Design of Data Warehouse for Minimarket's Sales Information Using Tabular Models

I Gusti Ngurah Agung Jaya Sasmita[1*], Made Sudarma[2], and Dewa Made Wiharta[3]

[1]Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
[2,3]Department of Electrical and Computer Engineering, Udayana University
*gillcoding23@gmail.com

Abstract Alongside with development of technology, more business need support of business intelligence tools to transform its data in a meaningful way which is accumulated rapidly in every transaction occurred. In order to achieved that, data warehouse with its online analytical processing (olap) can be considered as a solution. One tools for creating an olap from data warehouse is SSAS. This tools consist of two models, first one is multidimensional models or can be said the traditional one and the new one is tabular models. One best feature that tabular models give is the time needed for building it. The tabular models does not require the data in format of fact and dimension tables so reduce the time to change particular data into some scheme (star or snowflake). Hence, the purpose of this research is to design a data warehouse for minimarket's sales information using the advantage of tabular modeling for creating the analysis. Its also known that, the data analysis created using tabular, support reporting application such as pivot table excel, so the user not need to create a new application for reporting analysis.

Index of Terms—**Data Warehouse, OLAP, Reporting, Tabular Models.**

I. INTRODUCTION

Alongside with development of technology, more business need support of business intelligence tools to transform its data in a meaningful way which is accumulated rapidly in every transaction occurred. This data is a major asset for every business. One business that having many transaction data daily and accumulated in a rapid time is minimarket. The trend of processing this data into some meaningful information comes into data warehouse with its online analytical processing (olap) as a solution.

One tools for creating an olap from data warehouse is SSAS. This tools consist of two models, first one is multidimensional models or can be said the traditional one and the new one is tabular models. One best feature that tabular give is the time needed for building it. The tabular models does not require the data in format of fact and dimension tables so reduce the time to change particular data into some scheme (star or snowflake) and time needed to learn it before develop it [11]. However, some data type is more suitable to keep it using multidimensional rather than tabular. For example, data that have complexity many to many relationship, its better to use multidimensional as currently the tabular not support it yet [12].

Some researches before also found that data warehouse with olap can be a solution for business intelligence in selling batik [6] to figure its out the consumer behavior and characteristic in buying batik from the transaction record. Another research also add the easier access for querying data which is collected from several sources using centralized data warehouse [7]. In addition for that, research [8], [15], [16] also tells with data warehouse, the data that already organized in some way, can be used also as a desicision support system.

II. PURPOSE OF PAPER

Creating a design of data warehouse for minimarket's sales information using tabular model as an alternative way for faster and simpleness of development analytical processing data rather than multidimensional models from data warehouse.

III. LITERATURE REVIEW

A. Data Warehouse

Data Warehouse is a data collection that assembled in a way in order to have some meaning and later it can be used for a decision support making by analyzing it in a format of OLAP, data mining, querying, reporting and many more [1]. Data warehouse have four main characteristic such as, subject oriented, integrated, non-volatile, and time-variant [14]. This kind of data commonly being saved to dimensional or normalized data from an operational source system peridiocally. That means the data contains in data warehouse is a historical data of a business operation. [2].

Some benefits if business implement the data warehouse are [4],

1) Able to manage large data consistently.
2) Able to execute rapid analysis on the data.
3) Able to determine how well current business run using the analysis provided.
4) Able to monitoring business more effective.

B. Extract Transform Load (ETL) Process

When we creating a data warehouse, the data from source system need to be extracted, cleanse, transform and loaded to the data warehouse before it can be used for query or analysis [3]. This process needed as the needs to collect, filter, process, and combine of differences data types and structure that coming from source system [9]. ETL process is the most important process when creating data warehouse and take the most development time. It can be more explained as follows [5],

1) Extraction
   This is the initial process for taking what data will be needed in data warehouse.
2) Transform
   This process will change data types and structure to the new correct format and do some cleansing regarding the data noises.
3) Load
   This process will load the data into data warehouse structure commonly star or snowflake scheme.

C. Multidimensional Models

Data warehouse structure comes from dimensional modelling concept that being design for data querying and high volume data retrieval. It consist of fact table and dimensional table. Fact table is a descriptive attributes while dimension table is the classification and aggregation information about the fact table. How this two table linked is a scheme. Star scheme is the simplest and widely used that consist of one fact table surrounded by many dimension table. Illustration shown in figure 1 [1].
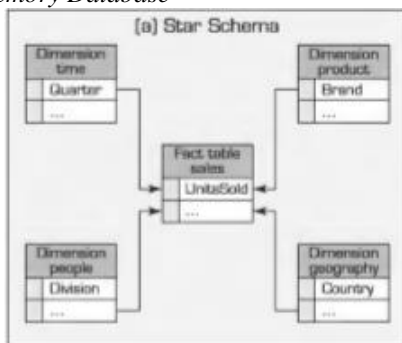
D. In Memory Database

Fig. 1. Star Scheme [1]

In-memory database is a database that using main memory (RAM) as for its storage, which is quite different with traditional database that used disk storage. Its commonly used for saving business transaction data, which later will be aggregated to the traditional database for reporting. Biggest reason why this type database used is querying data (insert, read, update and delete) as fast as accesing RAM and eliminate time needed for accesing data from disk drives like the traditional database. For example, with 1.8 GHz processor, read data approximately 5 microseconds, while insert around 15 microseconds [13].

E. Tabular Models

Around 2012, with the development of SQL Server Analysis Services (SSAS), Microsoft improve their ranks of tools for business intelligence. SSAS provided with two models, which are Multidimensional and Tabular [11]. Tabular models build on main memory for storing data in column oriented database (xVelocity) for upgraded performance in querying while it is quite different with multidimensional models that used disk drive like traditional relational database in row oriented format [12].

Tabular modeling using Data Analysis Expression (DAX) as its native language for query and calculations. In addition, it can also use MDX query that being used in multidimensional models. In contrast, multidimensional also can use DAX, however some features may not working correctly [12]. One of the best feature tabular models offer is no need to load the datawarehouse in scheme like star or snowflake.This improvement enable to quicker design and not wasting time in changing the data into the particular scheme [11]. This also means, ETL process is still needed when it comes from various sources, but we can dismiss the load process into scheme or if it is come from single sources we can even ignoring the transform and load process.

Some limitations tabular models compare with multidimensional models are, tabular still not support many to many relationship [12]. Tabular models that built on based in memory database needs sufficient memory for load all the data with its additional data structure for support query [10], which is current issue right now is the capabilities of memory in managing the data size cannot be as big as disk storage.

F. Tabular Models Principles

Development of tabular models consist of four main principles, such as [11],

1) Tables
   Contain columns and rows of data that will be populated into tabular model. Using model designer, this table later can be view as diagram or a grid view.
2) Columns
   When the data loaded into column of tabular model, it will automatically mapped into data type such as currency, date, decimal number, text, true/false, and whole number.
3) Measures
   How the data will be aggregated on the tables using some calculations DAX [12]

4) Relationship

How data in one tables connected with data in other tables. Currently only two supported, such as one-to-one and one-to-many with some rule of only one active relationship between tables.

One of additional principles are Hierarchies, this principle define how data between columns linked. This is used for better user interfaces.

G. SQL Service Analysis Services

SQL Service Analysis Services (SSAS) is developed by Microsoft in purpose for enhancing their bussiness intelligence tools via online analytical process (OLAP) database. With this type of database, it provides optimization for queries and calculations that often used in business intelligence. Later this result of queries and calculations can be used mainly for decision support and business analytics. SSAS support two types of OLAP, which are multidimensional model and tabular model [12].

IV.　RESULT AND DISCUSSION

A. Datawarehouse Design

How this minimarket data warehouse designed is illustrated on figure 2. The data comes from several source system which is needed to be processed using ETL to data warehouse then continue passed through tabular models before its reach the user.
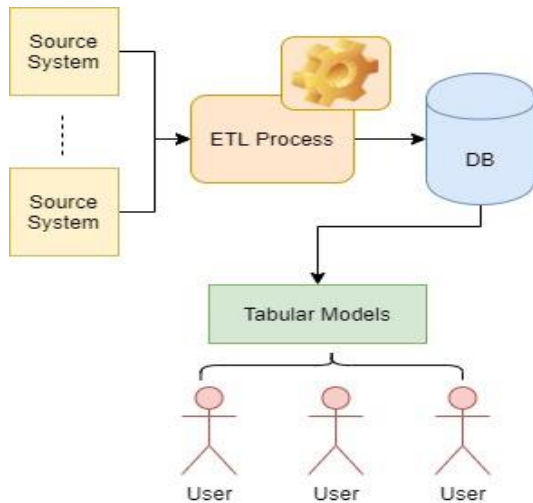

Fig. 2.　Data Warehouse Design

In this design, Firstly, ETL process runs for extracting the data from source system that needed for creating report. Second, the data will be processed through transform stage to make it clean, consivstent and standard format. Finally, the data will be simply loaded to data warehouse and finish. Because for tabular modeling, it is not require for the data loaded into scheme (star or snowflake).

This data warehouse then ready to be processed using tabular models. The data warehouse table is shown on figure 3.

Fig. 3. Datawarehouse Data Table

B. Designing Tabular Model

The data warehouse which is ready to be processed into tabular models will be following several steps such as,

1) Create connection

Data warehouse database server needs to be connected with the tabular models designer either use of windows authentication or SQL server login.

2) Importing tables

User will choose which tables will be used for business analysis queries, calculations and reporting, in this design all tables will be loaded into the designer.

3) Managing column

After all tables loaded into the designer there are some steps that needs to do in order for enhance the user friendliness such as,

- Change all colomn name for better naming, for example column 'ProductName' in table product can be changed to 'Product Brands' and so on
- Change column data type for 'TotalTransaction' and 'TotalCostOfGoods' from float to currency with zero decimal, hence it will from for example 3000,00 to Rp. 3000
- In this steps, if there is a date column with unique value, it can assigned as date table, in order to use the DAX Time Intelligence Function, for example in case of filtering data between dates. In this design the date is not unique, because in a day many transaction occurs, hence that date column in not unique.

4) Reviewing relationship

When the data being imported from data warehouse, it also imported the relationship between table according its foreign key. In this design no additional relationship is added.

5) Create calculated columns

Calculated columns can be describe as a custom columns based on calculation done in the data using DAX expression. Some calculated columns added for this tabular design such as,

- Added calculated columns for 'Total Gross Income', 'Total Net Income'
- Added calculated columns for separations of date

with its user friendliness such as, '2020-01-20' this will added to 'Display Date' = '20 January 2020', Month = 1, Quarter = 1, Semester = 1, Year = 2020. Some example implementation of the DAX expression are,

- 'Total Gross Income'
  =Transaction[TotalTransaction] - Transaction[Total CostOfGoods]
- 'Month'
  =MONTH(Transaction[Date])

6) Measures

Like calculated columns, creating measures is via DAX expression. The difference calculated column is for row level, this measures is for column level or can be said aggregated value.

- Added measures for 'TotalAllTransaction'
  'TotalAllTransactions:=SUM(Transaction[Total Transaction])'
- Added measures for 'TotalAllGrossIncome'
  'TotalAllGrossIncome:=SUM(Transaction[Total GrossIncome])'

This measures later will be used for the reporting for example total gross income grouped in month and product category.

7) Create KPI (Key Performance Indicator)

KPI can be defined as how to identified how well the business target achieved. For example, in we can calculated the KPI for 'TotalGrossIncome' every month. It is also can be view as a graph display.

8) Creating Perspectives

Perspectives is used for specified the target of its reporting result or analysis. For example, the sales information perspective only shows the one related with it, user then can create new perspectives for another business target. In this design, its just simple perspectives for sales information.

9) Creating Hierarchies

Hierarchies is used for grouping columns into levels. Previously the design added the date separation columns, it can be used for hierarchies. The data later will be grouped into Year > Semester > Quarter > Month.

10) Deploy tabular

After finishing its design, the tabular need to be deployed to the server instance for use in the application.

C. Usage of Tabular Models

There are several ways to use this tabular models, first like connected to reporting client such as pivot table in excel. It only need drag and drop. For example, drag the measures 'TotalAllGrossIncome' into values field, the x axis is branch name, the y axis is month. Later the 'TotalAllGrossIncome' data will be sliced into branch and month.

For custom application, it can get the data via dax query (native) or mdx query that used for in multidimensional models. Example like above, grouping data measures 'TotalAllGrossIncome' into values field, branch name and

month.

- MDX query
  "SELECT
      { [Measures].[TotalAllGrossIncome] } *
          [Transaction].[BranchName].MEMBERS ON 0,
      { [Transaction].[Month].[Month] } ON 1
  FROM [Model]"
- DAX query
  "EVALUATE (
      SUMMARIZE('Transaction',
          [Month],[BranchName],
      "Total All Gross Income", [TotalAllGrossIncome])
)
  ORDER BY [BranchName],[Month]"

V. CONCLUSION

Based on discussion that is stated above, tabular modeling data warehouse is expected as an alternative way for faster and simpleness development of analytical processing data. It also support some basic reporting tools such as pivot table in excel, so not giving user a hard time for developing new reporting application. As per data warehouse is used, now growing data must not become a problem in term of business inteligence and the result of analysis in tabular modeling is also expected to give user more broad horizon regarding its own business and give some decision support helper.

VI. SUGGESTION

Recommendations for further development is the design of minimarket's sales information data warehouse can be really implemented with more real world complex data and applied some change data capture (cdc) method in order to decreasing the time needed in ETL process.

REFERENCES

[1] Sharda, R., Delen, D. and Turban, E. 2015. Business intelligence and analytics. 10th ed. New Jersey: Pearson.

[2] Rainardi, V. 2008. Building a Data Warehouse: With Examples in SQL Server. New York: Springer Verlag.

[3] Kimball, Ralph. 2004. The Data Warehouse ETL Toolkit : Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Indianapolis : Wiley Publishing, Inc.

[4] Partha, I., Weking, P. and Mertasana, P., 2019. Data Center Data Warehouse Development at Z Bali Clinic Using the Kimball Nine-Step Method. International Journal Of Engineering And Emerging Technology, 4(1), pp.53-59.

[5] Denny, Atmaja, I. P. M., Saptawijaya, a. And Aminah, S. 2017. Implementation of Change Data Capture in ETL Process for Data Warehouse Using HDFS and Apache Spark. 2017 International Workshop on Big Data and Information Security (IWBIS).pp. 49-55

[6] Bhaskara, I. M. A., Suardani, L. G. P. And Sudarma, M. 2018. Data Warehouse Implemantation To Support Batik Sales Information Using MOLAP. International Journal of Engineering and Emerging Technology, 3(1), pp. 45-51.

[7] Pramita, M. D. P., Subiksa, G. B. And Saputra, K. O. 2018. Design of Intermediate Medical Record Information System in Electricity Patient Using Rolap Warehouse Data. International

Journal of Engineering and Emerging Technology, 3(1), pp. 56-61.

[8]   Pradhana, I. G. N. A., Giriantari, I. A. D. And Sudarma, M. 2018. Analisis dan Perancangan Sistem Pengelola Data Menuju Implementasi Data Warehouse Untuk Mendukung Administrasi E-Procurement. Majalah Ilmiah Teknologi Elektro, 17(2), pp. 245-250.

[9]   Iswardani, P., Pramana, I. and Saputra, K., 2019. Design of Data Warehouse for Monitoring Hotel's Food and Beverage Cost. International Journal of Engineering and Emerging Technology, 4(1), pp.1-5.

[10]   Docs.microsoft.com. 2020. Comparing Analysis Services Tabular And Multidimensional Models. [online] Available at: <https://docs.microsoft.com/en-us/analysis-services/comparing-tabular-and-multidimensional-solutions-ssas?view=asallproducts-allversions> [Accessed 18 May 2020].

[11]   Wilson, D., 2017. Tabular Modeling With SQL Server 2016 Analysis Services Cookbook. Birmingham: Packt.

[12]   Russo, M., Ferrari, A. and Webb, C., 2015. Microsoft SQL Server 2012 Analysis Services. United States of America: Microsoft.

[13]   Sawant, N. and Shah, H., 2014. Big Data Application Architecture Q&A: A Problem - Solution Approach. New York: Apress.

[14]   Narendra, A., Murpratiwi, S. and Sudarma, M., 2017. Design of E-Grant Application Data Warehouse. International Journal of Engineering and Emerging Technology, 2(1), pp.11-15.

[15]   Kotama, I., Adnyana, A. and Saputra, K., 2019. Design of Data Warehouse for University Library using Kimball and Ross9Steps Methodology Case Study: Udayana University Postgraduate Library. International Journal of Engineering and Emerging Technology, 4(1), pp.37-40.

[16]   Sumichan, A., Yudiana, I., Satrio, M. and Sudarma, I., 2019. Designing a Virtual Data Warehouse in Supporting Sales Information Needs (Case Study: National Scale Building Material Store X). International Journal of Engineering and Emerging Technology, 4(1), pp.45-48.