historical snapshots of the operational data in the data warehouse.
5) *Granularity:* In an operational system, data is usually kept at the lowest level of detail.

D. *Dimensional Modeling (Star Schema)*

The data warehouse used data modeling technique called dimensional modeling. Dimensional modeling is a model based on the query that supports access to high volume query. Star schema is a tool whereby modeling dimensional applied and contains a central fact table. Fact table contains descriptive attributes used to the process of queries and foreign key connecting to table dimensions. Fact table shows what data is supported by the warehouse to an analysis of the decision. Dimension table contains an attribute that outlines the data included in table the fact.

E. *Extract, Transforming, Loading (ETL)*

ETL is a process of taking and send data of data source to the data warehouse. The data in this process must be clean to get good quality data. At the conceptual approach, ETL process is taking data from the data source, put into the staging area, and then transform and load to the data warehouse [1].

F. *On-Line Analytical Processing (OLAP)*

OLAP allows users to analyze database information from multiple database systems at one time. While relational databases are considered to be two-dimensional, OLAP data is multidimensional, meaning the information can be compared in many different ways.

G. *Data Mart*

A data mart is a repository of data gathered from an operational data and other sources that designed to serve a particular community of knowledge workers.

H. *NDS + DDS Data Flow Architecture*

Data warehouse has four data flow architecture, single DDS, NDS+DDS, ODS+DDS, and Federated data warehouse. In NDS+DDS data flow architecture, there are three data stores: stage, NDS, and DDS. This architecture is similar to the single DDS architecture, but it has a normalized data store in front of the DDS. The NDS is in third normal relational form or higher. The purpose of having NDS is twofold. First, it integrates data from several source systems. Second, it can load data into several DDSs. Unlike the single DDS architecture, in the NDS+DDS architecture you can have several DDSs [1].

In the NDS+DDS architecture, NDS is the master data store, meaning NDS contains the complete data sets, including all historical transaction data and all historical versions of master data. The DDS ETL that loads data into the DDS is simpler than the one in the single DDS architecture because the data in the NDS is already integrated and cleaned. The flexibility of using a centralized NDS is that you can build a DDS that you need at any time with the scope of data as required. The ability to build a new DDS at any time is useful to satisfy requirements from projects involving data analysis.

3. METHOD

A. Software

This research uses some software in its implementation, that is:

- SQL Server Management 2014
- Pentaho Kettle – Spoon 4.4.0 Stable Release
- SQL Server Data Tools for Visual Studio 2013
- Microsoft Excel 2010

B. Research Procedure

The steps that will be taken in this research are:

1) *Determine the object of data warehouse:* One characteristic of data warehouse is subject-oriented so that the first step in building data warehouse is to determine the subject.
2) *Define the requirement:* Defining the requirement of the data warehouse and information from the data warehouse for decision support.
3) *Build Data Warehouse:* There are some steps to build a data warehouse. First, designing the architecture and the ETL process. NDS+DDS architecture design use three data store, which is :
   a) Stage, is an internal data store used transforming and preparing the data obtained from the source systems, before the data is loaded to other data stores in a data warehouse.
   b) Normalized Data Store (NDS), is an internal master data store in the form of one or more normalized relational databases for the purpose of integrating data from various source systems captured in a stage, before the data is loaded to a user-facing data store.
   c) Dimensional Data Store (DDS), is a user-facing data store, in the form of one or more relational databases, where the data is arranged in dimensional format for the purpose of supporting analytical queries.

   The purpose of NDS in this scheme not only to combines data obtained from the several source systems, but also it is aimed to make the system would contain data into some DDS.

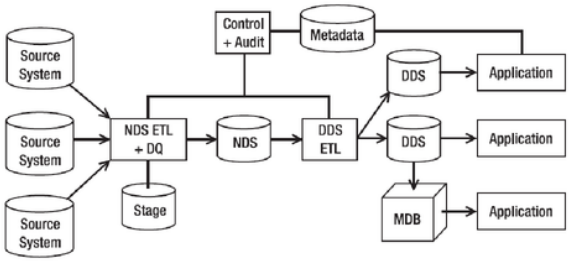

Fig. 1. NDS + DDS data flow architecture

The data warehouse use model dimensional data (star schema). Star schema has two parts, factual table and dimensional table. After the star schema made, the next is ETL (Extract, Transform, and Load) process from an operational database to a data warehouse. Before this process, is made the ETL design first.

The second step of building a data warehouse is ETL process from an operational database to a data warehouse. The ETL process periodically extracts data from the source system, transforming it into a form that is common and then loads it into the data store target (a data warehouse or a data mart). ETL process is very important for data integrating and data warehousing. This research step is based on NDS + DDS data flow architecture. And the last step for building data warehouse is built a cube and dimensional for data warehouse.

4) *Analyze the data warehouse use reporting*

5) *Data warehouse testing:* After the data warehouse is built using the ETL process and analyzed it, the next step is testing the data warehouse. The methods of data warehouse testing are:
   a) ETL Testing, is important because ETL brings the data from the source systems into the data warehouse. If the ETL is incorrect, the data in the data warehouse will be incorrect. Essentially, a data warehouse system has three primary components: ETL, data stores, and applications. Of these three, ETL development takes the largest portion of data warehouse development effort. A general view in the data warehouse community is that ETL takes 60-80 precent of development effort.
   b) Functional Testing is about making sure that all business requirements are satisfied.

4. RESULT AND EXPLANATION

A. Research Subject

The subject in this data warehouse research is student data in Muhammadiyah University of Yogyakarta.

B. User Requirement

The user requirements that referred in this research is the data for an accreditation process with the subject of student. The requirement data obtained from Borang BAN-PT accrediting document.

C. Building Data Warehouse

1) *Data Source:* The data sources used in this research is from some tables in s1makumyny4 database. The following tables are :

- dbo.MAHASISWA
- dbo.STATUS_TERDAFTAR
- dbo.STATUS_TRANSFER
- dbo.CLASS_PROGRAM
- dbo.FACULTY
- dbo.DEPARTMENT
- dbo.AGAMA

2) *Extract, Transform, Load (ETL) Process:* This process cleans the noise from the data source. Noisy data is the data that have random error or variance in a measured variable, such as duplicate records, incomplete data or inconsistent data. We clean the noise by queries. We execute the queries to check the noisy data. And then we take the duplicate records from databases and use a global constant to fill in the incomplete data e.g. "-" or "unknown". This process has three data store, which is:

- Stage, the tables from source systems stored in Stage data store with change the name of tables. Table II displaying

tables name change in Stage data store from source systems.

TABEL III
TABLES NAME CHANGES FROM DATA SOURCE

Data Source	Stage Data Store
dbo.MAHASISWA	dbo.buffer_mahasiswa
dbo.STATUS_TERDAFTAR	dbo.buffer_status_terdaftar
dbo.STATUS_TRANSFER	dbo.buffer_status_transfer
dbo.CLASS_PROGRAM	dbo.buffer_class_program
dbo.FACULTY	dbo.buffer_fakultas
dbo.DEPARTMENT	dbo.buffer_program_studi
dbo.AGAMA	dbo.buffer_agama

- Normalized Data Store (NDS), after the data collected in the Stage, the next step is the ETL process into NDS. In this process the data cleaned noise in each table. Noise can include duplicating data, null data and the others. There was a change of naming tables after ETL process that can be seen in table III.

TABEL IIII
TABLES NAME CHANGES FROM DATA STAGE

Stage Data Store	NDS
dbo.buffer_mahasiswa	dbo.clean_mahasiswa
dbo.buffer_status_terdaftar	dbo.clean_status_terdaftar
dbo.buffer_status_transfer	dbo.clean_status_transfer
dbo.class_program	dbo.clean_class_program
dbo.buffer_fakultas	dbo.clean_fakultas
dbo.buffer_program_studi	dbo.clean_program_studi
dbo.buffer_agama	dbo.clean_agama

The next step is making a relation in NDS. NDS relation uses the seventh tables that have been cleared of noise/dirty data. The NDS diagram relation can be seen in Fig. 2.

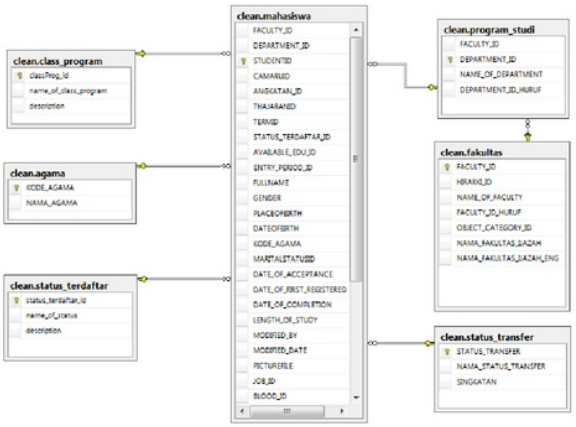

Fig. 2. NDS diagram relation

- Dimensional Data Store (DDS), building DDS needs the ETL process from NDS. Those seven tables turned into a factual table and six dimensional tables. The factual table consists of the measurements, metrics or facts of a business process. There is merger table between dbo.clean.program_studi table and dbo.clean.fakultas table with a newly named dbo.dim_program_study_fakultas. The following is the diagram relation of DDS on Fig. 3.

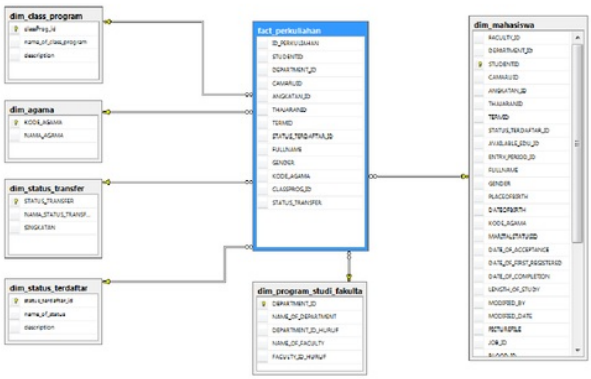

Fig. 3. DDS diagram relation

D. Data Warehouse Analysis

Data warehouse analysis in this research is done by making a report. The reports that have been made are as follow:

- The number of students based on faculty and department report. Table IV is the report that there are faculty and department in UMY with the number of students.

TABEL IV
REPORT BASED ON FACULTY AND DEPARTMENT

Faculty-Department	Total
AGAMA ISLAM	4080
D-II. PAI	23
Ekonomi & Perbankan Islam	1654
Komunikasi & Peny. Islam	471
Pendidikan Agama Islam	1932
EKONOMI	9789
Akuntansi	4112
Ilmu Ekonomi	1513
Manajemen	4164
HUKUM	4097
Ilmu Hukum	4097
ISIPOL	11205
Ilmu Hubungan Internasional	4705
Ilmu Komunikasi	3466
Ilmu Pemerintahan	3034
KEDOKTERAN	7801
Farmasi	525
Ilmu Keperawatan	2122
Kedokteran Gigi	1145
Pendidikan Dokter	4009
PASCA SARJANA	2716
Magister Ilmu Hubungan Internasional	77
Magister Ilmu Hukum	21
Magister Ilmu Pemerintahan	147
Magister Keperawatan	178
Magister Managemen	779
Magister Manajemen Rumah Sakit	633
Magister Studi Islam	881
PENDIDIKAN BAHASA	983
Pendidikan Bahasa Arab	108
Pendidikan Bahasa Inggris	739
Pendidikan Bahasa Jepang	136
PERTANIAN	2138
Agrobisnis	1134
Agroteknologi	1004
POLITEKNIK	645
D3. Akuntansi Terapan	132
D3. Teknik Elektromedik	218

D3. Teknik Komputer Jaringan	65
D3. Teknik Mesin Otomomif dan Produksi	230
PROGRAM DOKTOR	138
Politik Islam	27
Psikologi Pendidikan Islam	111
TEKNIK	9140
Teknik Elektro	1927
Teknik Mesin	2537
Teknik Sipil	4063
Teknologi Informasi	613
Total	52732

- The number of students based on student status report. Table V show some student status in UMY with the number of students.

TABEL V
REPORT BASED ON STUDENT STATUS

Student Status	Total
Eksekutive/Paralel	340
Non Transfer	340
Mahasiswa Baru	172
Melanjutkan Studi dari D3	168
Internasional Program	1116
Non Transfer	1116
Jalur Kemitraan	1
Mahasiswa Baru	1115
Reguler	51276
Melanjutkan dari D3 Non UMY	1
Mahasiswa Baru	1
Non Transfer	51275
Jalur Kemitraan	1153
Mahasiswa Baru	49929
Melanjutkan Studi dari D3	58
Pindah Program Studi	6
Pindahan dari S1 Perguruan Tinggi Lain	61
Tidak diketahui	68
Total	52732

E. Testing

Data mart in the data warehouse require a test. The testing that has been done is :

1) *ETL Testing*: Testing the data is done by comparing the result of data. And the result of comparison, the data is valid and has matched the number of data.
2) *Functional Testing:* This testing carried out by fill the requirement from the Borang document. Table VI displays the data on Borang document.

TABEL VI
TABLES STUDENT PROFILES ON STANDARD 3 BORANG DOCUMENT

Academic Year	Number of Student	
	Regular Non-Transfer	Transfer
TS-4	2453	0
TS-3	3279	0
TS-2	5146	0
TS-1	6750	0
TS	6072	0

5. CONCLUSIONS

The conclusion that can be taken from this research is the data mart that has been designed and built can meet the requirements of the institution. The data mart provide ease of getting information about student for Muhammadiyah University of Yogyakarta. And the data that was obtained from the data warehouse able to answer the requirements about students on Borang documents.

REFERENCE

[1] Rainardi, Vincent, *Building a Data Warehouse with Examples in SQL Server*. New York: Springer-Verlag, 2008.

[2] Ponniah, Paulraj, *Data Warehouse Fundamentals: a Comprehensive Guide for IT Proffesional*. New York: John Wiley & Sons, 2001.

[3] Kurniawan, Novianto Budi, *Desain dan Implementasi Data Warehouse Studi Kasus Pemetaan Daerah Rawan Bencana Provinsi Sumatera Barat*. Bandung

[4] S, Restia Rezalini P., Anggraeni, Wiwik, Wibowo, Radityo Prasetianto. *Perencangan dan Pembuatan Data Warehouse untuk Kebutuhan Sistem Pendukung Keputusan di Bidang Akademik pada Jurusan Sistem Informasi, ITS, Surabaya*. Surabaya: SISFO-Jurnal Sistem Informasi 2010.

[5] Windarto. *Pemanfaatan Data Warehouse sebagai Sarana Penunjang Penyusunan Borang Akreditasi Standar 3 pada Fakultas Teknologi Informasi Universitas Budi Luhur*. Jakarta: Jurnal TELEMATIKA MKOM, Vol.3 No.2. 2011.

Development of Student Data Mart using Normalized Data Store Architecture

ORIGINALITY REPORT

22%	%	22%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"Data Warehouse Architecture", Building a Data Warehouse, 2008 Publication	10%
2	"Testing Your Data Warehouse", Building a Data Warehouse, 2008 Publication	3%
3	"Introduction to Data Warehousing", Building a Data Warehouse, 2008 Publication	2%
4	Ponniah. "Trends in Database Technology", Database Design and Development, 04/04/2003 Publication	1%
5	Benoit Hudzia, Jonathan Sinclair, Maik Lindner. "chapter 80 Deploying and Running Enterprise Grade Applications in a Federated Cloud", IGI Global, 2013 Publication	1%
6	I R H T Tangkawarow, J P A Runtuwene, F I	

Sangkop, L V F Ngantung. "Three Tier-Level Architecture Data Warehouse Design of Civil Servant Data in Minahasa Regency", IOP Conference Series: Materials Science and Engineering, 2018
Publication

1%

7 Kai-Ying Chen, Teh-Chang Wu. "Data warehouse design for manufacturing execution systems", IEEE International Conference on Mechatronics, 2005. ICM '05., 2005
Publication

1%

8 William Yeoh. "chapter 14 Empirical Investigation of Critical Success Factors for Implementing Business Intelligence Systems in Multiple Engineering Asset Management Organisations", IGI Global, 2009
Publication

1%

9 Ponniah. "Data Warehouse: The Building Blocks", Data Warehousing Fundamentals, 08/24/2001
Publication

1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	Off		