

# Tableau Business Intelligence Using the 9 Steps of Kimball's Data Warehouse & Extract Transform Loading of the Pentaho Data Integration Process Approach in Higher Education

Indrabudhi Lokaadinugroho<sup>1\*</sup>, Abba Suganda Girsang<sup>2</sup>, Burhanudin<sup>3</sup>

<sup>1,3</sup> Computer Science Department, School of Computer Science,

<sup>2</sup> Computer Science Department, BINUS Graduate Program - Master of Computer Science,  
Bina Nusantara University,  
Jakarta, Indonesia 11480

indrabudhi.lokaadinugroho@binus.ac.id; burhanudin@binus.ac.id; agirsang@binus.edu

\*Correspondence:indrabudhi.lokaadinugroho@binus.ac.id

**Abstract** - This paper discusses about how to build a data warehouse (DW) in business intelligence (BI) for a typical marketing division in a university. This study uses a descriptive method that attempts to describe the object or subject under study as it is, with the aim of systematically describing the facts and characteristics of the object under study precisely. In the elaboration of the methodology, there are four phases that include the identification and source data collection phase, the analysis phase, the design phase, and then the results phase of each detail in accordance with the nine steps of Kimball's data warehouse and the Pentaho Data Integration (PDI). The result is a tableau as a tool of BI that does not have complete ETL tools. So, the process approach in combining PDI and DW as a data source certainly makes a tableau as a BI tool more useful in presenting data thus minimizing the time needed to obtain strategic data from 2-3 weeks to 77 minutes.

**Keywords:** Tableau; Business intelligence; Data warehouse; Nine steps of Kimball; ETL; Pentaho Data Integration; Higher education.

## I. INTRODUCTION

Organizations know very well that the vital information for decision-making lies in its operational databases. A large amount of data gets accumulated in various databases scattered around the enterprise. But the key to gaining a competitive advantage lies in deriving insight and intelligence out of this data (Silva, 2005). The concept of data warehousing and business intelligence is one alternative in harnessing the value and intelligence imbedded within those databases. During the last decade, organizations have invested a large amount of resources in building and implementing data warehouse and business

intelligence systems after realizing a positive return on investment that justified its existence (Silva, 2005). Studies show that only a small number of universities really implement a data warehouse to get an advantage in this tight competition (Golfarelli, 2010).

Previous studies have identified several major reasons related to this issue. Rudy and Miranda (2015) found that elements of globalization in the world of education have expanded and developed, and this has made a higher education institution market that has been developed as a global phenomenon, so that higher education institutions are aware of the importance to show their existence in global and high competition. Other studies reveal a design model for building data warehouse for a typical university information system. It is based on transforming an operational database into an informational warehouse useful for decision makers to conduct data analysis, predication, and forecasting (Bassil, 2012). Moreover, the management can use the data warehouse system to monitor their organizational marketing performance and to support the decision-making process (Rudy & Miranda, 2015). By implementing a data warehouse, a university can be more focused on what its students need and help its students to achieve their goals in the university.

The large variety of BI solutions on the market makes it very difficult for organizations to select one and evaluate the impact of the selected solution to the organization. The need of a strategy to help organization chose the best solution for investment emerges. In the past, BI market was dominated by closed source and commercial tools, but in the last years open source solutions developed everywhere. An open source business intelligence solution can be an option due to time-sensitive, sprawling requirements and tightening budgets (Tarnaveanu, 2012). Then the business intelligence should enable universities

to measure, monitor, and manage their performance more effectively (Muntean, Bologa, Bologa, & Alexandra, 2016).

The object of this study is the Marketing Division of BINUS University as one of the leading private universities in Indonesia with a growing number of students and its vision and mission to become a world-class knowledge institution, continues to implement technology to assist its operations. So that it can continue to compete with other competitors, especially through implementing data warehouse and business intelligence. Hopefully, this will reduce the time taken by Marketing Intelligence in preparing strategic data which previously took 2-3 weeks.

### 1.1 Literature Review

- Data Warehouse

According to Kimball and Ross, a data warehouse is a conglomerate of all data marts within the enterprise. Information is always stored in a dimensional model. Kimball views data warehousing as a constituency of data marts. Data marts are focused on delivering business objectives for departments in the organization. A data warehouse is a conformed dimension of the data marts. Kimball and Ross describe a data mart as being a subset of a data warehouse. A data warehouse is the sum of all the data marts, each representing a business process in an organization by the means of a star schema, or a family of star schemas of different granularity (Kimball & Ross, 2002).

- Fact Table

A fact table is a specialized relation with a multi-attribute key and contains attributes whose values are generally numeric and additive. A dimension table has a single attribute primary key (usually surrogate) that corresponds exactly to one of the attributes of the multi-attribute key of the fact table (Sen & Sinha, 2005). Fact tables contain many rows and relatively few columns; this is essential for ease of use and query performance.

- Star Scheme

The characteristic star-like structure of the physical representation of a dimensional model is called a star join schema, or simply a star schema. A dimensional model can be extended to a snowflake schema, by removing the low cardinality attributes in the dimensions and placing them in separate tables, which are linked back into the dimension table with artificial keys (Kimball, Reeves, Thornthwaite, Ross, & Thornwaite, 1998).

- Dimension Table

A dimension table has a single attribute primary key (usually surrogate) that corresponds exactly to one of the attributes of the multi-attribute key of the fact table (Sen & Sinha, 2005).

- ETL System

Achieving real-time data warehousing is highly dependent on the choice of a process in data warehousing technology known as extract, transform, and load (ETL). This process involves: 1) extracting data from outside

sources; 2) transforming it to fit operational needs; and 3) loading it into the end target (database or data warehouse) (Kakish & Kraft, 2012).

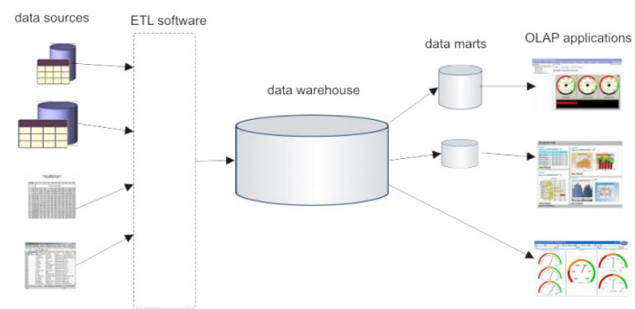


Figure 1. Data Warehouse Architecture (Waas, et al., 2013)

As shown in Figure 1, a traditional BI infrastructure is typically composed of five (5) layers:

1. Operational Data Sources (ODSs) layer
2. ETL layer
3. Data Warehouse
4. Data Marts
5. Online Analytical Processing (OLAP)

An operational data sources (ODSs) layer represents heterogeneous and distributed data sources. ODSs can include various relational databases as well as non-database storage systems. The second layer, an ETL layer, is responsible for loading data into a central database, called a data warehouse (DW) (Kimball & Caserta, 2004). The third layer includes an already mentioned data warehouse that stores integrated and summarized data. The fourth layer is composed of repositories (called data marts), typically storing subsets of aggregate data from the central DW. Each data mart includes data specific to a given business domain, e.g., marketing, finance, and HR (Golfarelli & Rizzi, 2009). The fifth layer — an online analytical processing (OLAP) layer, is responsible for various types of data analysis and visualizations (e.g., reporting, dashboards, management cockpits). OLAP applications typically execute complex queries to discover trends, patterns of behavior, and anomalies as well as for finding hidden dependencies between data. There are multiple source systems, a staging area where data is extracted to, a central warehouse for storing all historical data, and finally data marts that enable end users to work with the data. Between each of these building blocks, a data integration process is used, as shown by the ETL blocks (Casters, Bouman, & Dongen, 2010).

- Business Intelligence

The dashboards are the preferred method for delivering and displaying business intelligence to users. A university can use dashboards as the main components of a business intelligence solution (Muntean, Bologa, Bologa, & Alexandra, 2016). Eckerson defines a dashboard as “a multilayer application built on a business intelligence and data integration infrastructure that enables organizations to measure, monitor, and manage business performance more effectively” (Eckerson, 2006).

## II. METHODS

This study used a descriptive method, which is a research method that attempts to describe the object or subject under study as it is, with the aim of systematically describing the facts and characteristics of the object under study precisely.

In the elaboration of the methodology, there are three phases that include the identification and source data collection phase, the analysis phase, the design phase, and then following the phase results of each detail in accordance with the Kimball nine-step design methodology and Pentaho Data Integration.

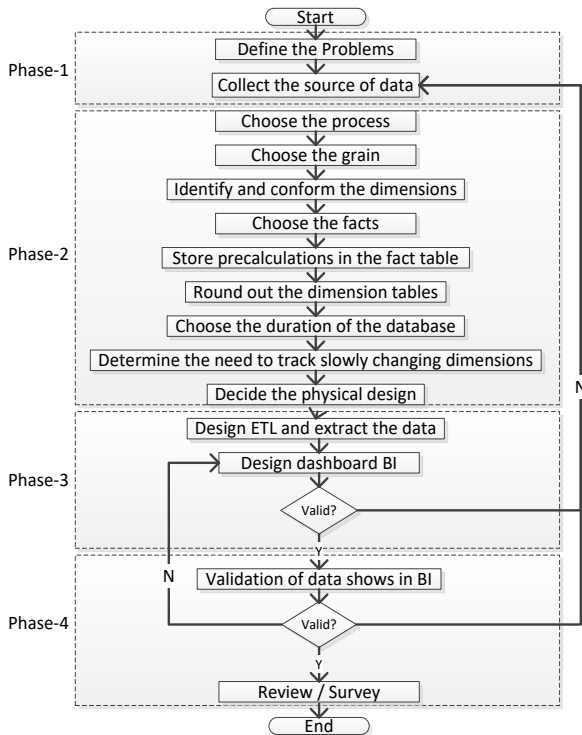


Figure 2. Design Process Approach of DW, ETL, and BI (Source: Authors)

### 2.1. Phase-1: The Identification and Source Data Collection Phase

The first step in this research was to formulate a problem. BINUS as a private higher education institution does not have DW as a single source of data, which results in BINUS operations often having problems with data differences between one application and another application. For example, when the admissions staff draws data from the application used, it will have a different report summary amount than the data released by Datacenter. Worse, every time there is a difference, it is very difficult to find the difference and it takes a long time just to find out what causes the dissimilarity in data differences. Another problem is the report that one wants to use when one wants to extract the data takes a long time because the database used to present the report is an operational database which is not designed for analytical reports.

This research focused on the marketing and admissions services, where DW and BI were made only to

meet their needs, especially in presenting reports that could be analyzed easily, quickly, and accurately.

### 2.2. Phase-2: The Analysis

In this phase, the Kimball nine-step methodology was used to create DW (Kimball & Ross, 2010). Kimball follows the mixed (top-down as well as bottom-up) strategy of data design. The goal is to create individual data marts in a bottom-up fashion, but in conformance with a skeleton schema known as the “data warehouse bus”. The data warehouse for the entire organization is the union of those conformed data marts (Sen & Sinha, 2005).

In choose the process step, Kimball said that to be successful in implementing a DW, the first step that must be done is to determine what business processes we want to make into DW. In this research, a marketing analysis process was one of the processes implemented, where the scope of the work was: student recapitulation and registrar profile.

In choose the grain step, in his book, Kimball stated that the grain here is what data we want to fill in the fact table that we will create. So, in this research the fact table that was created was filled with the profiles of prospective students who purchased forms from BINUS university or also called ‘applicants’ and ‘form sold’.

Identify and confirm the dimensions, in this step, the dimensions were determined to create or bring into DW that would be implemented. In this research, there were several dimension tables related to the marketing of BINUS University.

Choose the facts, from the second step of “choose the grain”, it was already decided that the applicant data would be used as the fact table. Besides that, the study kept track of another fact table that stored all the sales forms to find out the total number of forms sold.

In store pre-calculations in the fact table step, the items that were to be stored in a pre-calculation in the fact table were the status of the form sold, the status of the applicant, the status following the test, the status of passing the test, and re-registering until the status of resigning after they re-registered because of something or another condition. To speed up the calculation process and when it would be displayed in the BI dashboard, instead of we sum the total data during open the BI dashboard we do pre-calculations in the fact table that stored the total data per province, city, and year was divided into the table pre-calculation fact itself.

In round out the dimension tables step, after the fact table and dimension table were prepared from the previous step, then the next step was to provide an attribute description for each dimension table prepared previously.

Choose the duration of the database step, in order that the DW implementation was completed properly, at this stage the time limit for the database or data that was to be processed was determined. In this research, the duration of the data was analyzed and processed.

In determine the need to track slowly changing

dimensions step to maintain data consistency, slowly changing dimensions must be captured as early as possible. Kimball identified three types of slow changing dimensions:

1. Type 1 - Overwrite: or by directly overwriting existing data with new data.
2. Type 2 - Add a new dimension record: or by adding new data into dimensions with a new key and retaining old data with the old key.
3. Type 3 - Add a new field: if old and new data must be used simultaneously then type 3 is the most ideal type.

In decide the physical design step, because the source of the data came from several databases and a number of the legacy databases were made by different people, it caused its own constraints in combining the databases. So, a staging database was needed to accommodate and equalize the attributes of each data source. Then a cleansing process had to be performed before the data was channeled into DW.

### 2.3. Phase-3: Design

In this phase, the ETL process and BI dashboard were designed as user requirements. The user needs were aligned with the available data.

In design ETL and extract data step, in this process, all the data sources were processed into the DW. From the operational database source, the data was extracted using ETL tools. Afterwards, it was cleansed in the staging table. Then the final data was entered into the dimension table or fact table depending on what data was processed. The ETL process in this research used 'Pentaho Data Integration (PDI) version 6'. The data was taken from operational databases and Excel files obtained from the Marketing staff, which was then combined to become data that was meaningful in DW. Data Warehouse is a special database that is used as a "data warehouse" or data which has been consolidated from various data sources of existing information systems in an organization/company (Salaki, Tangkawarow, & Waworuntu, 2015).

In design BI dashboard step, in order to make the data that had been transferred into DW become more useful and able to be used as a decision maker, the BI dashboard was designed as a visualization tool so that it was easier to analyze. BI's dashboard design was made using a modern BI tool with the name 'Tableau 10' which processes data sources from DW, so that they can be displayed as meaningful data. Tableau Public is a web-based software and the open-source version of Tableau Desktop. This tool allows the creation of interactive visualizations and the option to embed them in a website, as well as publish them on Tableau Public Gallery or share them in the Tableau Public Community (Brandão, et al., 2016).

### 2.4. Phase-4: Results

This In this phase, the results of the BI dashboard were validated as a strategic alignment with the Marketing Intelligence requirements.

In validation of data shown in BI step, after the data was visualized in the BI application, it was re-validated

before being released to Operations. The data that had been validated was tested together with the operational staff, who in this study were the Marketing staff. If the testing results were approved by the user, the BI dashboard would be released to Operations so that it could be used as an analysis, for monitoring, and as a decision maker. When the testing process found perceptual mismatches, in this stage it was possible to revise the previous stages again.

Review / survey step, the final step in this study was to evaluate the results of the DW and BI implementation that had been released to Operations. Evaluations and reviews were carried out by measuring the processes that were running in presenting data, whether DW and BI could speed up the reporting process carried out by Marketing Intelligence so that decisions that needed to be made could be faster and more accurate, or if there were still things that needed to be improved and changed again. If there were still things that needed to be changed, it needed to be revised to the previous stage. The target to be produced in this research was the acceleration in producing reports made by Marketing Intelligence.

In addition to determining whether this design and implementation was successful or not, a questionnaire was created which was filled in by 3 BI users on the dashboard who were from the Marketing Intelligence section, which usually processes marketing data to be presented to the Marketing Supervisor. The questionnaire consisted of 7 questions, of which the questions used 5 types of assessment weights as revealed in Table 1 below:

Table 1. Success Rate of DW and BI Implementation (Source: Authors)

Total Average Score	Result
4.0 – 5.0	Success
3.0 – 3.9	Failure
< 3.0	Failure

## III. RESULT AND DISCUSSION

From the analysis and design phase, the authors tried to identify the best practice needed to develop the DW and BI of Marketing at BINUS University, so that it could achieve its goals and improve the efficiency and effectiveness of operational activities. Based on the results of the analysis and design phase previously, it was possible to formulate recommendations as follows.

### 3.1. Marketing Data Warehouse Design

Previously, the business processes were chosen that were needed to be transformed into DW, which were: the student recapitulation and registrar profile. From those business processes, the necessary fact tables were created, which were *applicant* and *formsold* as follows:



Figure 3. Fact Table  
(Source: Authors)

From those fact tables, then the dimensions needed were decided. There were 16 table dimensions used to support the grain:

1. D\_Period – Dimension that keeps the period.
2. D\_CampusLocation – Dimension that keeps the location of the campus that the university has.
3. D\_Province – Dimension that keeps the province data.
4. D\_City – Dimension that keeps the city data.
5. D\_StudentProfile – Dimension that is used for keeping the status of the applicant, and whether he/she passes the entrance test or not.
6. D\_BulkSchoolMapping – Dimension that is used to map all schools that contribute the most to the university.

7. D\_Agency – Dimension that keeps the agents of the marketing people who sell the registration forms.
8. D\_EnrollmentTrack – Dimension that keeps the history from where the students enroll to the university, such as from their school, the Admissions people, or certain events that are held by the university.
9. D\_ScholarshipType – Dimension that keeps the scholarships that are given to the applicants if they do re-registration to become students.
10. D\_Program – Dimension that keeps the programs that the students choose.
11. D\_Major – Dimension that keeps the majors that the students choose.
12. D\_Degree – Dimension that keeps the degree of which the students apply (undergraduate or graduate studies).
13. D\_Resign – Dimension that keeps the history of those who do re-registration and become students but resign for certain reasons.
14. D\_HighSchool – Dimension that keeps high school data from where the applicants came.
15. D\_Institution – Dimension that keeps the university branch if they have more than one branch.
16. D\_Exam – Dimension that keeps the results of the entrance exams of the applicants.
17. D\_ProgramGroup – Dimension that keeps the group of the program (faculty).

A detailed explanation from the applicant and formsold fact tables and what types of data were stored in the fact tables can be seen in Table 2:

Table 2. Detailed Description of the Applicant & Formsold Fact Tables  
(Source: Authors)

Name	Data Type	Remarks
ApplicantID	Int	Applicant ID key of the fact table (running number, not null).
PeriodID	Int	Period ID according to the Dimension table.
Batch	Char(4)	Batch code where the applicant registers, usually divided based on events carried out by the marketing and entrance examinations.
FormNumber	Char(12)	Form number when the applicant buys the form at Admissions.
UserName	Varchar(100)	The name used to log in via online admission. It is generally in accordance with the email address used by the applicant.
Name	Varchar(100)	Applicant's name.
PlaceOfBirth	Varchar(100)	Applicant's place of birth.
DateOfBirth	Datetime	Applicant's date of birth.
Gender	Char(1)	Applicant's gender.
Citizenship	Char(4)	Applicant's citizenship.
ProgramID	Char(4)	Program or major code chosen by the applicant.
MajorID	Char(3)	Specialization code chosen by the applicant, only for certain applicants.
DegreeID	Char(1)	Code for the level of academic career.

Name	Data Type	Remarks
GroupProgramID	Char(2)	Group codes of majors taken, whether Regular, International, or Online Learning.
InstitutionID	Char(2)	The institutional code chosen, because currently Binus has more than one (1) institution, such as BINUS Kemangisan, BINUS Bandung, or BINUS Malang.
CampusLocationID	Char(3)	Campus location code chosen by the applicant.
HighSchoolID	Char(10)	Code of the applicant's high school.
CityID	Char(5)	City code for the applicant's school.
ProvinceID	Char(5)	Provincial code of the applicant's school.
EnrollmentTrackID	Char(2)	The path chosen by the applicant, whether a scholarship or regular registration.
DaftarOnline	Char(1)	Whether the applicant registers online or manually through Admissions.
AgencyID	Char(5)	Agent code that gets applicants or makes them interested in registering with BINUS.
Year	Char(4)	The year the applicant registered with BINUS.
FormSold	Int	Sales Status Form, default = 1, because all applicants have definitely purchased the registration form.
Applicant	Int	The status of registering to become an applicant to take the entrance screening test.
Test	Int	Status where the applicant takes the entrance screening test or not.
Lulus	Int	Status where the applicant passes the entrance examination test or not.
DaftarUlang	Int	Status where the applicant re-registers or not.
UndurDiri	Int	The status where the applicant resigns after re-registering or not.
ProcessDate	DateTime	Date of data in process.

To speed up the calculation process and when displayed in the BI dashboard, the fact table that stores the total status per province, city, and year was divided into the pre-calculation fact table with the table name F\_SUM\_MAR or it stands for summary marketing table. As seen in Figure 4, the summary marketing fact table, this marketing fact table was created with a de-normalized model to speed up the process of data display on the BI dashboard.

F_SUM_MAR		
PeriodID	Int	NN (PFK)
Year	Char(4)	
ProvinceID	Char(5)	NN (PFK)
ProvinceName	Varchar(100)	
CityID	Char(5)	NN (PFK)
CityName	Varchar(100)	
FormSold	Int	
Applicant	Int	
Test	Int	
Lulus	Int	
DaftarUlang	Int	
UndurDiri	Int	
ProcessDate	Datetime	

Figure 4. Summary of the Marketing Fact Table  
(Source: Authors)

In addition to the applicant fact table, there is also a star scheme in the fact summary marketing table containing the total number of students per student profile per city per province de-normalized, as can be seen in Figure 5 below:

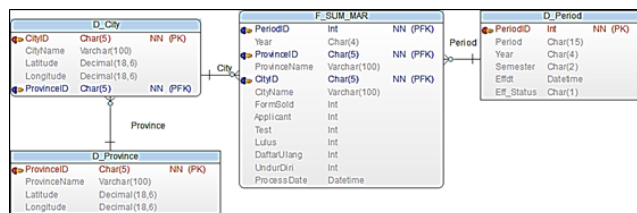


Figure 5. Star Scheme of the Summary of Marketing  
(Source: Authors)

The next step in this process was to provide an attribute description of each dimension table prepared previously, as in Figure 6 below:

D_StudentProfile dimension			D_Institution dimension		
No	Attribute	Description	No	Attribute	Description
1	FormNumber	Sales form number	1	InstitutionID	Key from institution
2	Re-registrationDate	Date of registering as a student	2	InstitutionName	Name of differences between BINUS Jakarta, Bandung and Malang
D_Agency dimension			D_Major dimension		
No	Attribute	Description	No	Attribute	Description
1	AgencyID	ID of agency agency that makes a sale	1	MajorID	ID from Major
2	AgencyName	Name of agency	2	MajorName	Name of Major
3	AgencyGroupID	ID of group of agency	3	DegreeID	ID from Degree
D_ProgramGroup dimension			D_Degree dimension		
No	Attribute	Description	No	Attribute	Description
1	ProgramGroupID	ID Group from Faculty	1	DegreeID	ID from Degree
2	ProgramGroupName	Name of Faculty Group	2	DegreeName	Name of Degree
D_Campus:Location dimension			D_HighSchool dimension		
No	Attribute	Description	No	Attribute	Description
1	Campus LocationID	ID of campus	1	HighSchoolID	ID of high school
2	Location	Campus location	2	HighSchoolName	Name of high school
3	Latitude	Coordinate latitude	3	CityID	ID of city
4	Longitude	Coordinate longitude	4	ProvinceID	ID of Province
D_ScholarshipType dimension			D_Province dimension		
No	Attribute	Description	No	Attribute	Description
1	ScholarshipTypeID	ID of Scholarship	1	ProvinceID	ID of Province
2	ScholarshipType Name	Type of scholarship name	2	ProvinceName	Name of province
3	EnrollmentTrackID	ID of Enrollment track	3	Latitude	Provincial latitude coordinates
4	Grade	The results of the entrance test	4	Longitude	Provincial longitude coordinates
D_Period dimension			D_BulkSchoolMapping dimension		
No	Attribute	Description	No	Attribute	Description
1	PeriodID	Surrogate key from Period	1	BulkSchoolMappingID	ID Bulk mapping of school
2	Period	Applicant's entry period	2	HighSchoolID	ID of high school
3	Year	Applicant's year of entry	3	BulkGroupID	ID Bulk group
4	Semester	The applicant enters the semester	4	PeriodID	Surrogate key from Period
5	Effdt	Effective period	5	CityID	ID of city
6	Eff_status	Status period active or inactive	6	ProvinceID	ID of province
D_EnrollmentTrack			D_City dimension		
No	Attribute	Description	No	Attribute	Description
1	EnrollmentTrackID	ID of Enrollment track	1	CityID	ID of city
2	EnrollmentTrackName	Name of enrollment	2	ProvinceID	ID of province
3	Scholarship	Name of scholarship	3	CityName	Name of city
			4	Latitude	City latitude coordinates
			5	Longitude	City longitude coordinates
D_Exam					
No	Attribute	Description			
1	FormNumber	Sales form number			
2	Grade	The results of the entrance test			
3	StatusLulus	Pass / Not Pass the Exam			

Figure 6. Attribute Description of Dimensions  
(Source: Authors)

Then the duration of the database was set from 2012 to 2016. The data was determined from 2012 because in the previous year there was no data stored in the Admissions database, so it was not possible to analyze it. The next step was to determine the need to track slowly changing dimensions in Table 3 as follows:

Table 3. Explanation of Slow Changing Dimensions per Dimension Table  
(Source: Authors)

No	Dimension Name	Remarks	Type
1	D_Period	To keep the naming period when looking at backward data and then add a new dimension record with an effective date	2
2	D_CampusLocation	History needs to be maintained	3
3	D_Province	History needs to be maintained	3

4	D_City	History needs to be maintained	3
5	D_StudentProfile	History does not need to be maintained	1
6	D_BulkSchoolMapping	History does not need to be maintained	1
7	D_Agency	History does not need to be maintained	1
8	D_EnrollmentTrack	History does not need to be maintained	1
9	D_ScholarshipType	History does not need to be maintained	1
10	D_Program	History does not need to be maintained	1
11	D_Major	History does not need to be maintained	1
12	D_Degree	History does not need to be maintained	1
13	D_Resign	History does not need to be maintained	1

14	D_HighSchool	History does not need to be maintained	1
15	D_Institution	History does not need to be maintained	1
16	D_Exam	History does not need to be maintained	1
17	D_ProgramGroup	History does not need to be maintained	1

Then the physical design for DW was determined. DW's physical design for this research is shown in Figure 7. Because the source of the data came from several databases and the number of legacy databases was made by different people, it caused its own constraints in combining the databases. So, it needed a staging database to accommodate and equalize the attributes of each data source, and then a cleansing process was performed before the data was channeled into DW.

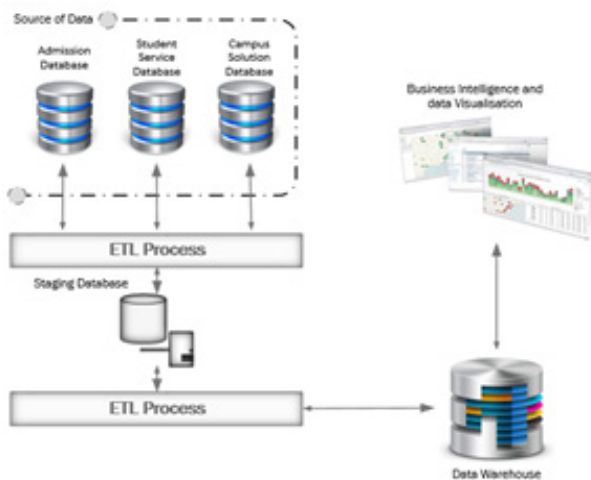


Figure 7. Physical Design of a Data Warehouse (Source: Authors)

The hardware and servers that were used in this study had specifications as shown in Table 4, Table 5, and Table 6 as follows:

Table 4. Specification of DWH Server Database

<b>Server Name</b>	DWH - DB
<b>Operation System</b>	Windows Server 2012R2
<b>Processor</b>	8 vCPU (Intel® Xeon® E5-2630v4 @2.20GHz)
<b>Memory</b>	8192 Mb RAM
<b>Hard Disk Size</b>	250Gb

Table 5. Specification of ETL Server

<b>Server Name</b>	DWH – Pentaho
<b>Operation System</b>	Windows Server 2012R2
<b>Processor</b>	8 vCPU (Intel® Xeon® E5-2630v4 @2.20GHz)
<b>Memory</b>	8192 Mb RAM
<b>Hard Disk Size</b>	200Gb

Table 6. Specification of BI Server

<b>Server Name</b>	DWH – Tableau Server
<b>Operation System</b>	Windows Server 2012R2
<b>Processor</b>	8 vCPU (Intel® Xeon® E5-2630v4 @2.20GHz)
<b>Memory</b>	8192 Mb RAM
<b>Hard Disk Size</b>	200Gb

The DWH - DB server was used as a DW database, while for ETL processes a different server was used, namely DWH - Pentaho, to maintain the performance of the ETL process itself and the performance of DW when accessed by application users or BI. In contrast, the DWH - Tableau Server was used for the BI dashboard. As seen from the memory of the BI server, it has the largest 32 gigabyte RAM. This was done because the feature in Tableau as BI Tools can be used In-Memory Computation when processing and displaying data, so that it requires larger memory than ETL servers and databases.

There were three important software programs in this research outside of the other supporting applications as described in Table 7. The software list is as follows:

Table 7. Software List

<b>Database for DW</b>	Microsoft SQL Server 2014
<b>ETL Tools</b>	Pentaho Data Integration 6
<b>Business Intelligence Tools</b>	Tableau 10
<b>Other Supporting Software</b>	Java Development Kit (JDK) version 1.7.79 Web Browser (Google Chrome or Mozilla Firefox)

For this research, the recommended web browser to use was Google Chrome and Mozilla Firefox. Web browsers such as Internet Explorer (IE) could not be used because Tableau Software is not supported to be opened using IE.

### 3.2. ETL Process and Transfer Database

After the DW was prepared in the nine-step design methodology, the next step was to design the ETL process. In this research, the repository data was taken from four different sources, namely from the Oracle database that used SAS campus solution applications, Application Admissions databases (AOB), legacy databases used by desktop application applications, and other sources that were still in the form of an Excel file that were extracted into DW, as seen in Figure 8:

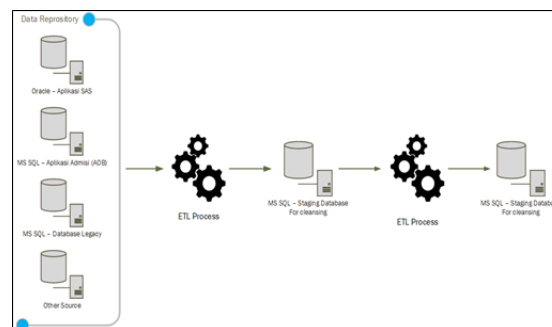


Figure 8. ETL Process Details (Source: Authors)



In general, for data sources, especially old data, it was very unclear because there was no validation from the application side. Therefore, data cleansing and mapping was needed. In the process of data cleansing and mapping, there had to be changes in the ETL process, so that it would be possible when this stage returned to the ETL process again. For the cleansing process to be repeated, the mapping of unclear data was needed. Mapping could also be used to standardize data; for example, the name of the city and the surrounding area when entering the BI dashboard had to be standardized in the city list according to the BI application.

Table 8. Table Total Running Time for ETL  
(Source: Authors)

No	Process	Duration (minutes)
1	Reload dimension tables and facts	60
2	Job marketing metric	17
<b>Total</b>		<b>77</b>

The total time required in running the ETL process can be seen in Table 8 above. It is quite different from 2 weeks (before) and then 77 minutes.



Figure 9. ETL Process Details for BI Using Pentaho

In Figure 9, it is seen that many stages were done to have DW for BI, and it took 77 minutes to do the ETL process.

### 3.3. BI Dashboard

In designing a BI dashboard, there are a lot of obstacles encountered in how to display data, so additional ETLs must be made as done in ETL F\_MARKETING (2). Additional ETLs should be made because when designing the BI dashboard the data cannot be displayed as expected, so that a container table must be created to hold pre-calculation data.

A dashboard marketing summary was used to monitor the total number of forms sold, as well as the applicants, who were already tested, passed, and re-registered, or intakes until they resigned. The dashboard can see details up to data per province and per city. The data source was taken from the F\_SUM\_MAR table which was processed in the ETL job F\_SUM\_MAR. A dashboard example for a marketing summary can be seen in Figure 10 to visualize the marketing summary:

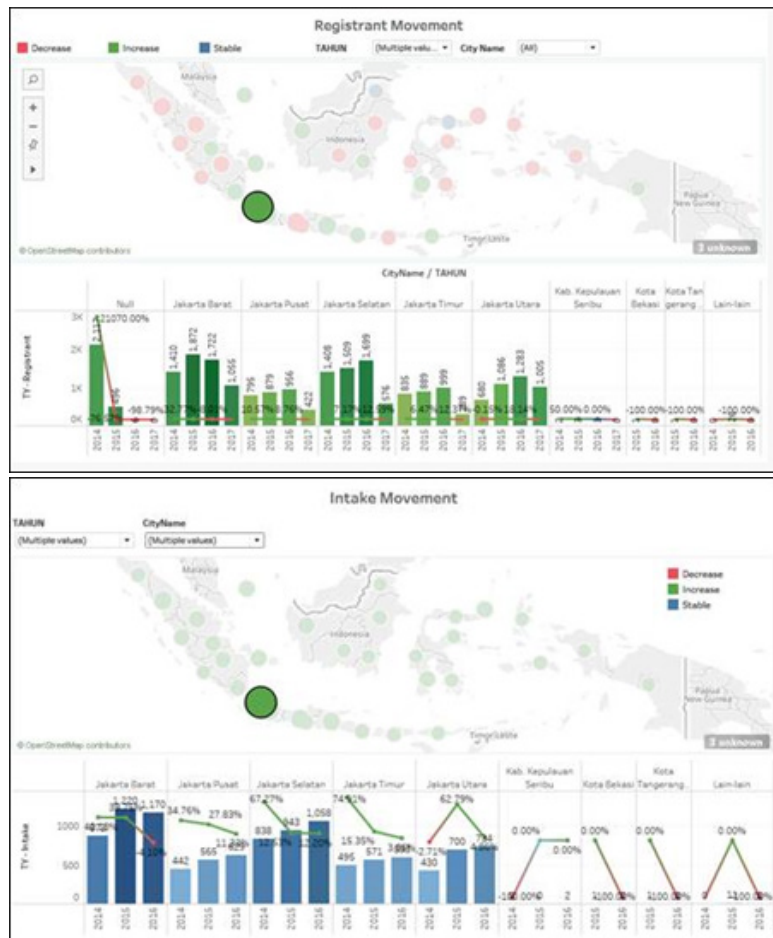


Figure 10. Visualization of Tableau Business Intelligence of the Marketing Division (Registrant Movement vs. Intake Movement)  
(Source: Authors)

As seen in the dashboard, the data can be viewed in detail per province by selecting the desired province to be seen. The Provincial Hierarchy can be setup on BI Tools where the data is in groups based on Year, ProvinceName, and CityName. This hierarchy can be unique per sheet dashboard according to one's wishes, but this study had a hierarchy like in each of its dashboard sheets. This hierarchy can be seen in Figure 11, which is setup in Tableau 10:

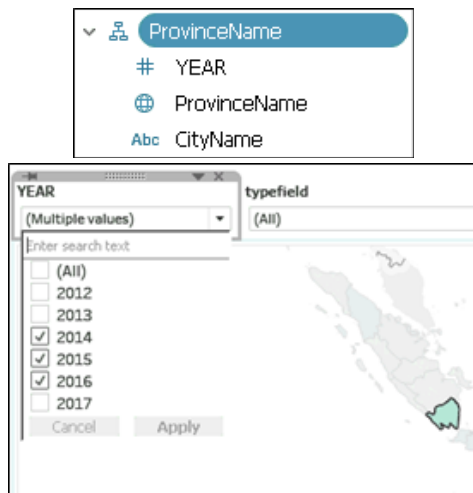


Figure 11. Hierarchy in Tableau 10  
(Source: Authors)

### 3.3. Review of Results

Prior to DW and BI, it took 2 to 3 weeks to do data processing by the Marketing Intelligence staff in producing reports that were used to monitor marketing activities and sales targets. With the presence of DW and BI, monitoring could be done quickly, especially in presenting data. In this study, with the availability of automation using ETL, DW, and BI, it obtained a much faster data processing process that only took 77 min, as shown in Table 4 to produce a similar BI report and the recapitulation process could also be done automatically.

To complete this study and find the success or failure, a questionnaire was distributed to BI users, as explained previously. The results of the questionnaire were collected from 3 BI users as shown in Table 9 below:

Table 9. Results of the Questionnaire  
(Source: Authors)

No	Questions	Average Score
1	Does DW and BI help marketing teams in processing data?	5
2	Is the presence of DW and BI in the process of presenting data able to be accommodated faster?	4.67
3	Is the data generated accurate and can be used primarily for Marketing Intelligence?	5
4	Can the BI dashboard be used as a decision maker tool?	4
4.a	If you agree, what decisions can be made with the BI dashboard that have been accommodated at this time?	

No	Questions	Average Score
5	Can BI be easily accessed by the BINUS Marketing team?	4.33
6	Is the link to access BI easy to remember?	5
7	Are the DW and BI implementations as expected by the Marketing team?	4
<b>Average Final Score</b>		<b>4.57</b>

From the results of the questionnaire, this study has a final value of 4.57 and is declared successful because it refers to Table 1, where the success rate of this implementation must have a value between 4 and 5.

## IV. CONCLUSION

Tableau is a tool which is well known for its modern name, as BI does not have complete ETL tools. So, the process approach in combining PDI and DW as a data source certainly makes Tableau as a BI tool more useful in presenting data. The results of the dynamic dashboard can also be broken down quickly according to one's needs, even though there are hardware limitations because ideally a server for BI should use a physical server instead of a virtual server in order to have fast performance, but with the presence of DW combined with BI these limitations can be overcome because the pre-calculation value is done in DW and BI only needs to consume data from DW. ETL uses PDI, equipped with data warehouse and BI tools, which are very useful for companies, especially some universities like BINUS.

Managerial implication for this study is that with BI monitoring the operational of Marketing Intelligence become more effective and efficient, because we change tabular format into more graphical and picture that made easier to understand by managerial level. Moreover, decision making become faster and it will increase the competitive advantage against another competitor. DW, ETL and BI also reduce the time needed to produce certain complex report, previously it take 2-3 weeks for generate complex report, now only in 77 minutes the report could be seen and the data is more easy to understand. In the other hand with DW and BI it also reduces incorrect data process because incorrect data in tabular form is harder to be founded comparing when we present it in graphical form. For further research, comparing ETL process using PDI and Tableau since they will have complete ETL feature in 2019.

## REFERENCES

- Bassil, Y. (2012, December). A Data Warehouse Design for A Typical University Information System. *Journal of Computer Science & Research (JCSCR)*, 1(6), 12-17. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1212/1212.2071.pdf>
- Brandão, A., Pereira, E., Esteves, M., Portela, F., Santos, M. F., Abelha, A., & Machado, J. (2016). A Benchmarking Analysis of Open-Source Business Intelligence Tools in Healthcare Environments. *Information*, 7(57), 1-16. doi:10.3390/info7040057
- Casters, M., Bouman, R., & Dongen, J. v. (2010). *Pentaho® Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration*. Canada: Wiley Publishing, Inc.
- Eckerson, W. W. (2006). *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Golfarelli, M. (2010). From User Requirements to Conceptual Design in Warehouse Design: A Survey. In L. Bellatreche, *Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction (Advances in Data Warehousing and Mining (ADWM))* (pp. 1-16). USA: Information Science Reference.
- Golfarelli, M., & Rizzi, S. (2009). *Data Warehouse Design: Modern Principles and Methodologies*. Italy: McGraw Hill.
- Kakish, K., & Kraft, T. A. (2012). ETL Evolution for Real-Time Data Warehousing. 2012 Proceedings of the Conference on Information Systems Applied Research (pp. 1-12). New Orleans Louisiana, USA: EDSIG (Education Special Interest Group of the AITP) .
- Kimball, R., & Caserta, J. (2004). *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. USA: Wiley Publishing, Inc.
- Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling* (2nd ed.). Canada: John Wiley & Sons, Inc.
- Kimball, R., & Ross, M. (2010). *The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence* (1st ed.). Indianapolis: Wiley Publishing, Inc.
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd ed.). Indianapolis: John Wiley & Sons, Inc.
- Kimball, R., Reeves, L., Thornthwaite, W., Ross, M., & Thornwaite, W. (1998). *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses*. USA: John Wiley & Sons, Inc.
- Muntean, M., Bologa, A.-R., Bologa, R., & Alexandra, F. (2016). Recent Researches in Educational Technologies. In N. Gavriluta, R. Raducanu, M. Iliescu, H. Costin, N. Mastorakis, V. Olej, & J. Strouhal (Ed.), *7th WSEAS / IASME International Conference on Educational Technologies (EDUTE '11)* (pp. 118-123). Romania: WSEAS Press. Retrieved from <http://www.wseas.org/multimedia/books/2011/Iasi/EDUTE.pdf>
- Rudy, & Miranda, E. (2015). Management Report for Marketing in Higher Education Based On Data Warehouse and Data Mining. *International Journal of Multimedia and Ubiquitous Engineering*, 10(4), 291-302. Retrieved from [https://www.researchgate.net/publication/283126375\\_Management\\_Report\\_for\\_Marketing\\_in\\_Higher\\_Education\\_Based\\_On\\_Data\\_Warehouse\\_and\\_Data\\_Mining](https://www.researchgate.net/publication/283126375_Management_Report_for_Marketing_in_Higher_Education_Based_On_Data_Warehouse_and_Data_Mining)
- Salaki, R. J., Tangkawarow, I., & Waworuntu, J. (2015). Extract transformation loading from OLTP to OLAP data using pentaho data integration. *International Conference on Innovation in Engineering and Vocational Education* (pp. 1-8). Bandung, Indonesia: IOP Publishing. Retrieved from [https://www.researchgate.net/publication/303515839\\_Extract\\_transformation\\_loading\\_from\\_OLTP\\_to\\_OLAP\\_data\\_using\\_pentaho\\_data\\_integration](https://www.researchgate.net/publication/303515839_Extract_transformation_loading_from_OLTP_to_OLAP_data_using_pentaho_data_integration)
- Sen, A., & Sinha, A. P. (2005, March). A Comparison of Data Warehousing Methodologies: Using a common set of attributes to determine which methodology to use in a particular data warehousing project. *Communications of the ACM*, 48(3), 79-84.
- Silva, F. D. (2005). Data Warehousing & Business Intelligence ROI. *International Journal of the Computer, the Internet and Management*, 13(SP2), 1-3. Retrieved from <http://www.ijcim.th.org/SpecialEditions/v13nSP2/pdf/p1.1-1.3-Data%20Warehousing.pdf>
- Tarnaveanu, D. (2012). Pentaho Business Analytics: a Business Intelligence Open Source Alternative. *Database Systems Journal*, III(3), 23-34. Retrieved from [http://www.dbjournal.ro/archive/9/9\\_3.pdf](http://www.dbjournal.ro/archive/9/9_3.pdf)
- Waas, F., Wrembel, R., Freudenreich, T., Thiele, M., Koncilia, C., & Furtado, P. (2013, April). On-Demand ELT Architecture for Right-Time BI: Extending the Vision. *International Journal of Data Warehousing and Mining (IJDWM)*, 9(2), 21-38. Retrieved from <https://pdfs.semanticscholar.org/5fc5/f7c59f881d536da27626f309927926626fa8.pdf>