The Use of Business Intelligence Tools for Health Data Analytics

Phatpicha Yochum[1] and Chutima Beokhaimook[2]

College of Information and Communication Technology, Rangsit University,
Pathumthani 12000, THAILAND

Abstract— **This research develops business intelligence (BI) for university health data analytics. The data warehouse has been constructed to store the four-year annual health checkup data of students and staffs. The reports are appropriately designed to present the relationships among data and forecast or predict the risks of disease. The results of the analytics from OHW-RSU BI system show that 95% of students have good health. However, the health problem is still found mostly in myopia. For staff health analytics, 90% of staffs have good health. Unfortunately, unhealthy staffs are in class 2 and 3 obesity.**

Index Terms— *Business Intelligence, Intelligence System, Data Analytics, Data warehouse, Healthcare Analytics*

I. Introduction

Business intelligence (BI) is modern techniques for data summarization in multidimensional model. BI technologies support decision making, forecasting and strategic planning process in the organization. By extracting knowledge from overall information immensely, BI is able to create various kinds of reports, for instance, dashboard, report, table, graph, map, etc. User can retrieve and use those reports in many ways such as comparing the related information or points of different views in the analytics, displaying relationship among data, predicting the trends of interesting values.

In a healthcare of today, the health problems may cause directly by genetic diseases or from behavioral health risk. In order to effectively manage healthcare, it must be planned to deal with the problems that may occur. Office of Health Welfare (OHW), Rangsit University (RSU) has annually collected health checkup data of students and staffs for many years. However, such data set is difficult to analyze and not beneficial. The long time stored health data have not been manipulated to present the overall health problems of the university as the executive board's needs. Therefore, one solution for this problem is the development of BI framework for health data analytics. BI can support decision making related to healthcare problems. BI is an assist tools not only for the university executive boards to access and understand the healthcare information, but also for the staffs to explore each data item in detail easily and effectively.

To construct right BI system, many aspects of knowledge are adopted to analyze the data set, for instance, health informatics, descriptive analytics [1] (to explain what is going on or what might happen), diagnostics analytics [2] (to deeply diagnose to the cause of the resulting effects) and predictive analytics [3] (to analyze the current and past trends to predict the risk of disease).

A data set of this research is 4-year annual health checkup data of students and staffs of RSU since year 2012. This dataset is stored as an information system of the university. In this study, BI is applied for the three objectives: 1) to analyze this dataset for exploring the health situation in RSU, 2) to design appropriate reports that can present the relationships among data, forecast or predict the risks of disease and 3) to develop the practical healthcare business intelligence system that can monitor and retrieve required information.

This research has been conducted under three consideration points. First, which are the appropriate tools and data warehouse techniques to be chosen to this research? Second, how to clean and integrate data from disparate sources into a data warehouse that allows multiple data visibility for each dimension of data? The last point is what is the right technology of the business intelligence for developing the RSU healthcare data analytics system?

This paper is organized as follows. Section 2 introduces the business intelligence and data warehouse concepts. Section 3 provides the users' requirement information. In Section 4, the data warehouse construction is presented. The results of OHW-RSU BI system are shown in Section 5. The last Section provides the conclusions.

II. Business Intelligence and Data Warehouse

BI technologies have been developed for years. The term "Business Intelligence" or "BI" and its concept were introduced by Professor Richard Miller Devens [4]. The BI helps users to access data easily and effectively transform such data into information and knowledge. BI can analyze data characteristics or data behaviors. The outcomes of BI

The authors are with the College of Information and Communication Technology, Rangsit University, Pathumthani 12000, THAILAND

[1]mink.phatpicha@gmail.com
[2]chutima@rsu.ac.th

become the precise information for business decision making which can enhance business competitiveness. Since BI is the self-service approach which enables business users to access and work with corporate data without requiring users' backgrounds in statistical analysis, business intelligence or data mining. End users can use BI and readily customize graphic user interface as several view points, relationships among data or predicted trends that are very useful for strategic planning. In the past decades, many researches employed BI widely. There are several works related in education business [5-6]. In [5] BI was proposed as the tools for university administration to analyze students' characteristics and figure out the numbers of students in each faculty and course. BI showed students' favorite subjects and provided tools for students' grades checking. While in [6] BI model was presented for analyzing the results obtained from the online tutoring processes. This model helped matching online teaching styles to each course. The students' successes were also predicted by the proposed model. In medical field [7], BI-based epidemiological alert system was constructed to monitor and display the trend of the spreading infectious diseases out in target area [7]. There was the work that adopted BI for analyzing revenue budget and presenting in data multi-dimensional reports that could show the data comparisons and trends for budget planning and allocating [8]. The voluntary automobile insurance applies BI to develop the decision support system which can enhance the business competitiveness [9]. BI was also applied in sugar cane production management to show multiple data perspectives and provide necessary information to manage enough sugar cane for domestic consumption and export of Thailand [10]. We also found the work of BI to perform sales analysis [11].

Data Warehouse is the central repositories of integrated data from one or more disparate sources. Data warehouse stores current and historical data in one single place that used for creating analytical reports for knowledge workers throughout the enterprise. In data warehouse system, three necessary processes to convert data from those data sources into data warehouse are extraction, transformation, and loading, so called ETL. The data extraction retrieves data from multiple data sources which may be internal or external data sources. Then, the data acquired from the extraction process will be transformed and formatted to the desired format. In loading process, the transformed data will be moved to data warehouse. In this decade, there are many products that provide ETL tools. Some popular ETL products are Pentaho Data Integration (Kettle) [12], Jaspersoft ETL [13], CloverETL [14], PowerCenter [15] or Talend Open Studio [16]. Data warehouse has been employed in many researches widely. In medical area [17][18][20], Kaul, el al [17] presented healthcare prediction systems using big data. The data sets in the data warehouse are used to perform clustering the patients as groups according to the similarity of the patient's symptoms. With the disease information and physical activities of the focused patients, the system can predict the future patient's health characteristics. In [18], the data warehouse was developed to store the collected diagnostic and illness information from the primary care units. It was applied for the policy planning decision support system. With Hadoop and

Hive [19], the patient treatment information could be obtained from the real time analytics on huge electronic health records (EHRs) [20]. Several works regarding to data warehouse also found in education business such as the student data warehouse that is developed for the decision support system of Chiangmai University [21] or data warehouse of Lampang primary educational service area [22] that collect the primary schools information to be used for policy planning. For other areas, data warehouse was found in managing the traffic which showed an accident statistics reports for police department [23]. Moreover, in factory's production unit, data warehouse that applied with BI could solve rework problems in wheel alignment process and figured out the exact causes of the problems and suggested effective solutions [24].

Designing and building the data warehouses for BI are very important since data warehouses are a repository of information that holds everything important to the organization. The more volume of data, the more advantages we can take from it. Meanwhile the BI acts as a reporter that presents various kinds of visualizations. The BI also has a self-service system. BI's data analytics tools are quite simple for all user levels to investigate and reveal patterns of the insight data. BI allows user to deter-mine the metrics and differentiate the data to the different data viewpoints or business strategies. Moreover, BI can supports various ways of using data, such as file uploading and database connecting to the business applications. The integration of data warehouses and BI could be confirmed to improve business decision making [25], forecasting [26] and strategic planning process [27] in the organization. The Gartner Incorporated which is the BI research leader company has compiled and compared the version of completeness, ability of data analysis and presentation reports among BI tools in this world [28]. Fig. 1 shows Magic Quadrant for Business Intelligence and Analytics Platforms 2016 by the Gartner Incorporated [29]. A chart shows leaders in BI area, which are Tableau [30], Qlik [31] and Microsoft [32].
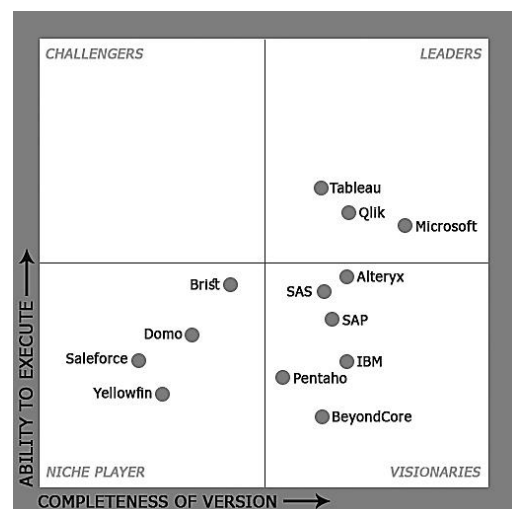

Fig.1 Magic Quadrants for Business Intelligence and Analytics Platforms 2016 [29]

The products from Microsoft Corporation are interesting. Since the Microsoft Corporation becomes leading company and has several tools in the area of BI such as Microsoft Office (Excel), Microsoft SQL Server, Microsoft SharePoint and Microsoft Power BI. Microsoft BI

tools are quite popular [33]. In addition, most of software used in RSU is Microsoft products. Therefore, in this research, Microsoft Power BI is applied to develop the healthcare BI system since it is the latest BI product from Microsoft Corporation and its data analytics tools is tool suite.

III. Requirement Analysis

User requirements have been collected from staffs and executives in the OHW. The users of this BI system are divided into two levels: executive level and operational level. For users who are in executive level will be shown as dashboards that comprises of many consolidation views of related data. While the users in operational level will see not only summarized reports, but also the deeper details of data. According to the health data collection of the OHW is separately gathered the health data of students and staff with different health checkup techniques. Therefore the two datasets which are student-health set and staff-health set have been used for develop the BI system, later named the OHW-RSU BI. The visualizations are designed to present in various report forms, for instance, frequency, relationships, comparisons and trends. In our BI system, student and staff health information reports are shown in three styles. They are the general information, physical examination, and behavioral risk as illustrated in Table 1 and Table 2.

Table 1. Three students health reports styles and their detail topics

General Information	Physical Examination	Behavioral Risk
Department	General physical	Cigarette
Sex	examination	Alcohol
ABO blood group	BMI	Exercise
Rh blood group	Blood pressure	
	Visual acuity	
	X-ray	
	Urinalysis	
	CBC	

Table 2. Three staffs health reports styles and their detail topics

General Information	Physical Examination	Behavioral Risk
Department	General physical	Cigarette
Sex	examination	Alcohol
Age group	BMI	Exercise
	Blood pressure	Drug Allergy
	Visual acuity	
	X-ray	
	EGK	
	FBS	
	Cholesterol	
	Triglyceride	
	Uric Acid	
	SGPT, SGOT	
	BUN, Creatinine	

IV. Construction OHW-RSU Data Warehouse

The OHW performs students and staffs health checkup annually. Each year the health data have been stored in scatter Excel files since the data was received from different part of body health-checks. All files need to be gathered to the data warehouse. The processes of constructing the warehouse are shown in Fig. 2.
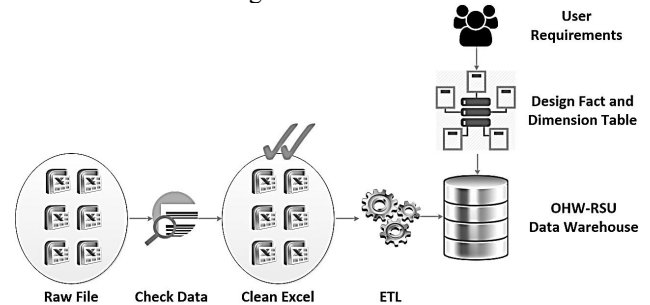

Fig.2 Workflow of constructing the data warehouse

The steps to creating the data warehouse, named OHW-RSU Data Warehouse, are iteratively done to ensure that the data are correct and right for our BI system. At first, all data in Excel files have been explored. In the meantime, the fact tables and dimension tables of the data warehouse have been designed regarding to the system requirements and characteristics of data in Excel file such as data value and data type of the attributes. All values of data and the attributes stored in Excel files have been checked their completeness. If the missing is detected, several ways to reduce those missing values are taken into account such as filling those missing values with 0.00 (if data type is real) or "0000000" (if data type is string of characters) or consulting with the data owner to suggest an appropriate value to substitute that missing. The cleaned data in Excel file then enters to the ETL process to be converted to the fact and dimension tables. At this step, those converted data in fact table and dimension table has been verified. If any defects are found, we are back to the step of cleaning data and modifying the fact and dimension table. In this research, OHW-RSU Data Warehouse is developed by preparing all structures on MS SQL Server. Pentaho Data Integration software is chosen to perform ETL process that will extract and load all cleaned data in Excel files to the prepared structures in data warehouse. OHW-RSU Data Warehouse consists of two fact tables with eleven dimension tables as shown in Fig. 3.
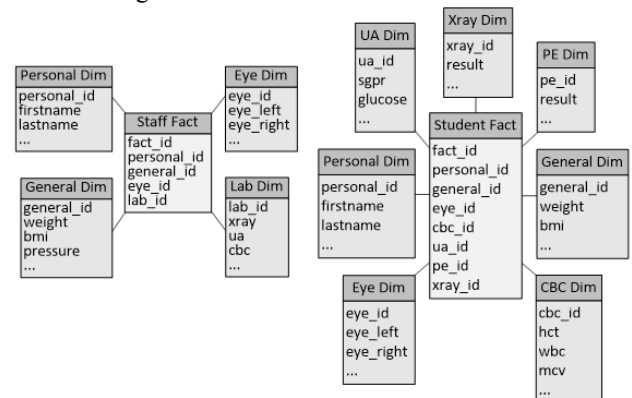

Fig.3 The two fact tables in OHW-RSU Data Warehouse

V. The OHW-RSU BI System

According to the requirement analysis in section 3, four report styles are selected to be shown in OHW-RSU BI dashboard. They are number or frequency style, data relationship style, data comparative or ratio style and data trend style. Moreover, The OHW-RSU BI system has slicer tool which allows users to select desired year or faculty. There are various visualizations for presenting the BI report, for instance, bar chart, column chart, combo chart, line chart, doughnut chart, pie chart, pyramid chart and multi-row chart as shown in Fig. 4.
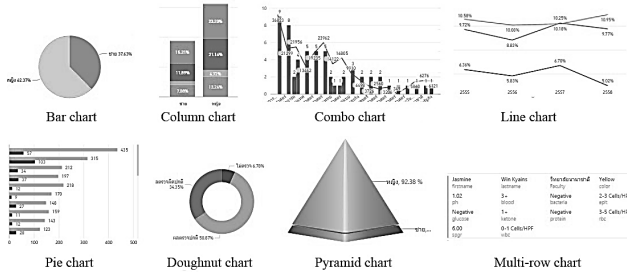

Fig.4 Various charts for BI report

The example of visualizations in dashboards of the OHW-RSU BI system is shown in Fig. 5. Every dashboard is comprised of three sections:- 1) slicer section which allows user to filter the information regarding to their selected year or faculty, 2) data amount section which presents the amount of records according to their selected year or faculty and 3) display section which demonstrates the results as the combination of charts.
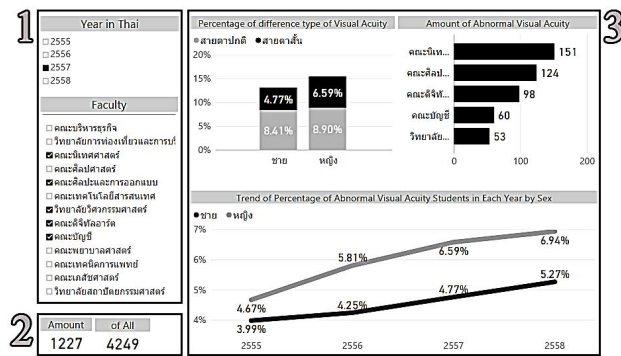

Fig.5 An example of OHW-RSU BI Dashboard and its sections
(Slicer, Data amount and Display)

From the OHW-RSU BI system, the results of the analytics shows the average amount of students are 4,000 per year. The average percentage of male is 38% and while those of female is 62%. The summarized statistics of health examinations obtained from student dataset are illustrated in Table 3.

The average amount of staffs is 1,500 per year. The average percentage of male is 45% and while those of female is 55%. The highest number of staffs falls in the staff with age range between 41-50 years age. The summarized statistics of health examinations obtained from staff dataset are illustrated in Table 4.

Table 3. The summarized statistics of students health examinations

The Examination of	Percentage of		Statistical Interpretation and details
	Normal	Abnormal	
Health	95	5	Most of students have good health.
BMI	75	25	Most of abnormal BMI students are in class 2 obesity.
Visual acuity	70	30	The abnormal visual acuity of students is about 20/20-20/25 myopia. The trend of abnormal visual acuity seemed higher myopia every year.
X-ray	99	1	Almost all students are normal.
Urinalysis	83	17	The abnormal urinalysis found more in female students than male students. The abnormal is 14% in female while those in male is 3%.
CBC	80	20	The abnormal CBC found more in female students than male students. The abnormal is 15% in female while those in male is 5%.

From the data of students and staffs behavioral risks, the analysis results discover the smoking behavioral risks that the average percentage of students who regularly smoke is 8% (male 7% and female 1%), while the average percentage of staffs who regularly smoke is 7% (male 6% and female 1%). Most of the staffs that regularly smoke are found in age range between 41-45 years old. In case of drinking alcohol behavioral risks, the average percentage of students who regularly drinking alcohol is 20% (male 15% and female 5%) where the trends of students that drinking alcohol seemed higher every year. About staff data, we found that the average percentage of staffs that regularly drinking alcohol is 10% (male 9% and female 1%). Most of the staffs that regularly drink alcohol are in age range between 31-40 years old. Accord-ing to the exercise behavior, the average percentage of students that regularly exercise is only 13% (male 9% and female 4%). While the average percentage of staffs that regularly exercise is only 12% (male 8% and female 4%). Most of the staffs that regularly exercise are found in the age range between 41-50 years old. The trend of men declined to exercise every year.

Based on our findings, OHW-RSU BI system could be used to track and present health information of staffs and students quite well. It also supports decision-making for the university's health campaigning and promoting. The OHW-RSU BI user satis-faction was rated after trying out. The evaluation had been focused in four aspects that are 1) the system meets all functional requirements 2) the validation of the system 3) the system usability and 4) the system security. The evaluation showed high user satisfaction level.

Table 4. The summarized statistics of staffs health examinations

The Examination of	Percentage of		Statistical Interpretation and details
	Normal	Abnormal	
Health	90	10	Most of staffs have good health. However, the staffs in age range between 41-45 years old found unhealthy.
BMI	40	60	Most of abnormal BMI staffs are in class 3 and 2 obesity.
Blood pressure	90	10	Most of staffs have normal blood pressure. All of 10% abnormal blood pressure is hypertension. The trend of blood pressure seemed higher every year.
Visual acuity	50	50	The abnormal visual acuity of staffs is presbyopia. The trend of abnormal visual acuity seemed higher presbyopia every year.
X-ray	97	3	Most of staffs are normal.
EGK	70	30	The abnormal EGK of male and female have the same ratio.
FBS	88	12	The abnormal FBS of male is 7% and female is 5%. However, the trend of female of FBS seemed higher every year.
Cholesterol	65	35	The abnormal cholesterol of male is 16% and female is 19%. However, the trend of cholesterol seemed higher every year.
Triglyceride	70	30	The abnormal triglyceride of male is 19% and female is 11%. However, the trend of triglyceride seemed higher every year.
Uric Acid	91	9	The abnormal triglyceride of male is 8% and female is 1%.
SGPT, SGOT	94	6	Almost all abnormal of SGPT and SGOT are found in male.

VI.   Conclusions

In this paper we propose the use of modern techniques of data warehouse, ETL and business intelligence to apply in health analytics of OHW-RSU. The system developed is name OHW-RSU BI system. The results of the analytics show 95% of students have good health. However, the health problem is still found mostly in myopia. For staff health analytics, we found that 90% of staffs have good health. Unfortunately, unhealthy staffs are in class 2 and 3 obesity. OHW-RSU BI system much assists in recognizing health problems found in the university. It is for not only the OHW officers to monitor the health problems and specify the unhealthy persons to be further medical treated, but also the executive board to retrieve health information for planning the health care campaigns and promotions of the university.

References

1. MarketingOops, https://www.marketingoops.com
2. BzInsight, https://bzinsight.wordpress.com
3. The Geek CEO, http://thegeekceo.com
4. CompareCamp, http://comparecamp.com
5. Gubalová, J.: The use of Business Intelligence Tools for leadership and university administration. In: 13th International Conference on Emerging eLearning Technologies and Applications, Slovakia (2015)
6. Valdiviezo-Díaz, P., Cordero, J., Reátegui, R.: A Business Intelligence Model for Online Tutoring Process. In: IEEE Frontiers in Education Conference, USA (2015)
7. Mungsang, C.: Epidemiological Alert System by Using Business Intelligence and Epidemic Forecast by Using Data Mining. Thai National Research (2015)
8. Ruengjutipopan, T.: Business intelligence system for analyzing revenue budget, Faculty of Medicine, Chiang Mai University. CMU Intellectual Repository (2012)
9. Saikaewtat, V.: Development of business intelligence system for voluntary automobile insurance, the union prospers insurance Co.,Ltd, CMU Intellectual Repository (2012)
10. Jitrapakdee, P.: Business Intelligence for Sugar Cane Production Management. Thai Library Integrated System (2010)
11. Boonthai, P.: Sales Analysis of Siam Global House Public Co.,Ltd. Wiangkumkam Branch by Using Business Intelligence Principles. CMU Intellectual Repository (2009)
12. Pentaho Data Integration (Kettle), http://www.pentaho.com
13. Jaspersoft ETL, https://www.jaspersoft.com
14. CloverETL, http://www.cloveretl.com
15. Power-Center, https://www.informatica.com
16. Talend Open Studio, https://www.talend.com
17. Kaul, C., Kaul, A.: Comparitive Study on Healthcare Prediction systems using Big Data. In: IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems, India (2015)
18. Kaewchompoo, T.: Data Warehouse System for Analysis of Diagnosis and Illness Primary Healthcare Units. KU Library Catalog (2010)
19. Apache Software Foundation, https://www.apache.org
20. Chennamsetty, H., Chalasani, S., Riley, D.: Predictive Analytics on Electronic Health Records (EHRs) using Hadoop and Hive. In: IEEE International Conference on Electrical, Computer and Communication Technologies, India (2015)
21. Surak, S.: Data Warehouse System for Student of Chiangmai University. CMU Intellectual Repository (2015)

22. Tathong, W.: Development of Schools Data Warehouse in Lampang Primary Educational Service Area Office 3. Thai Library Integrated System (2013)
23. Jatindan, N.: The Development of Executive Information System Using Technical Data Warehouse: A Case Study of Thai Traffic Police Official. Thai Library Integrated System (2012)
24. Suwanpinta, P.: Data Warehousing for Analysis of Rework Problems in Wheel Alignment Process. CU Digital Library (2012)
25. AVOSYS Technology, http://www.avosys.com
26. Data Made Good, https://www.datamadegood.com
27. NTT Data Americas, http://americas.nttdata.com
28. Donald Ham, http://www.infrics.com
29. Gartner, http://www.gartner.com
30. Tableau, https://www.tableau.com
31. Qlik, http://www.qlik.com
32. Microsoft, https://www.microsoft.com
33. G2CROWND, https://www.g2crowd.com

Phatpicha Yochum is currently a Master student in College of Information and Communication Technology at Rangsit University, Thailand. She graduated with Bachelor Degree in Software Engineering from Mae Fah Luang University, Thailand in 2009. She has been working as computer technical officer at Ministry of Education, Thailand.

Chutima Beokhaimook received her Ph.D. in Technology from Sirindhorn International Institute of Technology in 2006. She is now a Director of Master of Science Program in Computer and Information Management, College of Information and Communication Technology, Rangsit University. Her research work includes artificial intelligence, speech recognition and information management.