Business Intelligence for Education Management System

Abba Suganda Girsang, David Alfa Sunarna, Ahmad Syaikhoni, Aris Ariyadi
Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University
Jakarta, Indonesia
Email : agirsang@binus.edu, david.sunarna@binus.ac.id, ahmad.khoni@binus.ac.id, aris.ariyadi@binus.ac.id

***Abstract*— The main function of school operations is scoring and reporting. Each of them is unique in every school and cannot be generalized although using the same curriculum. School personnel such as headmaster, teacher, and staff also not only assigned to one school location and can have multiple roles in each school location. With that complexity, a single query is not fast enough to generate a scoring report. This paper proposed the system business intelligence which can analyze the progress of student learning. This paper used nine steps to design a data warehouse as a tool business intelligence. The On-line Analytical Processing (OLAP) is used to analyze the data and visualize them using dashboard. By using data warehouse, each school can answer business question relation to scoring and visualize them in a dashboard to make a decision.**

Keywords*— *Socrates System School, Business Intelligence, Report, Scoring, Data Warehouse

I. INTRODUCTION

Sokrates system is a school information system that aims at bringing a total quality of education management by ensuring an efficient way of day-to-day school operations. Sokrates system establishes under Bina Nusantara in Educational Technology. When it first started, Sokrates system is a web application implemented in one school with four-year level (ecy, elementary, middle school, and high school). Now Sokrates system is implemented in five schools (est. 2000 student). The next upcoming client is 28 schools under one foundation company. The main function of school operations is scoring and reporting. Each of them is unique in every school and cannot be generalized although using the same curriculum. Also, there is foundation companies that usually have a multiple school location with different management and curriculum. School personnel such as headmaster, teacher, and staff also not only assigned to one school location and can have multiple roles in each school location. With that complexity, a single query is not fast enough to generate a scoring report. To help teachers and school staff in each client, Sokrates System creates a Data warehouse to analyze the progress of student learning: this article focuses on analyzing the learning progress of the student.

There are two methods for creating the data warehouse, Inmon, and Kimball or more commonly refers to a top-down technique and down-top technique, but on this case, we used Kimball method which had nine steps to design a data warehouse

Extract, transform, Load (ETL) is a process where data from the operational database are extracted from a different type of database load it to one consolidate data warehouse. After that data have to be transformed like cleaning, join, filtering and validating. OLAP is the technique to analyze business data. It can make complex calculate, analyzes trend and data modelling and OLAP Cube can analyze business data with multiple dimension, it contains multiple table and base on those JPivot can make interactive visualization.

In this research, for data warehouse, we used On-line Analytical Processing (OLAP) to analyze the data and visualize student progress using OLAP cube and JPivot. Then, for design and implementation of data warehouse, we used Pentaho Schema Workbench, Pentaho Report Designer, Pentaho Business Intelligence. Pentaho is one of open source for Data warehouse and Business Intelligence tools, not only because it open source had all component that must have by Business intelligence solution some of them are able to extract, transform and load (ETL) from all type database, could analyzes data with on-line analytical processing (OLAP), data mining (DM) and any kind BI techniques. It also makes a report and visualize the result using Pentaho [1] [2] [3].

II. RELATED WORK

Pentaho had been used by many researchers as tools to implement data warehouse on a various field. For example, Pentaho is being used to create a data warehouse for a store and analyze benchmarking e-Government services [4]. The data warehouse was created by extracting data from information system on three public hospitals, and transforms it as it needs and load to the data warehouse. That data warehouse uses to monitoring antimicrobial resistance, measure antibiotic usage, detecting antimicrobial prescribing error and hospital that acquired bloodstream infection and measure the cost of the infection. Implementation Data warehouse for hospital infection control make decision maker increase precision in making a decision and they saved time and money.

Data warehouse prototype is also being used for integrating and visualize data from radiology [5]. Integrate data from various electronic source use to analyze and make a decision. Creating data warehouse from a various data source and extract into one consolidate data transform and load into the data warehouse, then using OLAP to visualize, calculate and analysis trend. The result is prototype can display visualize data from a various electronic source and analyze the trend.

At Bioinformatics, the data warehouse can be used to traditional Chinese medicine (TCM) clinical to use as medical knowledge discovery and decision support, Zhou et al [6] developed a data warehouse that integrates and analyze from the structure electronic medical and data for medical knowledge discovery and traditional Chinese medicine support. A data warehouse integrates data from

978-1-7281-5423-7/19/$31.00 ©2019 IEEE

clinical reference information model and TCM clinical data model and doing Extract, transform and load to the data warehouse and using OLAP to analyze. A Clinical data warehouse could be used to trend diagnosis for a syndrome.

There is also research which implements the data warehouse for integrating various database and services that related to small molecules bound to macromolecules, a data warehouse name Ligand Depot [7] is created. Ligand Depot is data warehouse that gives analysis and data visualization from many databases, services, and tools that related to small molecules.

Data mining and Data warehouse are also being used to analyze various aspect from healthcare company [8] [9] such as clinical data, administration data, and technical data, by integrated those data. and using OLAP to analyze those data to by business problem formula could help make a decision like operational efficiency, cost, and revenue.

III. PROPOSED METHOD

The method that used in this paper is nine steps from Kimball and Ross. The data from some division has been collected to do this method. The result information of data warehouse is presented by using Pentaho Business Analytics. The following are the nine steps of creating a data warehouse:

A. Choosing The Process

The first stage in building a data warehouse is to observe the processes that occur within the organization, in this case, the school. Selecting the main processor means determining the main subject of organization. In general, the main process that occurs in educational institutions in the process of learning and teaching, here students are required to attend at each session. Finally, at the end of every period, there will be test to determine the score of students. So the main process subject refers to an activity that can answer all important questions and have certain a characteristic is the process of assessment and absenteeism. This Case study emphasizes the design process of data warehouses in the scoring aspect.

B. Choosing the Grain

Choosing a grain is a step to determine what data will be represented in a fact table. After determining the grain of the fact table, we can determine dimensional tables related to the facts table. The grain on the fact table also determines the grain dimension table. In this case, the grain can be obtained from the score data.

C. Identifying and confirming the dimensions

Identifying and conforming dimension tables with fact tables starts with understanding that chosen dimension represent the data required in the data warehouse. Dimensions are an important set of perspectives to illustrate the facts contained in the fact table. In this case each row in the fact table containing dimension information such as the student dimension, time, class, teacher, and subject.

D. Choosing the fact

The grain of a fact table determines the facts that can be used. At this stage, we specify required measurement in the fact table. What kind of information that we wanted and what level of grain should it cover? in the end that the data can be analyzed and presented a report. Table I shows a fact table design created based on the available information.

TABLE I. SCORING FACT TABLE.

Column Name	Data Type
academic_calendar_id	integer
class_id	integer
student_id	integer
subject_id	integer
teacher_id	integer
student_subject_score	float

E. Storing pre-calculations in the fact table

After the fact table created, the next step is to analyze the data which one can be calculated. The Calculations that can be done include a comparison of data values with certain parameters with standard quality values. After that, the calculation result of an attribute can be stored in the database. It aims to reduce the risk of errors every time the program operates calculations on these attributes.

F. Rounding out the dimension table

From the dimensions that we have been identified, the next step is to determine and write a description of the attributes in the dimension table. The goal is to make sure our dimension table will be easily understood by end-user. The dimensions that we will be used are dim_academic_calendar, dim_attendance, dim_class, dim_student, dim_subject, dim_teacher, and dim_score.

G. Choosing the duration of the database

At this stage, we determine the data in what duration of data in a transactional database used for the source of the data warehouse. The duration of the data to be used in the data warehouse is data from 2016-2017 and 2017-2018 academic years. The database engine we used is PostgreSQL.

H. Tracking slowly changing dimensions

Dimensions of data that change slowly can cause many problems. There are three basic types of solutions to solve this problem such as: rewriting attributes of changed data, creating new records on dimensions, and creating alternative attributes to accommodate new values. In this project, we implemented the second type of slowly changing dimension. If there are data changes, the new data record will be made so that old data can still be tracked.

I. Deciding the query priorities and the query modes

In this final stage, we do not just design physical data warehouse but also identify the problems that may exist at the stage of physical design itself. Many things in deciding the query priorities and modes need more attention because it will involve the physical design of data warehouse that will affect to the end-user. At this stage issues that may occur such as indexing data, indexed view, storage, backup, and security should be overcome immediately.

IV. ANALYSIS RESULTS

A. Extract, Transform, Load

To create an ETL process in scoring, a star schema is created. Star schema is used to generate score fact through five dimensions of the scoring variable. Star schema in this case can be shown on Figure 1

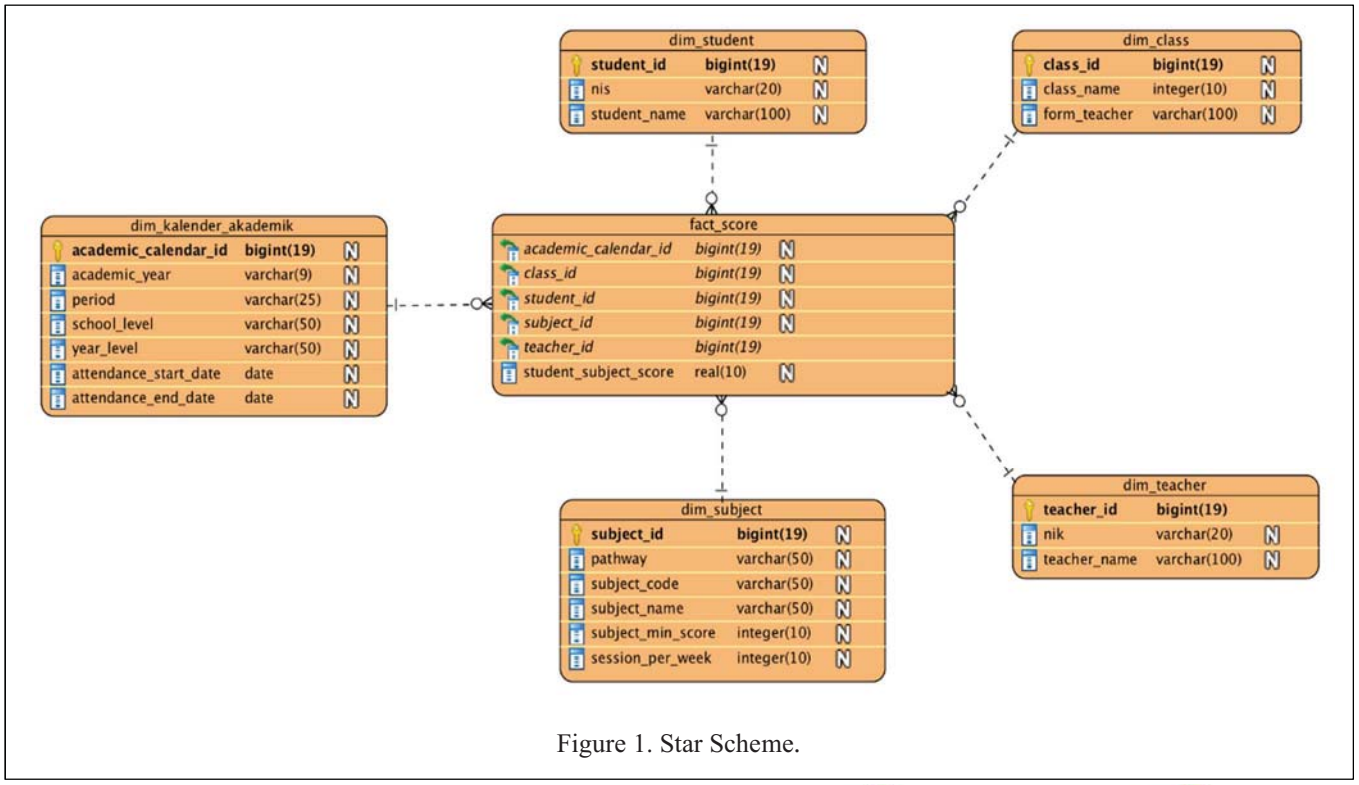

Figure 1. Star Scheme.

Fact score is generated with a granularity of every student. Every school is started within academic calendar in a certain period. In certain period, there must be school class. In every school class, there is a class member which is students itself and subject. Each teacher is responsible for their own assigned subject. For example, in the academic year 2017-2018 period II, student Ariz Rahma Azizah has belonged to class 7A. The Mathematics subject score is 70, and the subject teacher is Agung Haryanto.

An ETL process is generated through Pentaho Data Integration. Pentaho Data Integration is an open source software to generate ETL process with certain features. Please notice this research is not using a real data. Dummies data were used with original structure table.

Figure 2 shows that the dimension is extracted and transformed through a query process. We're using a query instead of data integration process because the data are very complex and difficult to get if not using query.

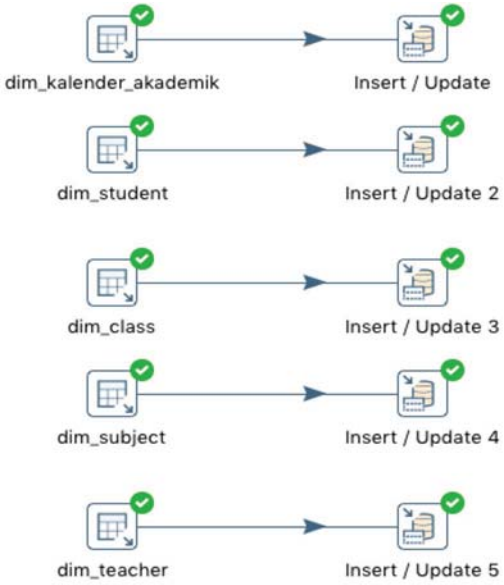

Figure 2. ETL Process for Dimension Table.

A fact score table is generated by joining student score component, academic calendar, class, and teacher subject. All ETL result is stored in the dwh_sokrates database with same structure table as the star schema as shown Figure 3. ETL process is taking 70% time of the data warehouse creation process. One of the purposes of the data warehouse is to access specific data instantly. By storing score data to the fact table, the data will be easily displayed and configured in the next process.

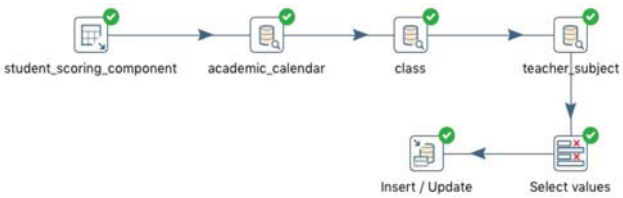

Figure 3. ETL Process for Fact Score Table

B. OLAP Cube

The next step is to create an OLAP cube to make the score fact multidimensional. In this research, aggregation for the score is average. But, it's not impossible to add another aggregation such as minimum and maximum score. OLAP cube is generated through Pentaho schema workbench.

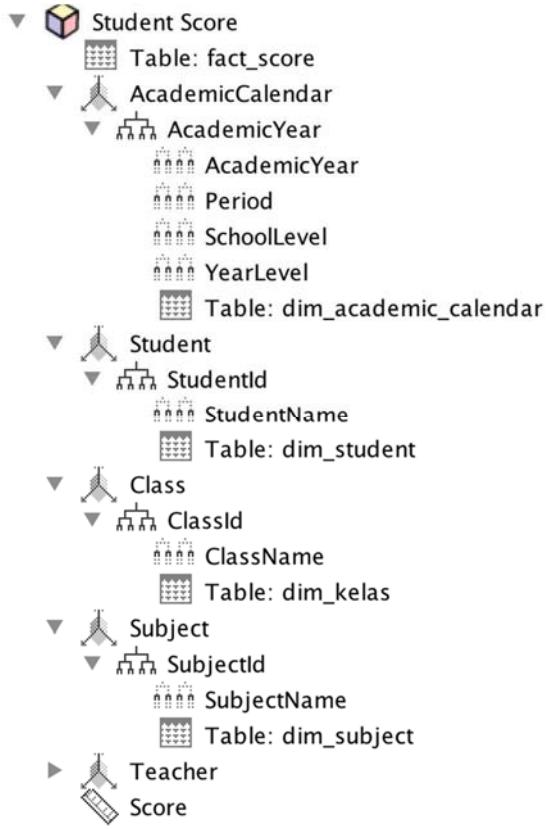

Figure 4. OLAP Cube.

An OLAP Cube is created just like the star schema as shown Figure 4. A measurement is a Score (fact score) with average aggregation. Measurement is obtained from the fact_score table. For example, Academic Calendar dimension contains Academic Year hierarchy. Academic Year hierarchy has four levels and obtained from table dim_academic_calendar.

C. Reporting

From the ETL table dwh_sokrates, many customized report about scoring can be created. The report is generated by Pentaho Report Designer as shown Figure 6 and Figure 7.

Name	NIS	Year	Period	Level	Class	Subject	Teacher	Score
Adele Caddy	1770006740	2017-2018	Term 1	Elementary Years	EL 5	Mandarin	Quintin Larkings	69

Figure 5. Pentaho report designer output in a row.

The Figure 5 is very basic report, all the student subject score with detail academic year, period, class, and teacher can be easily created in the data warehouse.

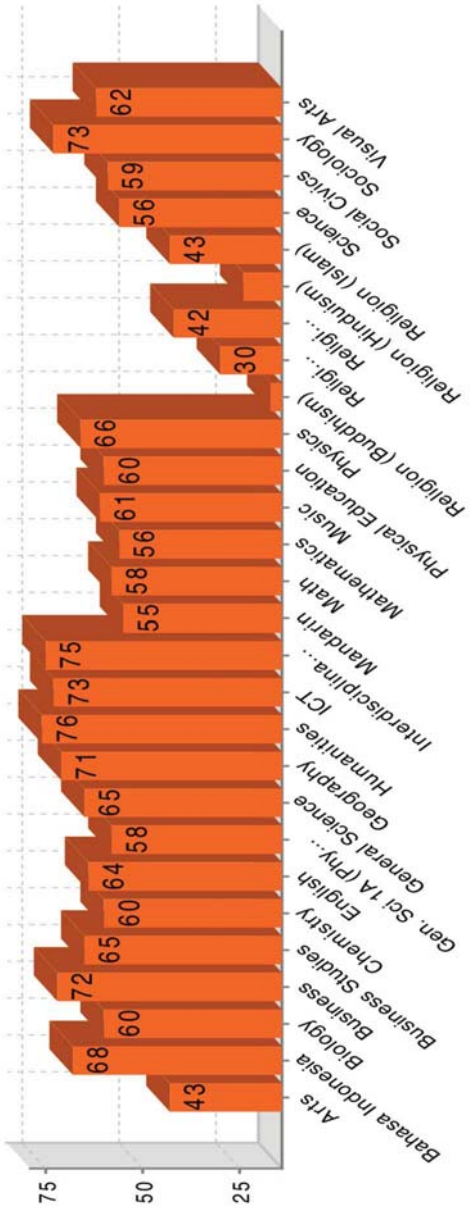

Figure 6. Pentaho report designer output showing average student score.

The picture above shows a customized report for showing average score in each subject. Showing data like this only takes a second because of all of the data needed already saved in table fact_score.

D. Dashboard

Data analysis from the Sokrates data warehouse can be done by using Pentaho Business Intelligence Server. With this software, we could create data analytics and dashboard. Data analytics and dashboard will answer the business question and track data. JPivot is used for creating a data analytics from the cube that already created before. The example dashboard can be seen on Figure 7.

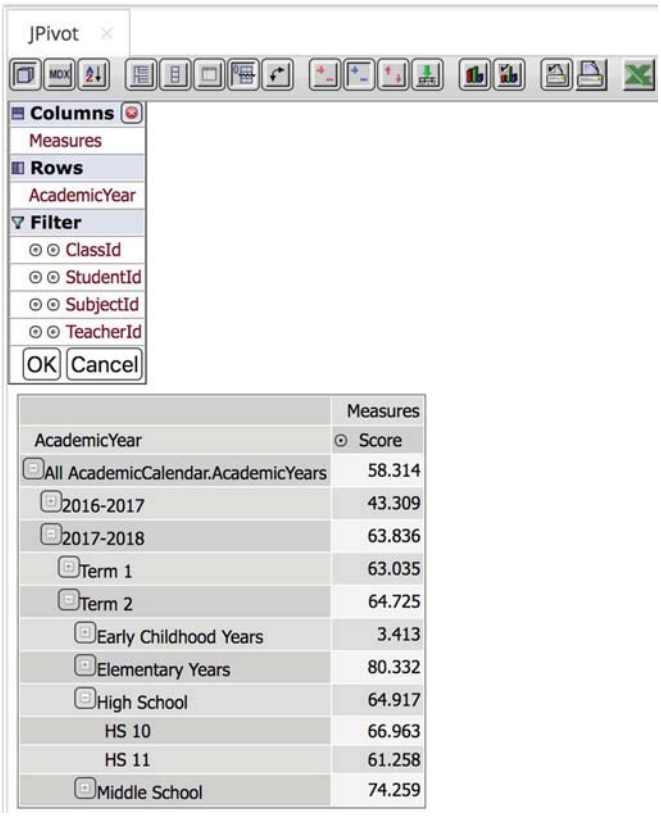

Figure 7. Jpivot average score by academic year.

By using JPivot in Pentaho Business Intelligence Server, data analysis can be done easily because the data is already described in a cube and we can define the granularity showed in the data instantly. The picture above shows detailed average score by the academic year and by the teacher subject. We can modify the data showed on the table based on the Columns (measurement), Rows, and Filter. Besides using JPivot, data analysis also can be done by using Saiku Analytics as shown Figure 8.

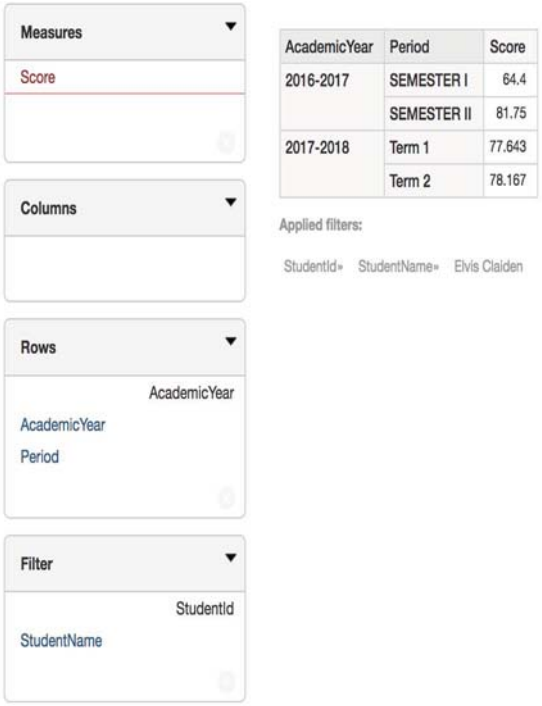

Figure 8. Saiku analytics.

The main concept of data analytics in Saiku remains the same as JPivot. Saiku added drag and drop features and more features to the data analytics interface. From the picture above, we analyze the average score of Elvis Claiden in each period. Measures set in Score, rows are set in AcademicYear and Period and using filter StudentName with Elvis Claiden name.

To visualize result in data analytics, a dashboard can be created by also using Pentaho Business Intelligence Server. A dashboard is a business intelligence tool used to display data visualizations in a way that is immediately understood even for a nontechnical user. Data is visualized on a

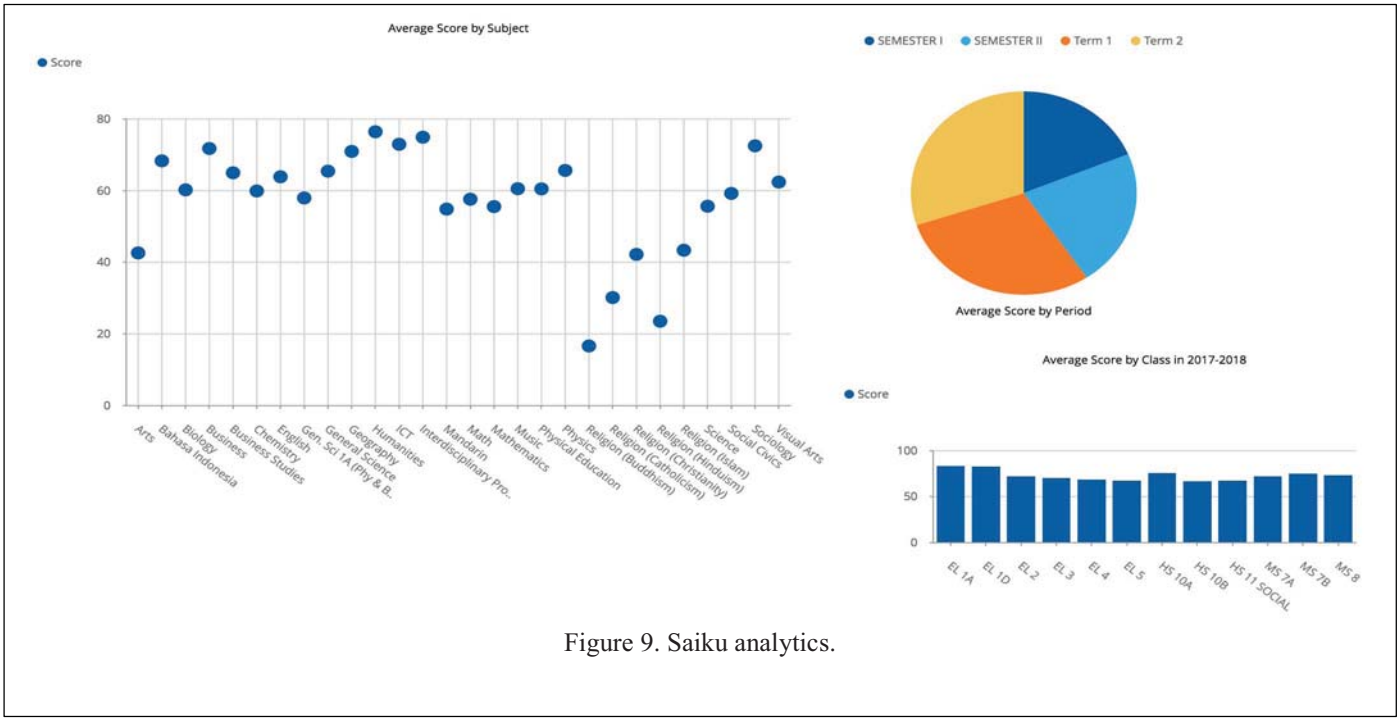

Figure 9. Saiku analytics.

dashboard as line charts, bar charts, and gauges so that school can track the score of their students against benchmarks and goals.

Figure 9 shows any information as a dashboard. First average score by subject is showed; second average score by period, an average score in class from the academic year 2017-2018.

V. CONCLUSION

By creating a data warehouse and implementing data mining for Sokrates School Information System, it really helps management to view summarized data based on a business question. It can also be done with querying data, but the data warehouse fast solution to access the data. Also querying data will be complex and time-consuming. The data warehouse in Pentaho come with a simple interface to do data analytics. A summarize of business question can be shown in the dashboard and provide data for the decision-making process. This research using Kimball approach (down-top) with a subject score as a start. In the future works, we can add more fact such as attendance, extracurricular score, behaviour score, and other process related to a decision-making process.

REFERENCES

[1] Tarnaveanu, D. 2012 Pentaho business analytics: A business intelligence open source alternative. ERP and E-Business Application Deployment in Open Source Distributed Cloud Systems

[2] He, Y., Lee, R., Huai, Y., Shao, Z., Jain, N., Zhang, X., & Xu, Z. 2011 RCFile: A fast and space-efficient data placement structure in MapReduce-based warehouse systems. In Data Engineering (ICDE), 27th International Conference on 1199-1208.

[3] Tuncer, O., & van den Berg, J 2010 Implementing BI concepts with Pentaho, an evaluation. Delft University of Technology. Netherlands.

[4] Liu, X., & Luo, X. 2010. A Data Warehouse Solution for E-government International Journal of Research and Reviews in Applied Sciences 4(1) 101-105.

[5] Prevedello, L. M., Andriole, K. P., Hanson, R., Kelly, P., & Khorasani, R 2010 Business intelligence tools for radiology: creating a prototype model using open-source tools. Journal of digital imaging, 23(2) 133-141.

[6] Zhou, X., Chen, S., Liu, B., Zhang, R., Wang, Y., Li, P., ... & Yan, X 2010 Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support Artificial Intelligence in medicine 48(2-3) 139-152.

[7] Feng, Z., Chen, L., Maddula, H., Akcan, O., Oughtred, R., Berman, H. M., & Westbrook, J. (2004). Ligand Depot: a data warehouse for ligands bound to macromolecules. Bioinformatics, 20(13) 2153-2155.

[8] Wisniewski, M. F., Kieszkowski, P., Zagorski, B. M., Trick, W. E., Sommers, M., & Weinstein, R. A. 2003. Development of a clinical data warehouse for hospital infection control. Journal of the American Medical Informatics Association 10(5) 454-462.

[9] Silver, M., Sakata, T., Su, H. C., Herman, C., Dolins, S. B., & O Shea, M. J 2001 Case study: how to apply data mining techniques in a healthcare data warehouse. Journal of healthcare information management 15(2) 155-164.

[10] Kimball, R., & Ross, M 2011. The data warehouse toolkit: the complete guide to dimensional modeling. John Wiley & Sons.

[11] Abba Suganda Girsang, and Candra Wijaya Prakoso 2017 Data warehouse development for customer WIFI access service at a telecommunication company International Journal on Communications Antenna and Propagation 7 (2) 114-124