Implementing of Data Warehouse Data Alumni using the Single Dimensional Data Store method

Asroni, Titis Bimbi Arimbi, Slamet Riyadi
Information Technology Departement, Universitas Muhammadiyah Yogyakarta, Jl Brawijaya Tamantirto Kasihan Bantul DI Yogyakarta

asroni@umy.ac.id, titis.ft@umy.ac.id, riyadi@umy.ac.id

Abstract. Data is Information that has not been processed or manipulated, which is not enough for some users. So that the info is useful, the data needs to be prepared and integrated first to become Information. Universitas Muhammadiyah Yogyakarta has a tool to monitor data on Alumni, namely Tracer Study. The data currently available is still not integrated and processed correctly. Now there is a solution for processing data and data by making base data. The purpose of creating this database is to facilitate users in checking data, data that will be stored neatly and integrated into the computer. The method in this study uses the Single DDS method.

Methodology This research starts from determining the subject of investigation, primary needs, creating baseline data, analyzing results. The backstage of need is carried out with two data analyzes, and interviews with LPKA (Lembaga Pengelola Kemahasiswaan dan Alumni) or Students and Alumni Management Institutions. Baseline data making in this study is single DDS, which has two data stores, namely Stage and DDS. Data analysis is done by creating reports with Microsoft Office Excel according to user needs. The testing phase is carried out with two processes, namely: ETL Testing and Functional Testing.

1. Introduction
Universitas Muhammadiyah Yogyakarta (UMY) is a private college that has the aim of creating a scholar of believers, devotees, and noble morality who can practice and develop science and technology and useful for the people, nations, and humanity. UMY is known as a college community that always prints qualified Alumni in academia. No wonder if these Alumni can work in large companies both public and private [1].

UMY itself already has/provides information system in the form of Tracer Study as a track record, where Alumni can enter the data of the name, Nim, year of graduation, IPK, the name of the company/institution of work, job title, section, and some additional Other data. With the information about good alumni can also improve the university's right image of the general public and can be used as a college benchmark for the wider community. However, to date, many Alumni of UMY who have not used the facility of the Study Tracer information system provided by UMY expected to make it easier to log and search for alumni.

Information system in the form of Tracer Study is designed to facilitate universities in managing Alumni data. This information system can also be used for various purposes, such as alumni search, activity information, and alumni job news. With so many alumni data going into this Study, Tracer can create opportunities to build a database. Before using the database, the existing data is only stored
without any further management. This makes it difficult to get information about the alumni performance from year to year [2].

The current problem is that databases are not yet used in UMY in monitoring alumni Trace records where still using data stored in database files. This is what makes writers interested in designing databases on Alumni Data. In developing this database, authors use the single-dimensional data store method [3].

The current problem is that databases are not yet used in UMY in monitoring alumni Trace records where still using data stored in database files. This is what makes writers interested in designing databases on Alumni Data. In developing this database, authors use the single-dimensional data store method.

2. Theoritical

2.1. Literature Review

The current problem is that databases are not yet used in UMY in monitoring alumni Trace records where still using data stored in database files. This is what makes writers interested in designing databases on Alumni Data. In developing this database, authors use the single-dimensional data store method:

1. Khusnul Khotimah dan Sriyanto in 2017, Planning titled “Perancangan dan Implementasi Data Warehouse untuk mendukung Sistem Akademik (Studi Kasus pada Sekolah Tinggi Keguruan dan Ilmu Pendidikan Muhammadiyah Kotabumi)”. The method used in this design was Kimball's method with the "nine-Step Methodology" methodology, for the design of data warehouse using the bottom-up approach. The purpose of this design is to serve strategic information such as the total number of new students per batch, student status, and several results based on PGA and predicate. The result of this design is the modeling of star schemes that make data organized, accessible and simplify in the process of querying analysis for data presentation as supporting the evaluation and reporting process [4].

2. Umi Fadilah, Wing Wahyu Winarno and Armadyah Amborowati in 2016 conducted a datawarehouse planning titled “Perancangan Data Warehouse untuk Sistem Akademik Sekolah Tinggi Manajemen dan Ilmu Komputer Kediri” The methods used are action research methods using action research, including Observe stage, Reflect, Plan, and Action. The purpose of using the technique is to plan the data subjects, design the star scheme, create the Data warehouse, and then implement the Extract, Transform methods. The result of this study is the draft of the academic data warehouse following the Indonesian STMIK and produce a report on the number of mahasiswa per generation, per gender, and per city of student origin [5].

3. Other research related to Data Warehouse berjudul “Pengisian Borang Akreditasi Standar 3 Menggunakan Model Data Warehouse” by Deppi Linda in 2015. This research is conducted in order to support the filling of accreditation form especially filling 3 standard form of student data and graduates as well as obtaining information relating to the number of prospective students, the number of new students In the faculty, the number of graduates are grouped by the index type of the cumulative achievement and the percentage. This research conducts Data Warehouse designing with Online Analytical Processing analytical methodology and Business Life cycle development Methods [6].

2.2. Data Warehouse

Data Warehouse consists of the data that has been stored within a specified period and is used to add information. According to Bill Inmon Data Warehouse is a complete/fixed (non-volatile) collection
used to support the decision-making process, which is subject oriented, crossed, and has a time dimension. He also stated that Data Warehouse is the basis of the decision-making process [7].

2.3. Characteristic of Data Warehouse
Data Warehouse According to Connolly and Begg has several characteristics [8]:
1. Subjected-oriented: Data Warehouse that is designed based on a specific subject of the organization, based on the main application area, not by function or application process. This is because the data in the Data warehouse is the data used to make decisions instead of data-oriented towards the application.
2. Integrated: Data Warehouse can store data derived from separate data sources and different applications, into a consistent format and can be integrated, so that the Data is not split and consistent with integrating from Data to the user.
3. Time-variant: Data Warehouse stores historical data or only stores accurate and valid data over a specified period.
4. Non-Volatile: Data Warehouse is not updated in real time and is refreshed from a conventional operating system. The latest Data is always added to the Database itself. The Database will gradually receive and absorb new data, and then it will be merged with the previous data.

2.4. Extract, Transform, Loading (ETL)
The definition of ETL is a process in the formation of a data warehouse. The purpose of the ETL process is to collect, filter, process, and combine the relevant data from various sources to be stored into the Data warehouse[9]. The ETL process consists of Extracting, Transforming, dan Loading. The following describes each process [10]:
1. Extract: The extract is the process of retrieving data extracted from the operational system, can use the query or ETL application itself. This process can last many times to get quality data.
2. Transform: Transform is the second process after the extract process. The data obtained from the extract will be cleaned and modified from the original data to the form of data that corresponds to the standard data Warehouse.
3. Load: The load is the last process by which it is used to insert data that has been transformed into the Data warehouse.

2.5. Architecture Data Warehouse
Data Warehouse has two architectures, namely, data Flow architecture and system architecture. Data flow design is about how to store data is stored in the data warehouse and how data flow from the user source system through this data store. In this study used data flow architecture that is DDS single as in Figure 1 [10].

![Figure 1. Single DDS Architecture](image-url)
On the single DDS architecture, there is only one DDS. The DDS consists of the one-dimensional data mart. The Dimensional Data mart contains a collection of interconnected fact tables and related dimension tables containing measurements. An ETL package extracts data from multiple source systems and puts it on stage. The fact table is a table that contains numerical values or measurable values, such as price, number of items, and others. This fact table also includes a collection of the foreign keys of the primary key in each dimension table. There is historical data on the table. The dimension table is a table containing detailed data describing the foreign key in the fact table. The existing attributes in the table dimension are created hierarchically to facilitate the query process. Map of flood risk level.

3. Method

3.1. Research Procedure

This research is carried out by following the steps used to design the data warehouse. Some of the steps that the author does include determining the subject of the database, defining requirements, creating a data warehouse, analysing the data warehouse by using reporting and the last doing testing on the data warehouse such as Figure 2 Research flowchart.

1. Determine the subject of data acquisition

This study uses the alumni database material owned by Muhammadiyah Yogyakarta University in the form of cdc_web. This alumni database is a transaction-oriented system directly to computers connected to computers or often known as OLTP (Online Transaction Processing) [11].

2. Defining needs

Defining needs is done with two events in the form of document analysis and interviews. At this stage to get a conclusion, an investigation is done by describing and analyzing the document. The research needs are also seen from Study Tracer, a questionnaire filled by alumni. Data derived from surveys filled in by alumni, facilitate researchers in decision making. Interviews are one method to get information and data by asking respondents through direct conversation. Discussions are the central part in answering a problem in research, also considered a complementary approach. At this stage, the researcher interviewed the LPKA (Student and Alumni Development Institution) to obtain accurate and accurate information in making the alumni database.

3. Creating a data warehouse

In Single DDS architecture, there is only one DDS. This DDS consists of one dimension or several data mart dimensions. An ETL packet is extracted from a different data source and then placed in a container or called a Stage. Furthermore, cubes and sizes are made, which is the last stage of the database creation process.
4. Analyse Data Warehouse with reporting

Data Warehouse that has been completed is then analyzed with the needs used to make a report. The cube and dimensions that have been made before are used to create reports. Statements are made with Microsoft Excel. Making reports in Microsoft Excel is done by importing data in the SQL Analysis Services Database. Imported data is a cube and dimensional data that is in SQL, then arranged in a diagram and information that fits the needs needed. The concept used is the concept of OLAP (online analytical processing), which is an agile approach that provides answers to multi-dimensional analytical queries, as shown in Figure 3 [12].

![Figure 3. Example Multidimensional Data Cube](image)

5. Perform Testing

The final step is to test data with ETL Testing and Functional Testing methods. ETL testing is done to process data from operational data sources to the data warehouse. If there is an error that occurs in ETL, then the data in the data warehouse will be wrong, so the data will produce invalid information. Functional Testing is done by checking whether the data can produce information according to needs or not.

4. Result and Explanation

4.1. Need Analysis

Cdc_web database obtained from the UMY Information Systems Bureau and discuss it with the necessary process. In the database Cdc_web that the author can, there are many invalid data. Because of this, then the author performs the process of selecting data. This is done so that the data can fit the required needs. The data selection process is done by discussion. The result, the author uses a table containing the complete data that becomes the point of reference doing the database design

4.2. Data Design

In the design of the database through several phases ranging from determining the source of the data, moving the data source to the stage, continuing the data transfer to the DDS and finally being able to obtain the data in the cube to test the results. To explain the steps in each process will be described as follows:

1. Data source:

Data source in this research derived from the table and views in the database Cdc_web from the University of Muhammadiyah Yogyakarta. The stage of the author is analyzing the needs first. The requirement in this research is the alumni graduate data required in the accreditation process. So to know the user needs in the accreditation process is done method of interview. The interview method that the author does with the informant gets the data needs results that can be seen in table 1.
Table 1. Parameter and indicator to evaluate flood hazard level

| Category                  | Type of Information                  |
|---------------------------|--------------------------------------|
| Alumni Graduate Data      | Alumni profile                       |
|                           | IPK Alumni                           |
|                           | The waiting period gets work          |
|                           | Salary Received                      |
|                           | Conformity of course with employment |

The research subject refers to the data needs of the accreditation process regarding graduates, namely the UMY study tracking data. The issue of this study was focused on alumni tracking data from 2001 to 2014.

2. Stage
At this Stage, the author performs an analysis of the converted table into its Views form. The investigation is tailored to exist needs and data sources. The ETL process is also done at this Stage using the Single DDS architecture, which includes the Stage and DDS. Once the extract process is performed at this Stage, the data is saved into the Stage. A stage is a place or container as the data store that has been extracted. Data change from data source to Stage can be seen in table 2.

Table 2. Parameter and indicator to evaluate flood hazard level

| Source Data    | Stage              |
|----------------|--------------------|
| dbo.alumni     | dbo.st_alumni     |
| dbo.fakultas   | dbo.st_fakultas   |
| dbo.jenis_pekerjaan | dbo.st_jenis_pekerjaan |
| dbo.job_position | dbo.st_job_position |
| dbo.prodi      | dbo.st_prodi      |
| dbo.salary_range | dbo.st_salary_range |
| dbo.send_jobs  | dbo.st_send_jobs  |
| dbo.user       | dbo.st_user       |
| dboUserRole    | dbo.st_UserRole   |

In this Stage, the source data has gone through the first ETL process. Data check is done to find out whether there are duplicated data, incomplete data, or invalid data the dbo.st_alumni table has 643 data and 32 columns, regarding the alumni's data. To do the process can be seen in Figure 1, where the process starts FROM DB Source is moved TO DB Destination after the decline process as per the data requirement as in Figure 4.
3. Dimensional Data Store (DDS)

The data that is used to create this DDS retrieves data from the stage constructed in the previous process. Based on the previous explanation, the specified fields and data will be used in the DDS Data Store. This ETL DDS process uses the SQL Server Data Tools (SSDT) application. The ETL process can be seen in Figure 5.

![Dimensional Data Store Diagram](image1)

**Figure 5.** Proses stage pada tabel dbo.st_alumni

After all the dimension and fact tables are created, the next process is to negotiate the schedules. A fact table relationship Diagram can be seen in Figure 6.

![Fact Table Relationship Diagram](image2)
4. Cube

The data that has been created needs to be organized into another form of multidimensional. Information is stored on databases in a multifaceted way after a multidimensional analysis that provides the ability to query and create reports (reporting). To see that multidimensional data, we should use the cube. Cube is the primary structure of OLAP for viewing data. Here's the data source view in the form cube as in Figure 7 [13].

![Diagram Relational DDS](image)

Figure 6. Diagram Relational DDS

![Relationship of Cube Data Source](image)

Figure 7. Relationship of Cube Data Source

Inside the cube, there are the terms measure and dimension, and Measure is the numeric or quantitative data that traces the value sought, which is calculated by the dimensions of the cube itself. While aspect is the viewpoint of Measure itself, so it can define a transaction. The element itself has a hierarchy or a position whose determination depends on the drill down process and roll up that you want to do when the OLAP process is done. Here are the Measure and dimension of the cube can be seen in Figure 8.
4.3. Data Warehouse Testing

The analysis process in this research is done by reporting. The presentation of the information will be analysed on this database shaped cube in SQL Server Analysis Service (SSAS). Information on the number of graduates/alumni UMY can be seen from various dimensions, namely alumni, Prodi, faculties, departments, sections, salaries, and so forth. The creation of reporting results on the data warehouse, authors use Microsoft Office Excel. Analysis results in Microsoft Excel data can be created with graphs, tables, and diagrams. The analysis needs of the information available on the cube alumni are as follows:

1. Alumni profile
2. Student GPA Value
3. Alumni Waiting period Get jobs
4. Salary Size
5. The relevance of the study

At this stage, the data used to sample there is data on the faculty of FISIPOL. Here is a view of the number of students that can be used for analysis according to the needs as in Figure 9 and Figure 10:

![Figure 8. Measure and Dimensional Cube](image)

![Figure 9. Alumni Salary Display](image)
5. Conclusion

The development of the Alumni database on UMY alumni data using the architecture of Single DDS has been successfully created and implemented to support providing one variable to support the strategic decision for quality improvement UMY graduate students.

References

[1] J. Yusanti and S. Riyadi, “Dashboard-based Alumni Tracer Study Report Using Normalized Data Store Architecture,” presented at the Proceedings of the 2019 8th International Conference on Software and Computer Applications, 2019, pp. 70–74.

[2] F. Rianda and R. Adrian, “Building a Data Mart Using Single Dimensional Data Store Architecture with Student Subject,” Advanced Science Letters, vol. 21, no. 10, pp. 3391–3394, 2015.

[3] R. Z. Bakhtiar and A. Asroni, “Implementasi Arsitektur Operational Data Store (ODS) dan Dimensional Data Store (DDS) dalam Pembangunan Data Mart Lulusan,” Semesta Teknika, vol. 18, no. 1, pp. 55–64, Jan. 2016.

[4] K. Khotimah, “Perancangan Dan Implementasi Data Warehouse Untuk Mendukung Sistem Akademik (Studi Kasus Pada STKIP Muhammadiyah Kotabumi),” Jurnal Teknologi Informasi Magister, vol. 2, no. 01, pp. 94–107, 2017.

[5] U. Fadilah, W. W. Winarno, and A. Amborowati, “Perancangan Data Warehouse Untuk Sistem Akademik STMIK Kadiri,” SISFOTENIKA, vol. 6, no. 2, 2016.

[6] D. Linda, “Pengisian Borang Akreditasi Standar 3 Menggunakan Model Data Warehouse (Studi Kasus Ibi Darmajaya),” Jurnal Informatika, vol. 15, no. 2, pp. 89–101, 2016.

[7] B. Inmon, “BUILDING THE DATA WAREHOUSE: GETTING STARTED,” p. 19, 2000.

[8] T. Connolly, “dan Begg, C. 2002,” Database Systems: A Practical Approach to Design, Implementation, and Management.

[9] M. Madhikermi, A. Buda, B. Dave, and K. Främling, “Data Model Logger-Data Discovery for Extract-Transform-Load,” presented at the 2017 IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 2017, pp. 629–630.

[10] V. Rainardi, Building a data warehouse: with examples in SQL Server. John Wiley & Sons, 2008.
[11] C. Yu, “Sale Decision Analysis System Design Based on Data Warehouse,” presented at the 2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 2016, pp. 468–471.

[12] J. M. P. Martinez, R. Berlanga, M. J. Aramburu, and T. B. Pedersen, “Integrating data warehouses with web data: A survey,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 7, pp. 940–955, 2008.

[13] B. Golub, A. Hudz, A. Dudnyk, and A. Bushma, “Production of Biotechnological Objects using Business Intelligence,” presented at the 2019 9th International Conference on Advanced Computer Information Technologies (ACIT), 2019, pp. 200–204.