IMPLEMENTING DATA WAREHOUSE AS A FOUNDATION FOR DECISION SUPPORT SYSTEM (PERSPECTIVE : TECHNICAL AND NONTECHNICAL FACTORS)

Tanty Oktavia

Bina Nusantara University, School of Information Systems, Jakarta – 11480, Indonesia

E-mail: tanty_oktavia@yahoo.com

ABSTRACT

A company needs a system to conduct the process of analyzing data in order to support decision maker takes the best decision. Data warehousing system as one solution can accommodate that need. The system has been accepted as a key which enables company or organizations to improve their abilities in the data analysis, the decision support for managerial, and the automatic extraction of knowledge. With the developing of information included in the decision making process, the considered data become more complex, in both structure and semantics. The aim of this research is to identify the classification technical and non-technical factors on data warehouse implementation to help the management decide what to be prepared for the company if they want to develop data warehouse project. It also explores on how to integrate the data warehouse system into strategic business process. The study results show that the successful implementation of data warehouse influenced by how the company aware about technical and non-technical factor when implemented the project.

Keywords: *Data Analysis, Decision Support, Technical, Non-Technical*

1.   INTRODUCTION

For over a decade, the developing of information compared to previous data quantities is much different. Many data from multiple sources grew number of reports with goal to reach competitive advantage in industry. Transferring and sharing of information within the company, among divisions in different locations as well as among business partners is important [1]. Therefore, company begins to initialize a system to accommodate and synchronize all of data in order to support strategic decision.

Analyzing data from a lot of point of view is not simple for many organizations because those manage to understand any kinds of aspects both technical and non-technical, such as characteristics and structure of data transaction to combine with the company needed. The most successful companies are those that capable to respond quickly to market. It can happen, if the company can process their core business values become new initiative as a guide.

Based on that idea, company can not only focus on transaction, but also must realize on analyzing and monitoring data. According to this problem, a key to response is the effective and efficient to manage data become information that valuable for the company. Data warehouse can be the solution for this matter. Data warehousing system is an integrated data repository system that designed for consolidating data become report and for keeping a collection history of the company sustainable. This system will solve problems for many divisions in the company that oriented with the strategic area.

Current data warehouse systems are not only to support reporting but also to provide mode advanced analysis to meet the needs of different users. That system collaborate data from all heterogeneous information sources, which part in the core business process and transform them into a multidimensional representation for support decision maker in the company. Data Warehouse as a foundation that facilitates business intelligence area, such as data mining, Online Analytical Processing (OLAP), and decision support system. Apart from a complex architecture, involving operational data sources, warehouse manager, ETL manager, query manager; is also characterized by a complex design lifecycle because a database designer must identify decision maker perspective align with the data warehouse system. Therefore, data warehouse become long duration projects.

In a design phase, the database designer has to employ and maintain a conceptual model from Online Transaction Processing (OLTP) and a logical schema, accompanied by a physical design become a new structure for a data warehouse. The database designer must also deal with data warehouse procedure to meet the needs of decision maker, which are complex in mapping structure, numerous number of data and hard to explore the data.

In recent years, data warehouse concept has developed to allow for real time synchronization between operational data and warehouse data. Attempting to reduce the time delay between the creation of operational data and extraction data in the warehouse is not easy like common thought. There are issues and challenges that must solved. The problems categorized into two groups:

1. Technical

This problem referred to what infrastructure the company used, such as selection data, warehouse architecture, schema, data integrity, security, data availability, etc

2. Non Technical

According to this, the problems are procedure, management, user, and skills.

For those reason, the research will describe what components that must be noted for the company to apply data warehouse system become streamline of the business process. This research differs from previous, because it emphasizes what the significant technical and non technical factors according to the implementation of data warehouse.

2.   THEORETICAL BACKGROUND

Information and decision-making are related to each other. Making good decisions requires good information because with a lot of information decision maker can forecast the potential alternatives that can be choose as an outcome. All information can be obtained from operational system, but information from operational systems is too detailed to be prepared as information for decision maker. As data processing systems record more and more data, it is not enough for analyst to produce more management reports to be viewed by decision makers [2]. Therefore, a company needs a single data warehouse system that generates summary data for management.

Operational systems that connect with database operational were never designed to support decision-making activities. Because it is not an easy way to produce strategic report using Online

Transaction Processing (OLTP), complicated queries might take a long time to answer and also might degrade the performance of the OLTP [3]. Data warehousing or Online Analytical Processing (OLAP) is focus on process analyze data from many point of view or user's perspective. Unlike Online Transaction Processing (OLTP) databases, which are business application-oriented, detailed, and operational [4], Data Warehousing (table 1) is "a subject-oriented, time variant, integrated, and non-volatile, non-updatable collection of data to support management decision-making processes and business intelligence system." [5]. The information warehouse was proposed to allow company to use data archives widely for acquiring information from multiple sources and delivering it to manager in order to gain a business advantage. [6]. Moreover, all levels of management decision-making processes are supported by Data warehouse through the collection, integration, extraction, transformation, and interpretation of both external and internal data.

Table 1: Characteristics Of Data
Source : (Inmon, 2005)

Characteristics	Description
Subject-Oriented	Warehouse is organized around major subject of the enterprise rather than major application areas
Integrated	The data warehouse integrates corporate application-oriented data from different source system, which often includes data that is inconsistent
Time Variant	Data in the warehouse is only accurate and valid at some point in time or over some time interval
Non-Volatile	Data in the warehouse is not updated in real-time but is refreshed from operational systems on a regular basis

Before exploring into the main objectives, it is crucial to differentiate a Data Warehouse and operational data stores/OLTP. OLTP systems are designed to maximize the transaction processing capacity, while data warehouse are designed to support report query processing. Table 2 describes a comparison of the characteristics of OLTP Systems and Data Warehousing Systems.

Generally, a company has a number of different operational database systems for support business processes. These systems optimized for a high number of transactions that large number of concurrent operational, detailed, and subject to change. In contrast, a company will need a data warehouse system, which holds data that is historical, summarized, and rarely modify. The data warehouse is organized the potential queries and supports analytical requirements of a lower number of user.

Table 2 : Comparison Of OLTP Systems And Data Warehousing Systems
Source (Connolly & Begg, 2010)

CHARACTERISTIC	OLTP SYSTEMS	DATA WAREHOUSING SYSTEMS
MAIN PURPOSE	SUPPORT OPERATIONAL PROCESSING	SUPPORT ANALYTICAL PROCESSING
DATA AGE	CURRENT	HISTORIC (BUT TREND IS TOWARDS ALSO INCLUDING CURRENT DATA)
DATA LATENCY	REAL-TIME	DEPENDS ON LENGTH OF CYCLE FOR DATA SUPPLEMENTS TO WAREHOUSE (BUT TREND IS TOWARDS REAL-TIME SUPPLEMENTS)
DATA GRANULARITY	DETAILED DATA	DETAILED DATA, LIGHTLY AND HIGHLY SUMMARIZED DATA
DATA PROCESSING	PREDICTABLE PATTERN OF DATA INSERTIONS, DELETIONS, UPDATES, AND QUERIES. HIGH LEVEL OF TRANSACTION THROUGHPUT	LESS PREDICTABLE PATTERN OF DATA QUERIES; MEDIUM TO LOW LEVEL OF TRANSACTION THROUGHPUT
REPORTING	PREDICTABLE, ONE-DIMENSIONAL, RELATIVELY STATIC FIXED REPORTING	UNPREDICTABLE, MULTIDIMENSIONAL, DYNAMIC REPORTING
USERS	SERVES LARGE NUMBER OF OPERATIONAL USERS	SERVES LOWER NUMBER OF MANAGERIAL USERS (BUT TREND IS ALSO TOWARD SUPPORTING ANALYTICAL REQUIREMENTS OF OPERATIONAL USERS)

Data warehouse contains cleaned, aggregated, consolidated large volumes of data that is accumulated in multidimensional data structure to support multidimensional analysis. Not only a Data warehouse recognizes the need for current and next future data, but also recognizes the need for historical data, for example trend analysis requires a great deal of historical data regardless the company size. Data warehouse allows a business organization to manipulate a great deal of data in ways that are useful to it. Examples of typical data warehouse queries:

1. What was total revenue for Indonesia in two quarter 2013 ?
2. Which type of property sells for prices above the average selling price for properties in the main cities of Jakarta ?
3. What was total revenues for property sales for each type of property in Sumatera in 2013 ?
4. What are the three most popular areas in each city for the renting of property in 2013 and how does this compare with the figures for the previous two years?
5. What is monthly revenue for property sales, compared with rolling 10 monthly prior figures?

These queries can be produced in OLTP, but with OLAP the query process will be made to be simple without checking complicated structure or disturbing operational process. Because many ad hoc query processing will use substantial resource of the system to generate the output that user need. (Figure 1)
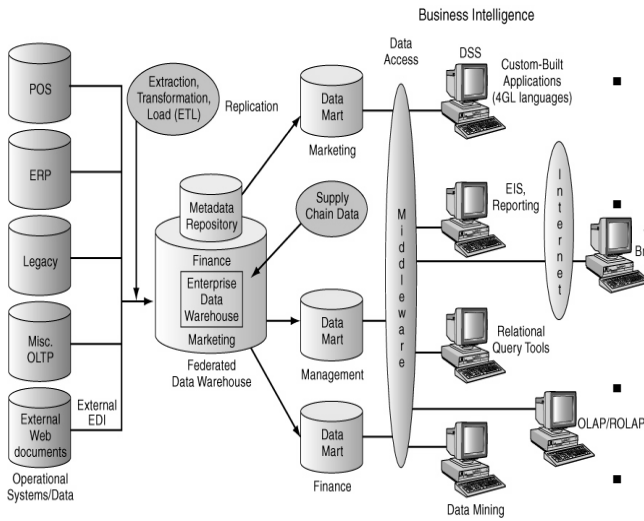
Figure 1: Typical Architecture Of A Data Warehouse
Source : (Rainer Jr. & Cegielski, 2013)

Architecture of a data warehouse (Figure 1) consists of:

- Operational Data Store
  This store is a repository of current transactional data, which detailed and real time data.
- Extract Transform Load manager. It is to performs all the operations associated with the extraction and loading of data into the warehouse. Complexity and size will vary between data warehouses and may be constructed using a combination of vendor data loading tools and custom built programs
- Warehouse manager. It is to perform all the operations associated with the management of the data in the warehouse. It constructed by using vendor data-management tools and custom-built programs. Operations performed include :
  o Analysis of data to ensure consistency
  o Transformation and merging of source data from temporary storage into data warehouse tables
  o Creation of indexes and views on base tables
  o Generation of de-normalizations
  o Generation of aggregations
  o Backing-up and archiving data
- Query Manager. It is to perform all the operations associated with the management of user queries. The complexity determined by the facilities provided by the end user access tools and the database.
- Detailed Data. It is an order to stores all the detailed data in the database structure. In most

cases, the detailed data is not stored online but aggregated to the next level of detail.

- Lightly and highly summarized data. It is an order to store all the pre-defined lightly and highly aggregated data generated by the warehouse manager.
- Archive/Backup Data. This area stores detailed and summarized data for the purposes of archiving and backup. It is necessary to backup online summary data if this data is kept beyond the retention period for detailed data.
- Metadata. This area stores all the meta-data (data about data) definitions used by all the processes in the warehouse.
- End user Access Tools. These tools are used as an intermediary between user and data warehouse to interact each other.

Although, there are a lot of benefit can be taken from data warehouse, but company must be considered about a few potential problems associated with the data warehouse, such as performance, scalability, availability, which must be identified when defining the system. In some cases, a company is confined with company structure or internal problem that not support for the system, such as procedure, management, user, and skills. Therefore, it is so necessary for the company to understand and take apart when data warehouse project is running so that all of the perspective can be summarized into the project.

Furthermore, there are more features to support the data warehouse, such as monitoring data loading, integrity checker, updating metadata, purging data, archiving and backing up data, security management, etc. These feature as complement for the data warehouse, which integrated in the DBMS.

3. RESEARCH METHOD

There are a lot of methods that can be used to build data warehouse system, but the important thing must be considered is the consistency the method with the data warehouse architecture implemented should be linear to collaborate the functional system. Sometimes the methodology is referred to as top down approach and or otherwise, as bottom up. The pattern will have an impact when structuring or mapping the process to place considerable emphasis on initially putting the infrastructure and focus on delivering a solution that addresses a current business need. The top down approach start with the importance of

developing a few high level models incrementally. The bottom up approach begins at the fundamental of having an enterprise plan for integrating the incrementally developed data marts. Kimball's methodology that would directly be analyzed problem question about loading, cleaning, computation, aggregation, and integration. The methodology comprises a set of stages, involving the design of the quality goal, the evaluation of the status, the analysis and improvement of this situation, and finally the re-evaluation of the achieved plan. The data from one company were analyzed following a content analysis structure. For this study, the structure can be defined for making inferences by systematically and objectively identifying shared common properties.

4. RESULT AND DISCUSSION

Depending on the critically of the data warehouse, most of business organizations appear to have implemented this system to support strategically function, but on average the stage to its accomplishment has been overwhelmed with failures. There are variety technical and non-technical problems facing business organizations when implementing Data warehouse projects.

Based on survey respondents and experts personal insights, there are several factors defined, such as cost consideration, a need to implement, politics, E-R data modeling techniques and normalization procedures, expert influence, security and performance, the need of data accuracy, consistency, and control; the complexity of existing decision support system, the ability to find new information requirement, and nature of independents data marts. [7]

For this research, these factors will be classified into five classifications. It can be seen at Figure 2:
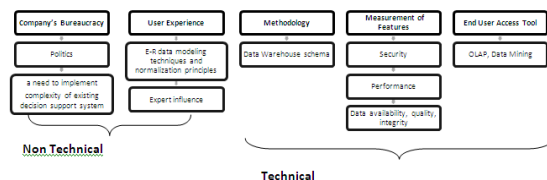

Figure 2: Classified Factors Of Data Warehouse

This could be referred to the fact that Data warehouse project is not simple project, that project needs organizational readiness, legalization and endorsement all parties involved, complete risk management, familiarity with technology, and skills of system analysts that can support the project.

In the following section, both technical and nontechnical issues have been described in detail:

Non-Technical

❖ **Company's bureaucracy**

Building data warehouse for the company is a large investment project. It might consideration planning decided to invest in this project. Otherwise, company commitment through executive management's endorsement, involvement, and support from all division of company is an important aspect for an innovation to be adopted and subsequently used efficiently and effectively to business value in order to sustain favorable implication.

❖ **User Experience**

As a broad based project, the data warehouse certainly requires a strong capability of personnel specialization in organize this system. According to [8], managing the data warehouse requires three kinds expertise:

- Business Expertise
  An understanding of the company's business process and business goals to help in determining what data should be mapped in the data warehouse system.
- Data Expertise
  A comprehension of the company's transactional data and databases to design and manage data transformation for the acquisition of enrichment data
- Technical Expertise
  An understanding of handling data warehouse project, include unique requirements for security, backup and recovery, being split across multiple disk devices.

This part is thus crucial especially for the first stage when the analyst need requirement from the user about the scope and detail of project. Users must give all information about the process or knowledge that they have, so the analyst easily identified what their expectation for the new system. In order to get the most benefit from Data Warehouse, users must endure gradually, formal, and systematic training for such large-scale necessity, this will enable users to understand functions of Data warehouse much better as well as being more accountable for making data warehouse produce higher quality information as an outcome.

Technical

❖ **Development Methodology**

Scoping of the data warehouse project may help analyst to establish which methodology the more appropriate to use. To identify what system to be built, how large the system and other consideration

can direct to the proper methodology. The two top researcher of the data-warehousing field, Bill Inmon and Ralph Kimball, have invented data warehouse methodologies [7]. Inmon advocates the use of the hub and spoke architecture, while Kimball discovers the data mart architecture. There are other methodology alternatives but these two options are fundamentally different approaches. If the company refer to functional division, the will build data mart using bottom up methodology, otherwise the company will use enterprise data warehouse using top down approach.

Data Warehouse Schema

There is not a best single architecture that suitable for all situation and company. If it were that simple, there would not be disagreements over architecture selection. The company should identified what schema can be applied as data warehouse structure, certainly according to what kind of information that to be displayed as an outcome. Not all structure fit to the company business process. In the logical perspective, the schema is concerned with the functionality of an activity, describing what this particular activity is about in terms of production of information. In the physical perspective, the details of the execution of the process are the center of modeling.

❖ **Measurement of Features**

Data warehousing system as analyzing function, must be completed with complement features, such as :

Security

Almost all company data will be stored in data warehouse. However, the data in data warehouse should not be threatened, lost, or manipulated [9]. Many techniques can be applied to prevent data warehouse from any possible risk of losing company data, such as backup activity, disaster recovery plans, authentication policy, Intrusion Detection Systems (IDSs), encryption. Furthermore, data warehouses system need authorization and authentication to protect from user that does not have access right so only the right and secure user can access all data in data warehouse.

Performance

Data warehouse application with its supporting middleware and end user access tools (i.e. OLAP, business intelligence, data mining, dashboards, etc) needs a suitable platform to elaborate data warehouse system. Naturally, company that want to implement data warehouse, should have integrated operational system to support business

process, such as Transaction Processing System (TPS). If the company desire to run data warehouses well, the first point to be thought is compatibility the application and the current platform to support each other so the OLTP and data warehouse performance can operate with excellent capacity.

Data Availability

Data is the most important factor to build data warehouse system. Insufficient information has been identified as the most failure problem for this project. In this reason, a company should analyze, whether data can be involved to the data warehouse, or originally viewed and used only by a particular department of business area. Many user think some of data is sensitive or private to be implemented in data warehouse, so when the data warehouse is implemented, loss information occurred. Therefore, lack of operational system, caused data is not available to convert to the warehouse. In the end, data warehouse system cannot support company needs. This problem can be solved if there is cooperation to support data warehouse project from all part of company.

Data Quality

Data quality involves the accuracy, completeness, and consistency of the data used by company. High quality data in data warehouse can be created with ensuring that the source data from operational data source contain high quality data so it can be combined and loaded into the data warehouse.

Data Integrity

After creating schema, next step is transformation the detailed data residing in the operational data store must frequently be consolidated in order to generate and store. In many cases, data records concerned with the object resides in multiple source systems. Integration with data quality software is often the only realistic way of matching these records. Each system could contain its own variation.

To deal with the complexity of data loading process, there is ETL (Extraction, Transformation, and Loading) Tools. This tool tasks include the identification of relevant information at the source, extraction information, customization and integration of the information coming from multiple sources, cleaning the resulting data set, the propagation of the data to the data warehouse. After that, company must combine the data warehouse system with the user access tool that suitable with company need.

❖ **End User Access Tools**

Establishment of end user access tools that capable of supporting the information requirements of the management also become consideration. This application is the interface used by data warehouse user to access the data. It contains a series of tools, such as OLAP, query applications, analytic applications, data mining, reporting, etc [10]. The essence of the data warehouse is not in storing data and support transactional process, but in making the information available to decision makers through using suitable tools, such as OLAP and data mining, and so on. This process is important because it will decide the data warehouse is completely success or need to improve. Some case, user does not want to use the application because it is difficult to operate, if it happened the analyst must identify what the failure in the system. After that stage, the project can be implemented as planned before.

5. CONCLUSION

Developing a Data Warehouse project proved to be difficult and problematic, but highly return. Data Warehouse is recognized as an infrastructure, many applications can run over it. The information processing requirements and the source of data then combine to determine the view of the data warehouse. Furthermore, Data Warehouse meant to be used by managers since they support decision-making process. It has been observed that there are various issues available when implement data warehouse project whether technical or non-technical, but only company can evaluate what factor definitely needs to be taken into consideration when designing data warehouse so data warehouse can be implemented as well as company need to support strategic analyzing. The limitations of this research are difficult to classify the aspect from many research objects because every project has various uniqueness factors so we have to decide the general point that consider with technical and nontechnical aspect.

REFERENCES

[1] Velicanu, Manole, and Gheorghe Matei. "Database Vs Data Warehouse." *Revista Informatica Economică*, 2007: 91-95.

[2] Singh, Ajit, D C Upadhyay, and Hemant Yadav. "The Analytical Data Warehouse: A sustainable approach for Empowering Institusional Decision Making." *International Journal of Engineering Science and Technology*, 2011.

[3] Rainer Jr., R. Kelly, and Casey G Cegielski. *Introduction to Information Systems.* Singapore: John Wiley & Sons, 2013.

[4] Ramamurthy, K R, A Sen, and A P Sinha. "An Empirical Investigation of The Key Determinants of Data Warehouse Adoption." *Decision Support Systems*, 2008: 817-841.

[5] Inmon, William H. *Building The Data Warehouse.* Canada: Wiley, 2005.

[6] Connolly, Thomas M, and Carolyn E Begg. *Database Systems : A Practical Approach to Design, Implementation, and Management.* Boston: Pearson Education, 2010.

[7] Watson, Hugh J, and Thilini Ariyachandra. "Data Warehouse Architectures: Factors in the Selection Decision and the Success of the Architectures." July 2005. http://www.terry.uga.edu/~hwatson/DW_Architecture_Report.pdf (accessed May 8, 2012).

[8] Gillenson, Mark L. *Fundamentals of Database Management Systems.* USA: Wiley, 2012.

[9] Al-Debei, Mutaz M. "Data Warehouse as a Backbone for Business Intelligence:Issue and Challenges." *European Journal of Economics, Finance and Administrative Sciences*, 2011.

[10] Khan, Md. Ilyas, R K Singh, and P K Dey3. "Parameter Based Quality Measurement of Data Warehouse." *InternationalJournal of Electronics Communication and Computer Engineering*, 2011.