Business Intelligence for Product Defect Analysis

To cite this article: A S Girsang et al 2019 IOP Conf. Ser.: Mater. Sci. Eng. 598 012117

View the article online for updates and enhancements.
Business Intelligence for Product Defect Analysis

A S Girsang¹, S M Isa¹, A L Haris, Arwan¹, K Mandagie¹, L R Ariana¹ and V Ardinda¹

¹Computer Science Department, Bina Nusantara University, Jakarta, Indonesia 11480
agirsang@binus.edu

Abstract. The problem of reporting service in PT.XYZ is the complexity of the query to take the data from the table in oracle. There are many table to be processed for the data, because of that it’s hard to show the report quickly. PT XYZ needs fast and reliable report to be shown, especially for control or audit. The suggested solution is to use data warehouse, because it has more structured storage and the data can be analysed by OLAP analysis that has the capability to show the data quickly. Data warehouse development was done through nine step that has been designed by Kimball and Ross. With this solution, PT XYZ has the capability to maintain quality of the product by monitoring the process and the data.

1. Introduction

PT. XYZ is one of FMCG Company that operates in Indonesia. PT. XYZ has produced many products that have been distributed to all of places in Indonesia. To ensure the customer feel satisfied with the product, the company always make innovation of new product and maintain the quality of the product. Report service is used by quality service employee every day to maintain the quality of product. Report service in PT. XYZ faces some problem in its daily use. Report service can’t show the data quickly, because of that it’s a trouble when the monthly result is needed. The problem occurs because of the huge amount of data in database and the amount of time that report service need to take the data. It’s a huge problem for daily usage because of the amount of time employee need to see the data and to monitor the production line. Currently, employee use manual report to record the data. Although manual report has no problem to show the data daily but it needs extra work when the data is needed for the monthly statistical report.

The solution offered in this case study is to set up and move the data to a data warehouse. A data warehouse is a special kind of a database that is organized according to different design principles then databases are. Its purpose is to integrate (physically) data from different sources and provide a trusted place of important pieces of information [1]. The source of data that used in the report service comes from many different sources. In data warehouse, various sources of data can be called data mart. Diversity of data source can cause some problem in the process of integrating data to a data warehouse because of the difference in data type or data definition [2]. Once successfully integrated, data can be found in data warehouse and decision making becomes much simpler and much faster as well [1]. The data warehouse provides the solution needed by the manager to make decision about the process of the production. To fulfill the need of a report service, the data in data warehouse can be processed to Online Analytical Processing (OLAP) analysis. OLAP is a technology for building multidimensional data structure called “cube”, it help the decision makers to explore data in a multidimensional way, and browse the aggregates according to the different hierarchical levels of each dimensions in order to
extract relevant information depending on the level of granularity chosen [3]. The aggregates and relationship of data from each dimension of database is determined by admin.

2. Related Work

According to Kimball and Ross, database management system is a computer application whose main objective is to save, retrieve, and modify data in a very structured way. Data within the DBMS are usually mutually shared with the application [4]. Data warehouses have the potential to provide business intelligence solutions for companies looking for competitive advantage [5]. Data warehouse can be built from the top division to bottom division in the company or vice versa. The top-down approach starts with overall design and planning. It is useful in cases where the technology is mature and well known, and where the business problems that must be solved are clear and well understood. The bottom-up approach experiments and prototypes. This is useful in the early stage of business modelling and technology development [4].

A data warehousing system is required to integrate data from all different aspects of the corporation. This data warehouse can then process complex repetitive tasks using Extract-Transform-Load (ETL) code [6]. After ETL process has been done to all the data, data warehouse needs to be modelled for report. The entity relationship (ER) data model is commonly used in relational database design, where a database schema consists of a set of entities and the relationship between them. The ER model is used to demonstrate detailed relationships between the data elements [7]. The database schema is a design that contains description of relation of each database. Data warehouse schema design usually contains at least one fact table and some dimension table. Fact table contains the data of transactional process in database. Dimension table contains data explaining in more detail from fact table. Dimension table is build based on the specific business process.

OLAP is a process that can provide analysis for all of data that has been collected. Type of analysis depends on requirement. To facilitate various kind of analysis, the data is stored in multidimensional table. The data in multidimensional table has a relationship that connect each of the data. OLAP is a technology for building multidimensional data structure called “cube”, it help the decision makers to explore data in a multidimensional way, and browse the aggregates according to the different hierarchical levels of each dimensions in order to extract relevant information depending on the level of granularity chosen [3].

The source in an OLAP system is the server that supplies the data to be analyzed. Depending on the use of the OLAP product, the source could be an data warehouse, a legacy database housing corporate data, a collection of spreadsheets that holds financial data, or a combination of any of the above [8]. Logically, OLAP servers present business users with multidimensional data from DW or data marts, without concerns regarding how or where the data are stored [9].

3. Methodology

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 Analytical. The following are the nine steps of creating a data warehouse:

Choosing the Process. The selected process is the process of monitoring production line from factory in PT. XYZ. In general, activity of monitoring production process is divided into three category which is the incoming, production and laboratory. Monitoring incoming involves checking the resource coming from supplier. If it passes the quality standard, the resource will be received and delivered into production line. Monitoring production line involves checking the production process and result of machine production. If it passes the quality standard, employee will take sample from the production to the laboratory. In laboratory, employee checks the quality standard in the term of chemical and nutrition. This case study focuses on the process of integrating data from production line to data warehouse. The purpose is to generate report quickly and dependable for audit or control operation.

Choosing the Grain. Choosing the grain is the process to decide what a record of the table to represent in each row of fact table. In this case of study, data comes from the production line parameter.
To meet the requirement of the report, grain from the fact table is “one line of data contain information of one check condition quality meets the standard or not”.

**Identifying and conforming the dimensions.** Dimension is selected according to the requirement of data warehouse to the report service. Each row in fact table stores information of time, day, employee id, factory, product, and check name.

**Choosing the fact.** This phase creates fact table based on the collected data and the selected grain. Each data can be analyzed and presented as report. Table 1 explains that fact table is created based on the available information.

| Table Name | Description |
|------------|-------------|
| Fact_report | Id           |
|            | Id_date      |
|            | Id_time      |
|            | Id_username  |
|            | Id_productcode|
|            | Id_plantcode |
|            | Id_headercheckcode|
|            | Id_checkcode |
|            | Subcheckcode |
|            | Grade        |

**Storing pre-calculations in the fact table.** Once the fact table has been created, each row of data re-examined to determine whether there are opportunities to use calculation. Calculation was done by compare the value of check parameter and its quality standard.

**Rounding out the dimension table.** This phase determines what properties to include in dimensional table. This phase covers information such as descriptions of the dimension tables and the list of tables. The dimensions are dim_date, dim_time, dim_user, dim_plant, dim_product, dim_headercheck, and dim_check

**Choosing the duration of the database.** This phase determines the duration of the data in data warehouse. Duration of data in this case study will use data of monitoring in April, May, June and July. The database used in PT. XYZ is Oracle Sql Developer.

**Tracking slowly changing dimensions.** This scheme uses slowly changing dimension type two. If the data change, the new data become a new record on table in data warehouse. So, the old data were still there on the table as history.

**Deciding the query priorities and the query modes.** This step considers most critical physical decision issues that affect the end-users perception include physical sort order of the fact table on disk. This step also determines additional physical design issues include administration, indexing data for performance, indexed views, storage, backup, and security.

Based on the nine steps to design a data warehouse, a star schema can be built with one fact table and some dimension tables as seen in Figure 1.
4. Analysis Result

4.1. Extract, Transform, and Load

ETL processes are responsible for the extraction of the appropriate data from the sources and their transportation to a special-purpose area of the data warehouse where they will be processed [10]. As one can observe, an ETL process is the synthesis of individual tasks that perform extraction, transformation, cleaning or loading of data in an execution graph [10]. The required data for ETL process is taken from transactional database which contains raw data. Data from transactional database is processed to entity data warehouse. The process includes copy, insertion, delete, calculation, etc.

Figure 2 shows one of dimension ETL process, time dimension. The process generate time for a whole day and add sequence id for each of it after that the value split into hour and minute before loaded into time dimension table. There is another dimension date. This dimension is same as time dimension. The process generate date for two years and add sequence id for each date after that the value from the process split into day, month, and year before loaded into date dimension table. ETL process of user dimension, plant dimension, product dimension, header check dimension, and check dimension have same process before loaded into their dimension. Once all the dimension tables have been created, ETL
process for fact table is performed. The process load data from transactional table and looking up standard value from standard transactional table. Foreign key from is loaded to fact table representing value from each dimension as seen in Figure 3.

4.2. Design OLAP Cube

OLAP analysis was done to show report analysis from each dimension tables. OLAP cube was needed to build OLAP analysis. In OLAP cube scheme, there is a hierarchy of dimension and measure in accordance with the requirement of PT. XYZ.
Figure 4. Date dimension hierarchy

Figure 5. Time dimension hierarchy

Figure 6. Product dimension hierarchy

Figure 4 shows scheme hierarchy of date dimension table. The hierarchy consists of level year, month, and day that allow user to drill-down or roll-up data based on year, month, and day. Beside date dimension, there are time dimension that has same hierarchy with date dimension as seen in Figure 5, and the other dimension which have simple hierarchy as seen in Figure 6.

4.3. Report and Dashboard Design

Based on OLAP analysis, report and dashboard can be made to present the data. For example, product defect report based on product dimension as seen in Figure 7. From this report, management can see total of product defect based on product. This information is useful to determine what follow-up action that needed to take. Report may also be presented in the form of a pie chart as seen in Figure 8. This report contains information based on factory where production defect happened. By using pie chart for this information, management can see the data easily.

Figure 7. Product defect by product name

Figure 8. Production defect by plant dimension

Figure 9. Production defect by date dimension
Another example is production defect report based on day and month as seen in Figure 9. This report provides information about production defect pattern in day and month scale. Management can use this information to see the day production defect usually happen and compare it with another month.

The way to build multiple charts and report on one page is to use dashboard. Dashboard allows management to see a brief overview of production defect based on various dimension. The dashboard consists of four panels: which are based on plant, product, date, and time dimension as seen in Figure 10. First panel show data about total product defect in each plant. Second panel shows data based on product in production line. Third panel show data based on monthly production data. Fourth panel show hour of the time that productions defect usually happened. This show vital information for management to determine what action needed to take to solve this problem.

![Product Defect Report](image)

**Figure 10. Dashboard report**

5. **Conclusions**

Data warehouse help PT.XYZ quickly present data or information that is helpful when necessary such as when there is a control or audit. In addition to being the solution for the existing problems, PT. XYZ could have benefited more from data storage in warehouse comparing to only storing data from manual paper. Presentation of data can be done in two forms: in the form of report that has been built based on the needs of company, and data in data warehouse dashboard so the data can be stored in data warehouse so that user can monitor the data history.

The recommendation for the next research is to make the application to change how to take the data by creating mobile application that can directly put the data into data warehouse. Putting the data directly into data warehouse can make a report system that has real-time data for immediate action by superior.

Take an immediate action for every production defect that happened can make product quality increased and customer happy.

**References**

[1] K. Rabuzin. Deductive Data Warehouses. 2014 Int. J. Data Warehous. Min. 10 16-31.
[2] M. O. Olaru and M. Vincini. 2014 *J. Digit. Inf. Manag.* 12 255-66.
[3] R. Erratta, M. Fakir, and F. Z. Salmam. 2014 *J. Inf. Technol. Res.* 7 63-78.
[4] A. S. Girsang and C. W. Prakoso. 2017 Data Warehouse Development for Customer WIFI Access
Service at a Telecommunication Company. 7 114-24.

[5] N. Rahman and Nayem. 2016 *Int. J. Knowledge-Based Organ*. 6 21-37.

[6] N. Rahman and D. Rutz. 2015 *Int. J. Intell. Inf. Technol*. 11 1-22.

[7] J. Yao, J. Wang, Q. Chen, and R. Xing. 2013 *Int. J. Robot. Appl. Technol*. 1 57-66.

[8] A. H. Slaughter. 2001 Introduction OLAP System Components – MIT pp 1-8.

[9] A. Parekh. 2013 *Int. J. Eng. Comput. Sci*. 2 2569–73.

[10] P. Vassiliadis. 2009 *Int. J. Data Warehouse. Min*. 5 1-27.