Design and analysis on data warehouse of personnel administration system using time series algorithm

S Alviana[1], B Kurniawan[2]

[1]Departemen of Magister Information System, Universitas Komputer Indonesia, Indonesia
[2]Departemen Electrical Engineering, Universitas Komputer Indonesia, Indonesia

sopian.alviana@email.unikom.ac.id

Abstract. The purpose of this study was to design and analysis on data warehouse of personnel administration system using time series. The success of the company in performing its business activities is definitely supported by reliable system administration job. Abundant data within the company; employee information, attendance data, leave the data, the increase in the classroom, increase salaries and allowances requires good management in the order. To assist in the management of personnel data in the study, analysis and design on data warehouse is done in the administration section personnel to facilitate the collection of employee data, monthly and annual reports. The method used in the data warehouse design, data retrieval is done form related sources through extraction, transforming, and loading (ETL) process. After the data warehouse is formed, then conducted data analysis of the presence of the employee with a period of time, either the monthly attendance data analysis as well as using the annual time series algorithm. Process analysis is used to look at trend data is the presence of any employee that can be used to analyze the performance of each employee. This process then continues with the creation of reports on performance of employees which can be used for strategic decisions at the management level. The results obtained from this study showed that the application of data warehouse in staffing information systems conducted to obtain information that can support strategic decisions at the management level. Thus, the design and implementation of data warehouse in power system can help collect data, analyze data, create a report and present the data in a form of employment that can be used to aid strategic decisions at the management level.

1. Introduction

The existence of information technology in the era of globalization many significant changes in the world economy and business, the company must be attention to the development of technology with business processes [1]. Personnel administration system is one system that requires the support of information technology in the administration of the staffing process data, employee data, i.e. data on leave, attendance data, and data related to the process of administration employee.

In generally the existing personnel administration system are still manual, i.e. a system that functioned only keep staffing data and display it. The manual system raises the risk that can disrupt business processes such as a search of the old data, duplication of data, the report is not composed, and data that are not yet integrated. Previous research has many that use the data warehouse as well as the time series as a data storage and analysis of data.

Data warehouse is a technology that aims to facilitate in decision making from a variety of data by doing OLAP [2,3]. Sonal sharma uses a data warehouse to analyze each of the criteria used in the application of business intelligence, research generates data warehouse that can be used to assist in decision making and forecasting data period ahead with the existing history data [4]. Marcos Aurelio using the data warehouse to support automation website, such research generates data warehouse that can perform automation in different contexts [5].

Solomon addisu using time series analysis to measure and identify the impacts of climate change by analyzing the time trends in temperature and precipitation time series analysis, the results can determine the mean temperature maximum and minimum averages that occurs so it can determine strategies and plans that address [6]. Thus, this research aims to make the design of the data warehouse as a data storage system administration personnel are integrated and can make a quick report by using time series analysis for use in a more appropriate decision making at the management level.

2. Research methodology

Research methodology includes the following phases of building a data warehouse and analysis time series. Data warehouse development process consists of the identification of data source, ETL processes, build the cube data, as well as the creation of datamart [2,4,5,7]. After the data warehouse is formed, to look at trend data based on time can be used in time series analysis [6]. The results of the analysis can help in retrieval strategies and associated decisions relating to such data [8].

2.1. Data warehouse

The process of building data warehouse staffing administrative system are described in the following steps, as illustrated in **Figure 1**:
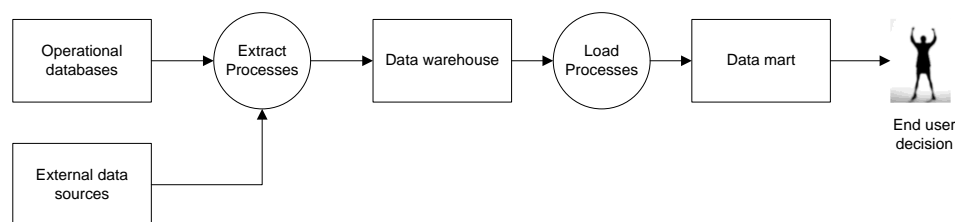
Figure 1. Data warehouse architecture.

- Identification of data sources
  On the process design of data warehouse system required the staffing process of the identification of any data source [3]. Table data source is needed into the data warehouse is the employee data table, data on leave, attendance data.
- Extract, Transform, and Loading (ETL)
  Data warehouse has different data as well as from different sources. The difference in the data required uniformity of data by performing the process of extraction, transform, and loading (ETL) [3]. Extraction process used to obtain the relevant data before it is stored in the data warehouse. Transform phases needed to maintain consistency of the data. Whereas, the loading process is loading the data into the data warehouse. The process of data preparation it is ETL, regularly either daily, weekly, monthly or yearly [9].
- Build cube
  The making of the cube is the process of the establishment of a database to identify each dimension table to the fact to be made to the need for data storage in the data warehouse. Build the cube is stages in the process of Online Analytical Processing (OLAP) that is used to collect the data that has been put together so that it can speed up the query results [10].
- Create datamart

Data warehouse creation process of making datamart. Datamart can be summed up as a data warehouse used for support in the process of reporting and analysis on dimension data is already built in the cube.

2.2. Analysis time series

Time series analysis is used to view data trends that occurred in the staffing system. The data analyzed includes employee data, attendance data, data on leave, and promotion. Historical data is the data from previous years will be analyzed and made report. The results of the analysis used to support strategic decisions at the management level related to the administration of the staffing process everything.

3. Results and discussion

On the design of the data warehouse system administration personnel is divided into several sections, namely the administration of staffing, administrative leave, and attendance data administration. Following the results of the process of the data warehouse:

3.1. The results of the identification of data sources

On the process of identification of the source data that is showing the tables that the data source being used [12]. The data from the table which will be entered into the data warehouse administration staffing. The following table of the data source that is used (Table 1).

Table 1. Table of data source.

Table Personnel (Tb_Staff)		Table Section (Tb_Section)	
Field Name	**Description**	**Field Name**	**Description**
emp_no	Number of employee	id_section	Employee ID section
name_emp	Employee name	name_section	Employee name section
place_birth	Employee place birth	**Table Leave (Tb_Leave)**	
date_birth	Employee date birth	**Field Name**	**Description**
sex	Employee gender	id_leave	Leave ID
religion	Employee religion	name_leave	Name of leave
marital_status	Employee marital status	type_leave	Type of leave
Tb_Period			
Field Name		**Description**	
id_periode		ID period	
year_period		Year of period	

On table 1 describes the tables data source are used for the construction of a data warehouse. Personnel tables used for employee data, the table section is a description of the data portion of each employee, the table leave is leave data explanation of each employee, the table period data for the period of attendance of employees. Identification of the data source table is part of the process of operational system on data warehouse architecture [9]. Study on the identification of the source of data is the initial stage of the analysis database, while carried out by Moody's is a form of operational databases.

3.2. ETL process results

ETL process is the process of the selection of the field needed to be admitted into the tables facts and dimensional data warehouse. After the extraction process, conducted the process to transform that is changing the selected data into the data as required in the administration of personnel. Then, the last process is the process of loading, where do the addition of data is already extracted and at transform into data warehouse system administration personnel. The ETL process is done from the tables data source then do the extraction to get the dimensions and the fact that it takes on the process of data warehouse system administration personnel. Then the process transform is used to generate the needed facts in analyzing data warehouse, as well as the loading process to store that data into the data warehouse. The ETL stage performed almost the same with research conducted by Stephen Gardner, namely data transformation in which every data be transformed to fit the needs of each data [11]. Similarly, the process ETL i.e. adjust each data format is needed in building a data warehouse.

3.3. The results of cube

Process cube is one of the processes in the data warehouse to determine any fact which dimension will be used. Model dimensions are listed in table 2.

Table 2. Dimension model cube.

Modul	Dimension	Description
Employee administration	Employee data	Employee data is used to see the spread of employee data based on gender, age, educational status, as well as married.
	Title	Information about the office employees, parts, and the other.
	Violation	Information of violations, the type of violations, as well as employees who commit violations.
	Increase	Information showing the time of the promotion or the employees, the employees who do promotion or the.
Leave	Employee leave information	Used to view the type of leave used employees, time off work, and information employees who take leave attendance.
Employee attendance	Information the number of attendance	The number of attendance employees, the number of not present, sick, duties, and so on.

Table 2 shows the dimensions tables used this fact in the system administration personnel. Dimension consists of employee data, title, violation, either increase or the salary, leave, and the employee attendance data. Any dimensions that form is used in order to make of the fact that this mediation will be used in the data warehouse. In building a data warehouse, the process of building a cube is a very important thing that is to facilitate to processing of data with Online Analytical Processing (OLAP) [12]. The construction of a cube is the basis in determining each table the fact that this will be used in making the staffing system administration reports with data warehouse.

3.4. The results of the datamart

The model of the schema used in the datamart that is formed is the fact table that is formed on a data warehouse that is:
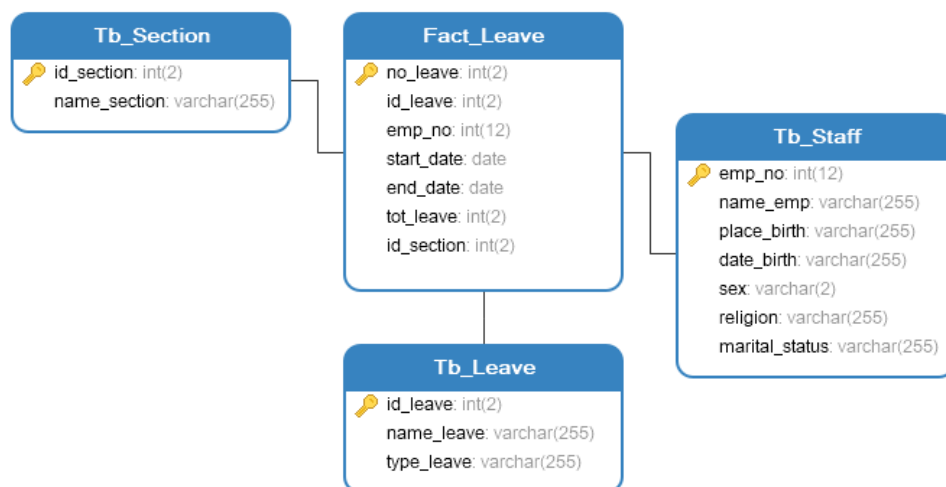
3.4.1. The fact of employee data (Figure 2).

Figure 2. Fact of employee data.

From the fact table employee data in Figure 2 shows that the visible information employee data that include the number of employees based on gender, relationship status, religion, place of birth, and the old work (Figure 3).
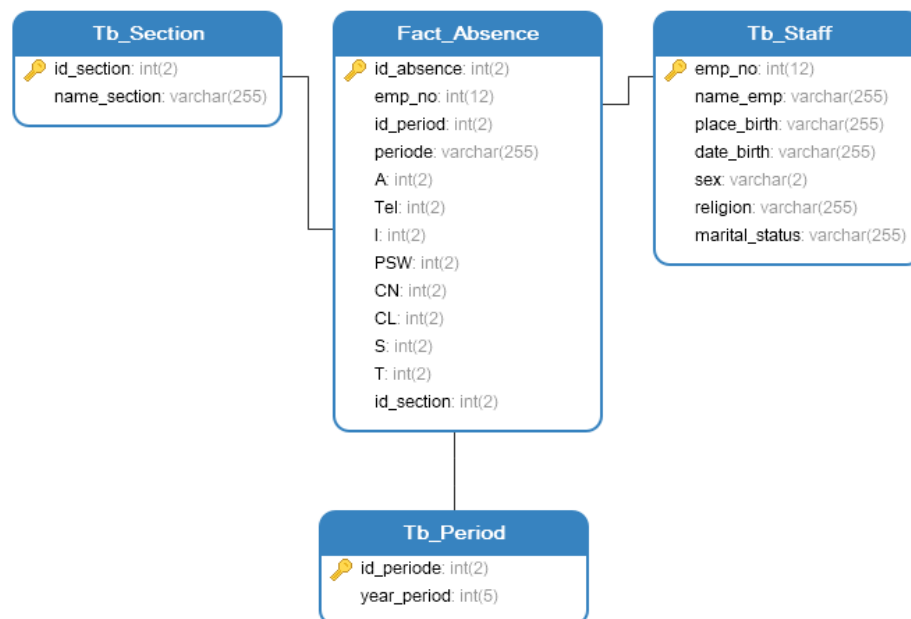
3.4.2. The fact of leave data

Figure 3. **Fact of leave data.**

From the fact table data for the indicated figure 3 on leave can be obtained information includes the number of employees who do leave, the leave taken, type the old on leave, and time off work employee.

3.4.3. The fact of attendance data (Figure 4).

Figure 4. Fact of attendance data.

Figure 4 shows the fact table data for an attendance used to determine the amount of information the attendance of any employee, the type of attendance, as well as the period of attendance employee.

3.4.4. Analysis time series. In the last stages of the design of the data warehouse built on the system of the administration of staffing that is using time series data analysis. Time series analysis is used to view data based on number of employees began to work, attendance data, data on leave either annual or specific period. Here are some of the design of the graphs based on time series analysis are used namely data with time over five years (Figure 5).
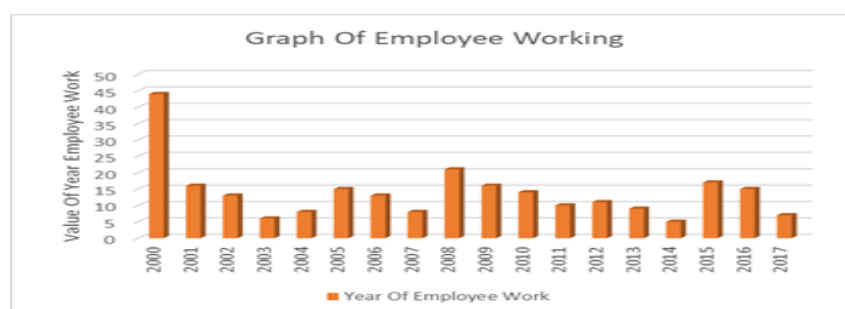
Figure 5. Graph value of employee.

In figure 5 shows the staffing data namely the amount of each employee in each year of the time span of data with more than ten years. Within the last five years can be seen in the information occurs fluctuations in addition and reduction of number of employees, it can be used as information for the management level in taking decisions.

4. Conclusion
The use of data warehouse in system administration personnel are used for the needs of the human resources management information administration staffing. Such information can be used to speed up the creation of employee data reports, attendance data, data on leave.

Acknowledgements

References
[1]	Tarigan R E and Dewi K S 2015 Web-based implementation of e-marketing to support product marketing of chemical manufacturing company *CommIT (Communication and Information Technology) Journal* **9** 2 p 73-82
[2]	Sharma S and Jain R 2013 Enhancing business intelligence using data warehousing: A Multi Case Analysis *International Journal* **1** 7 p 40-56
[3]	Domingues M A, Soares C, Jorge A M and Rezende S O 2014 A data warehouse to support web site automation *Journal of the Brazilian Computer Society* **20** 1 p 11-20
[4]	Boutkhoum O and Hanine M 2017 An integrated decision-making prototype based on OLAP systems and multicriteria analysis for complex decision-making problems *In Applied Informatics* **4** 1 p 11-15 (Springer Berlin Heidelberg)
[5]	5 Aissi S, Gouider M S, Sboui T and Said L B 2015 A spatial data warehouse recommendation approach: conceptual framework and experimental evaluation *Human-centric Computing and Information Sciences* **5** 1 p 30
[6]	Addisu S, Selassie Y G, Fissha G and Gedif B 2015 Time series trend analysis of temperature and rainfall in lake Tana Sub-basin, Ethiopia *Environmental Systems Research* **4** 1 p 25–40
[7]	Faisal S, Sarwar M, Shahzad K, Sarwar S, Jaffry W and Yousaf M M 2017 *Temporal and Evolving Data Warehouse Design. Scientific Programming*
[8]	Parmar K S and Bhardwaj R 2014 Water quality management using statistical analysis and time-series prediction model *Applied Water Science* **4** 4 p 425-434
[9]	Moody D L and Kortink M A 2000 From enterprise models to dimensional models: a methodology for data warehouse and data mart design *In DMDW* **4** 3 p 5–20
[10]	Han J, Pei J and Kamber M 2011 *Data mining: concepts and techniques* (Elsevier)
[11]	Gardner S R 1998 Building the data warehouse *Communications of the ACM* **41** 9 p 52-60
[12]	Han J, Stefanovic N and Koperski K 1998 Selective materialization: An efficient method for spatial data cube construction *In Pacific-Asia Conference on Knowledge Discovery and Data Mining* **5** 2 p 144-158 (Springer, Berlin, Heidelberg)