Proceedings of the 33rd International Conference on
Information Technologies (InfoTech-2019)
19-20 September 2019, Bulgaria

ENERGY PROVIDER DATA WAREHOUSE DESIGN AND IMPLEMENTATION – CASE STUDY

***Digest of paper*[1]**

Nertila Kryeziu, Florije Ismaili, Jaumin Ajdari, Bujar Raufi, Xhemal Zenuni

Faculty of Contemporary Sciences and Technologies, South East European University, e-mails: {nk21430; f.ismaili, j.ajdari; b.raufi; xh.zenuni}@seeu.edu.mk
Republic of North Macedonia

Abstract: An important issue in data warehousing is to design suitable multi-dimensional data models to support querying, exploring, reporting, and analysis as required by organizational decision making. This paper shows the design and implementation of data warehouse within context of Energy Provider Company which collects various data from seven different transactional databases and unifies them under single schema, in order to improve the reporting process.
Keywords: data warehouse; ETL; reporting system.

1. INTRODUCTION

Reporting and exchange of information within the enterprise as well as between enterprises is one of the key factors in the outcome and success of the work [2].

The paper focuses mainly on the design and implementation of a data warehouse, respectively in the part of the report generation system, where a theoretical approach has been made and as a case study, 'KEDS' – energy provider company in Kosovo was taken. In the practical implementation part, it clarifies the system architecture which enables the integration of data from transactional databases located in different company departments. The developed prototype demonstrates how a variety of different data bases are subject to the Extraction, Transformation and Load (ETL) process by creating a star schema to make reporting easier. At the end of the work, through a set of experiments, the performance of query execution time in separate transactional databases versus data warehousing, is compared. The main benefits from data warehouse implementation are: users will save time as historical data are retrieved

[1] The full paper is proposed for including in the IEEE Xplore Digital Library

faster, the system can answer many questions simultaneously, users can generate new reports using existing queries or by creating their own queries, the database administrator can easily update the new conditions in the data warehouse tables.

Hence, DW must evolve to fulfill the constraints and criteria allocated by the various people who need the assistance of information preserved in the data warehouse; a survey of various techniques of DW designing is done [3,4,5]. As can be concluded from related work, the work presented in this paper is still valuable for discussion.

2. DATA WAREHOUSE HIGH LEVEL ARCHITECTURE

The Enterprise Data Warehouse is designed conform to studied and current methodologies of data warehousing, where the result is several star schemas combined. The data warehouse design is based on the work of Kimball [1], who introduced dimensions and fact type tables as the manner of data arrangement into data marts and reporting. The figure below provides the way company data moves from transactional databases, through the "ETL" to data warehouse itself, providing to the users the ability to report their data as needed.

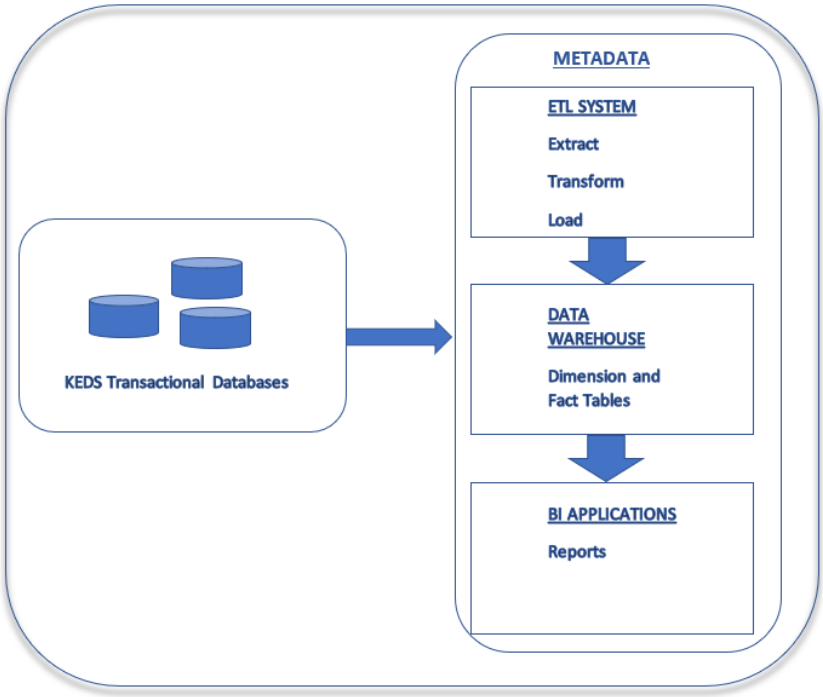

Fig.1. Data Warehouse High Level Architecture

According to Kimball, four components should be considered: operational source systems, data staging area, data presentation area, and data access tools.

The operational source is the transactional databases that are in seven different databases within company.

Data staging area involves ETL principles regarding data processing, namely: Importing data from transactional databases, performing calculations/alterations to the data set imported to fit operational needs, exporting it into the end target - data

warehouse. The data presentation area is the actual data warehouse full of data that is in dimension and fact tables, the standard formatting method of data warehouses.

Finally, the data access tools represent the BI Applications that provide users with the data in a format they need to make decisions effectively.

3. DATA WAREHOUSE PROTOTYPE OF THE ENTERPRISE

The purpose of the "KEDS" company data warehouse design, is to show how the data warehouse in the company can integrate data between existing different systems. The design of company data warehouse began by taking in consideration data sources, technical infrastructure, customer expectations, and budget. The Data warehouse is the SQL Server 2012 database that uses Server Integration Services as a tool for ETL, Server Analysis Services as a Cube Creation Tool, and SQL Server Reporting Services as a reporting tool. The prototype presented in this paper includes a sample implementation of the KEDS Data Warehouse on a significantly smaller scale, but which is appropriate to demonstrate the full-scale implementation.

4. PERFORMANCE EVALUATION

The purpose of this section is to take the real company data to demonstrate the possible improvement in query running time, comparing a transactional database and a de-normalized star schema in a data warehouse. In a large data base, the time of the report generation depends on the complexity of the views and multiple joints to the tables. The expectation in this testing set is that the data returned from the data warehouse will be less time consuming than from running the same query against the transactional database. To test the differences between the two systems, tests are performed to compare the query execution. The tool used for testing is T-SQL. The same queries will be executed in both systems and the results obtained will be shown below.

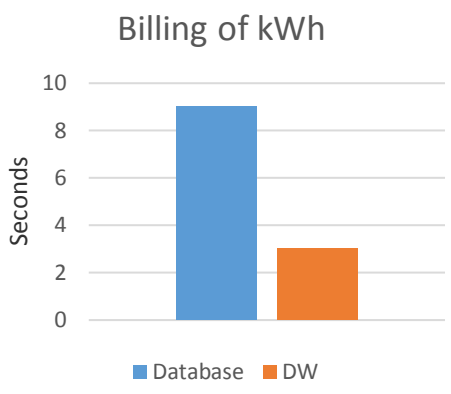

Fig.4. Billing of kWh and number of customers

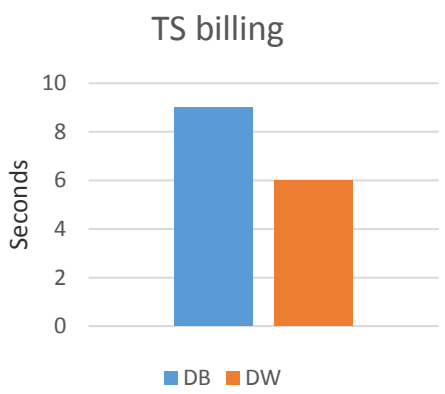

Fig.5. Comparison Diagram of TS billing

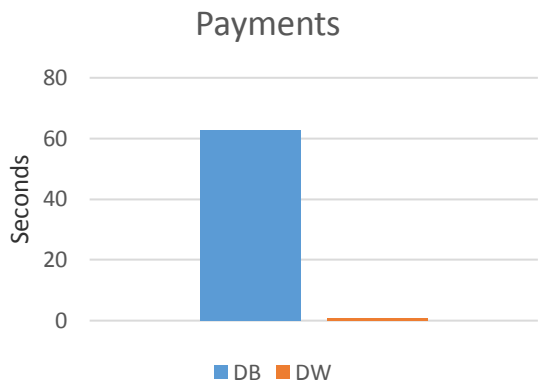

Fig.6. Diagram of Payments Results

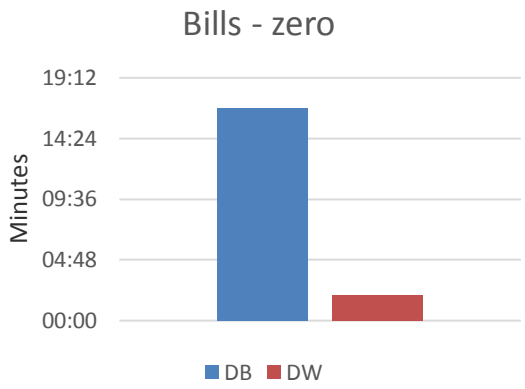

Fig.7. Diagram of bills - zero

As it can be seen from the results, the query execution time in data warehouse outperforms in all cases the same query execution in transactional databases.

6. CONCLUSION

This paper described the design and implementation of an Enterprise Data Warehouse within the context of a specific energy provider environment, with objective to better integrate existing systems for faster, simpler and improved reporting. Through a series of practical experiments, it was shown that query execution time for targeted indicated processes in the data warehouse outperforms the same query execution time on transactional databases

Evaluating the user experiences, especially in the area of decision making is left as future work.

REFERENCES

[1] Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite, (**1998**). *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses*, ISBN: 0-471-25547-5, Published by John Wiley&Sons Inc.

[2] M. Rifaie, K. Kianmehr, R. Alhajj and M. J. Ridley, "Data warehouse architecture and design," *2008 IEEE International Conference on Information Reuse and Integration*, Las Vegas, NV, USA, 2008, pp. 58-63.

[3] Abdullah, Zina A. S. and Obaid, Taleb A. S. (**2016**), Design and Implementation of Educational Data Warehouse Using OLAP. *International Journal of Computer Science and Network - IJCSN*, vol. 5, n. 5.

[4] M. E. Houari, M. Rhanoui and B. E. Asri, (**2017**), "Hybrid big data warehouse for on-demand decision needs," *2017 International Conference on Electrical and Information Technologies (ICEIT)*, Rabat, pp. 1-6.

[5] Xishui Pan, X. Zhou, Hongmei Song, Runshun Zhang and Tingting Zhang, (**2012**), "Enhanced data extraction, transforming and loading processing for Traditional Chinese Medicine clinical data warehouse," *IEEE 14th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Beijing, 2012, pp. 57-61.