Implementing Business Intelligence in Grocery Store with Tableau

Atika Firdaus – 13514009
Informatics/Computer Science
School of Electrical Engineering and Informatics
Bandung Institute of Technology (ITB), Ganesha 10 Bandung 40132, Indonesia
13514009@std.stei.itb.ac.id

***Abstract*—Recently, business intelligence has been on the priority list of enterprise CIOs as it is able to help business owners in making their business competitive. This paper will discuss the implementation of business intelligence in the grocery store using Tableau to help top management in decision making.**

Keywords—business intelligence; data warehouse; Tableau

I. Introduction

Data has always been the most important thing for business owners to keep their business alive and remain competitive. Data has become bigger nowadays, people have a term for it, "Big Data". There is a concept explaining what the characteristics of big data are, named 3Vs. A Gartner analyst named Doug Laney introduced the 3Vs concept in a 2001 MetaGroup research publication, "3D data management: Controlling Data Volume, Variety, and Velocity" [1]. This concept is based on these facts: 1) Volume, 90% data in this world are created in these last couple years; 2) Velocity, most of the recent data are time sensitive; 3) Variety, the source of data are evolving and become more and more. Both plain data and big data are needed by businesses to be processed. But, in fact, what business owners will use to drive their company is not data, but the information extracted from it.

In these recent years, business intelligence has become more popular among CIO of enterprises. Business intelligence, BI in concise, is reported as the number one priority in CIOs requirements lists. BI has been the enterprise primary supporting system by providing information for the enterprise which helps them in every decision they make.

One example of a business that deals with a lot of data in its daily process is a grocery store, particularly the big ones. They have to store many data about the products they are selling, who their suppliers are, how their branch outlets are performing, etc. As we all know, there are so many grocery stores and they compete each other for the sake of their business life. Therefore, it will not be a surprise if a business owner, or in this case a grocery store owner, want to do anything to keep their business alive. Being able to predict their business performance in the future turns out become one of the things they need, which a business intelligence is able to do.

II. Basic Theories

A. Data and Information

There are some terms we have to understand when we are talking about data. First, data itself is a stored representation of meaningful objects and events. Data is also known as an unorganized material. There are two types of data according to its form: structured and unstructured. Structured data are data which formats are persistent such as numbers, text, and dates. Meanwhile, unstructured data are the opposite of structured data, its format is not constant and may change. Images, videos, and documents are the example of unstructured data.

Second, there is information. Information is processed and organized data which extracted to increase the knowledge of the person using the data. Therefore, the value of an information may be different from one user to another. An information about the price of fish feeds in the market may be useful for a fish farmer, but a strawberry farmer will not get his knowledge increased by knowing the same information.

B. Business Intelligence

In business intelligence, there are some terms that we have to understand as well. Data integration, data warehouse, and business intelligence. To build a solution from BI, we need a data warehouse [2]. To construct a data warehouse, we need to get through data integration. Here is the further explanation of those terms.

Business intelligence solution is built from operational data which are produced by the lowest management level in the enterprise. A really big enterprise produces a big number of data daily and these data come from many different sources. Among all of the data, there might be some data which have a different format compared to the rest. There might be some missing values too. To make these data usable, we need to integrate the data and cleanse them. Therefore, data integration is needed so that the data have the same formats and consistent. Some of the sub processes included in data integration are data gathering, data transforming, data cleansing, and data aggregating.

After data integration is done, the integrated and cleansed data will be moved to the data warehouse. This process is usually called as data warehousing. Data warehousing is a process of storing information in a physically separated place with operational database and optimizing the information so that the information can be analyzed by the enterprise. There are some characteristics of data warehouses which differs data warehouses to the usual operational databases:

1. Subject-oriented

   A data warehouse is built based on the subject who is going to utilize it. A data warehouse provides data briefly and densely for a specific matter, without providing the data which unnecessary for decision making.

2. Time Variant

A data warehouse has a broader time dimension compared to an operational database. Mostly, an operational database is used for a matter in present time. Meanwhile, a data warehouse provides information from the historical perspective. A data warehouse usually stores data from 5 to 10 years backward.

3. Non-volatile

A data warehouse is physically separated to operational databases. There is not any data updating process in a data warehouse as well.

4. Integrated

As explained before, the information stored in a data warehouse have been integrated before through data integration. A data warehouse is built by integrating multiple distinct data sources. Afterward, the data are cleaned to make sure they are consistent and there are no missing values.

Business intelligence itself is a presentation of data which is made for the importance of business so that business owners can increase their knowledge. BI accommodates information to the business users. BI is the face of enterprise data which can be seen by the business users, while the data warehouse acts as the backroom. The final form of BI is either a visualization, a dashboard, or a report.

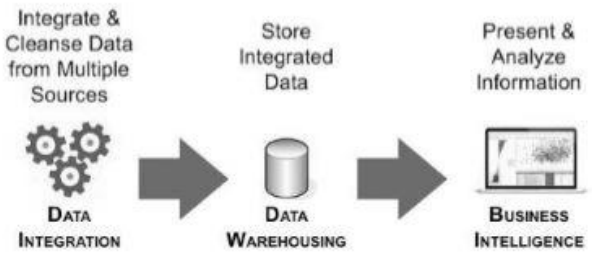

Figure 1 Processes in Building Business Intelligence

C. Data and Dimensional Modelling

A well-designed data model is required to build a data warehouse and business intelligence that is able to give values to the enterprise. Some kinds of the logical data model are entity relationship and dimensional model.

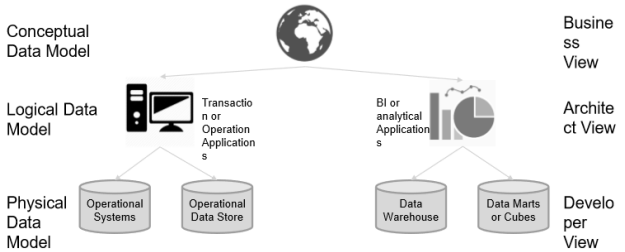

Figure 2 Level of Data Models

For business intelligence, dimensional model suits better than entity relationship. Dimensional model is able to describe the whole business process in the enterprise and manage the data structure at the same time. There are three main concepts in dimensional modeling: facts, dimensions, and attributes. These concepts are designed through a schema.

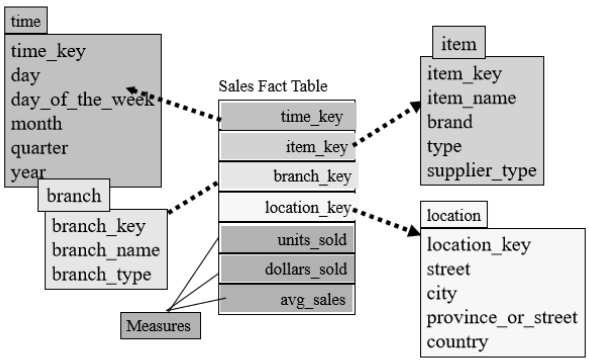

Figure 3 Example of Dimensional Data Model

In dimensional data modeling, there are two kinds of a table. The first one is fact table, and the other is dimension table. A fact table describes measurements of a particular business activity. There are two parts of a fact table: fact keys and fact measures.

Fact keys are a set of foreign keys which is related to the dimension tables. The relationship between a fact table and a dimension table is one-to-many. Meanwhile, the fact measures are numeric measurements of the business activity. For example: the sum of income, the average number of consumers per day, and the order quantity. Every time you are going to define a measure in your fact table, make sure each row in the table has the same unit of measurement.

There are three kinds of fact measures. The first one is additive, meaning the measure can be added, such as product quantity and total profits. The second one is semi-additive, meaning the measure can be mathematically operated in a limited dimensions/units, such as bank balance. The last one is non-additive, the opposite of additive. A non-additive measure also called as a value-per-unit measure.

In the other hand, a dimension table is an entity which describes the business contexts for particular measures in an enterprise. A dimension table is built from reference tables in operational databases. Four characteristics of dimension tables are descriptive, complete, unique, and valid.

D. Tableau

Tableau Software is a software company which produces interactive data visualization products focused on business intelligence [3]. Tableau provides several ways to connect to a data source. Some of them are Excel, JSON, MySQL, and Oracle. Tableau has a relatively easy way of usage. A simple click and drag-and-drop are the actions we have to do to interact with Tableau. Tableau also provides some types of data visualization, including fill maps, histogram, and tree view.

Figure 4 Home Screen of Tableau

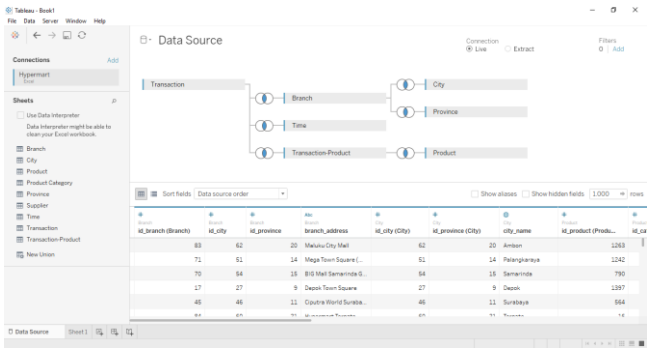

Figure 5 Joining Tables in the Data Source in Tableau

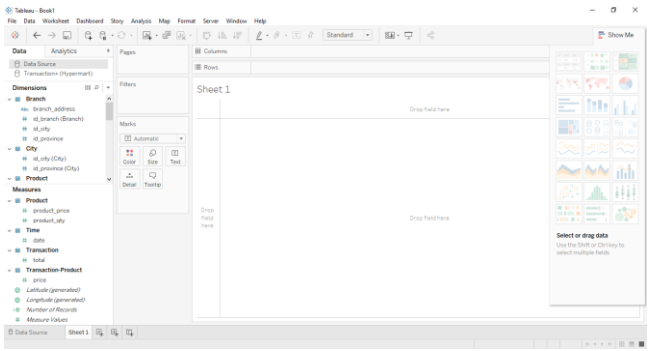

Figure 6 Tableau Worksheet

III. METHODS

When building a BI, the very first step we need to do is deciding what are the decisions we will give help to in its making by providing information related to them through BI. Keep in mind that BI is generally used by the top managements, so the decision that will be helped by the information provided by BI should be a big and general decision which affect the enterprise in general. In this paper, I will pick the decision "the location where a new branch will be built" as the case.

After that, we need to analyze what are the information required in the process of making the related decision. The information will be gathered from operational data and referential data. The required information for the strategic decision I have chosen before as seen in Table 1.

Table 1 Required Information for Strategic Decision

Strategic Decision	Required Information
Decide the location where a new branch outlet will be built	Income of the grocery store per province
	Income of the grocery store per city
	Income of the grocery store per branch
	Number of branch outlets per province
	Number of branch outlets per city

By knowing the information, top management will get some knowledge about which location gives the highest income to the enterprise. A high income indicates a high demand from the customers in the related location, which can be a reason why the enterprise shall build another branch outlets in that particular location. The number of branches in each location may help the top management to decide whether they shall build a new branch outlet in a new location or just add another outlet near the existing branches.

The next step after we know what are the required information is to do the dimensional modeling. A well-designed data model will help us making a well-built data warehouse and business intelligence. The dimensional data model can be seen in Figure 7.

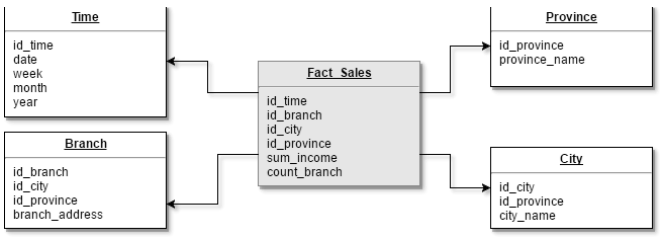

Figure 7 Dimensional Data Model

Furthermore, I generate some random data to be the object of this study case. To generate the data, I use mockaroo.com and the feature to randomize some values on Microsoft Excel. Since Tableau is able to connect a data source in the form of Excel file, I decided to save my data in an Excel file. Each sheet acts as a table in a database.

IV. RESULTS

The main visualization can be seen in Figure 8. It shows the grocery store income per province in the fill map view. Initially, I grouped the dimension province name, city name, and branch name in a hierarchical. I did this so that the data visualization can be drilled down. After that, I dragged the hierarchical group I made before to the column pane and the measures of total under transaction to the row pane. Then, I changed the view to the fill map view. I also added count distinct of id_branch of Branch to the Marks card so that it shows the number of branches in each location.

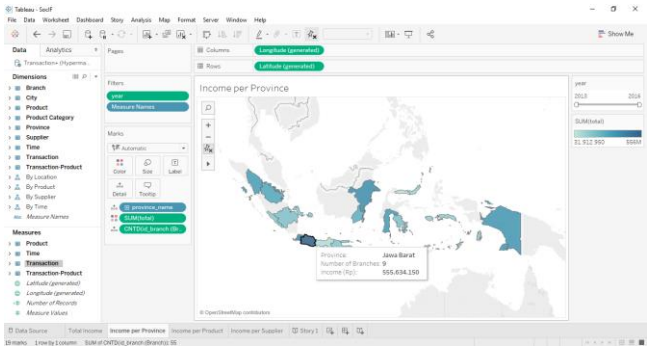

Figure 8 Visualization of Income per Province

To drill down the data, I clicked the plus symbol in the hierarchical group which I have dragged to the card before. For a better appearance, I changed the view to horizontal bars. The data visualization that shows income per city along with its province can be seen in *Figure 9*.

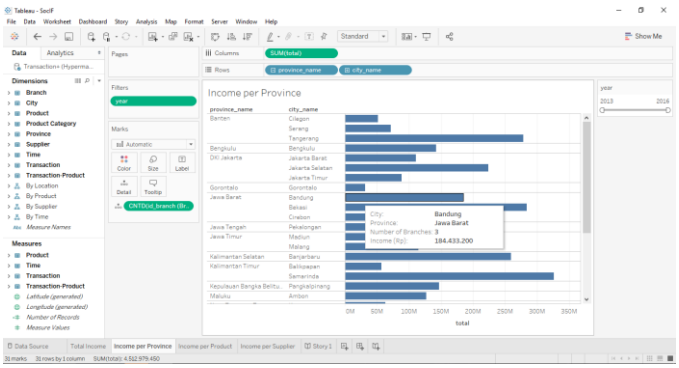

Figure 9 Visualization of Income per City

The final visualization I made is for showing the income per branch. I simply did the same thing as I did before to drill down the data. The data visualization that shows income per branch along with its province can be seen in Figure 9.

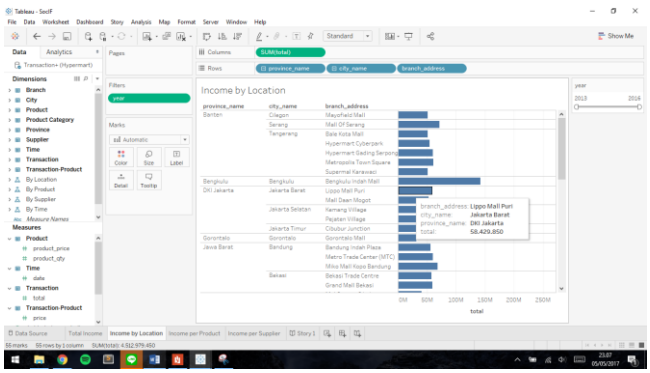

Figure 10 Visualization of Income per Branch

V. Conclusion

To keep a business remain competitive, business intelligence may give some great help to the business owner. Although, business intelligence solution is not a cheap thing to buy. So, business owners have to think carefully whether purchasing a business intelligence is a good investment for their business.

Acknowledgment

I would like to thank Allah SWT for His guidance since I started making this paper until it is finally done. I would also express my gratitude to our three lecturers of IF3280 Socio-Informatika dan Profesionalisme, Dr. Eng. Ayu Purwarianti, ST.,MT., Dr.Dessi Puji Lestari, ST.,M.Eng., and Dr. Ir. Rinaldi Munir, MT. Thank you to my parents who have been supporting me. Last but not least, thank you to all of the people around me who have accompanied me through my hardest times.

References

[1] What is 3Vs (volume, variety and velocity)? http://whatis.techtarget.com/definition/3Vs accessed on Thursday, May 4, 2017 at 21.32 WIB.

[2] Sherman, Rick. 2014. *Business Intelligence Guidebook: From Data Integration to Analytics*. Burlington: Morgan Kauffman.

[3] Tableau Business Intelligence. http://www.tableau.com/business-intelligence accessed on Friday, May 5, 2017 at 22.27 WIB.

Statement

I hereby state that this paper I have written is original. It is not an adaptation, a translation of one's paper, nor a plagiarism.

Bandung, May 5, 2017

Atika Firdaus
13514009