DOI: 10.7250/eunis.2013.027

Building data warehouse in Metropolia University of Applied Sciences

Antti Tikka[1]

[1]Metropolia University of Applied Sciences, Bulevardi 31, PL 4000, antti.tikka@metropolia.fi

Keywords
Data warehouse, data mart, architecture, Inmon, Kimball.

1. INTRODUCTION

Universities use many different data systems that store data in different databases. To get maximum benefit from this precious data it is necessary to build a data warehouse. In this building project data from different sources is integrated and processed to form that can be easily used in reporting and analyzing.

There are two different widespread data warehouse architectures: Inmon's and Kimball's. Metropolia started first with Inmon's architecture. There were many problems and after many years we could not get much useful data to reporting. Later we changed to Kimball's architecture and got good results with it.

2. STARTING WITH INMON'S ARCHITECTURE

Building of data warehouse was started in 2006. At this moment there was no data warehousing expertise among our own employees. Development was outsourced to a consulting company and they chose Inmon's architecture. Building of data warehouse was started with creating a conceptual model of centralized data warehouse. The conceptual model was created in co-operation with consultants and our workers of different fields. It was ready in 2008. The complete conceptual model included about 120 different concepts of studying, teaching, HR, organization and finance.

One server was installed with SQL Server 2005 and later SQL Server 2008 R2. Tables were created directly from the conceptual model. After that ETL-processes were developed with SQL Server Integration Services. Those processes loaded data to the most important tables. Other tables were just left empty. Development was sometimes done passively and it had breaks.

One data mart was built on top of the centralized data warehouse. It was designed to report data from course feedback system.

3. PROBLEMS APPEARED

Conceptual model was carefully designed but it was so complicated that there were difficulties to develop ETL-processes. These difficulties appeared as erroneous data and delays in the development process. Major effort in data warehouse development focused on the centralized data warehouse and little on data marts with star schemas. Because the centralized data warehouse was normalized it made querying difficult because you needed to join many tables. It was not suitable to reporting. Also empty tables in data warehouse made data utilization more difficult. Sometimes when you started to develop a report you later realized that there was no data in one of the tables needed for that report.

At this time we also realized in practice that star schemas were needed to reporting. In Inmon's architecture two different data models are needed: the normalized data model in the centralized data warehouse and the star schema model in data marts. Also two ETL-processed are needed to get useful data to reporting: one process to load from operative databases to the centralized data warehouse and the second process to load data to data marts. Developing two data models and two ETL-processes needed a lot of work. Because of two ETL-processes there is double probability of error in one those processes. We noticed many errors in data in reports. Before you can correct

DOI: 10.7250/eunis.2013.027

erroneous data you have to track whether the error is in the first or in the second ETL-process. Tracking and correcting errors consumed large amounts of work time.

We have databases where data is already in denormalized form. So it could be easily processed directly in dimension tables in star schema. However in Inmon's architecture you have to first normalize this data, load it to the centralize data warehouse then denormalize it close to original form and load it to the data mart. This work seemed futile.

4. CHANGE TO KIMBALL'S ARCHITECTURE

In 2010 Metropolia hired own business intelligence developer. Two years data warehouse was developed in co-operation with own business intelligence developer and consultants. In 2012 it was still difficult to get useful data to reporting. We understood that data warehousing is not gainful with this approach. We changed data warehouse architecture to Kimball's.

Using Kimball's architecture we built three data marts in less than a year. Those are course feedback, studying and finance data mart. Structure of the data warehouse is presented in bus matrix in Table 1. In this table there is one row per each star schema and one column per dimension. Note X means that there is that certain dimension in the given star schema.

Data mart / fact table	**Fact**	Date	Month	Quarter	Semester	Year	Implementation	Course	Degree program	Cluster	Employee	Student	Grade	Question	Reason	Cost center	Account	Function
Studying																		
Studying accomplishment	Study credits	X	X	X	X	X	X	X	X	X	X	X	X					
Presence	True / false				X	X												
Course feedback																		
Feedback to question	Numerical assesment	X	X	X	X	X	X	X	X	X	X			X	X			
Open feedback	Textual assesment	X	X	X	X	X	X	X	X	X	X							
Finance																		
Accounting event	Amount of money		X	X	X	X										X	X	X
Budget	Amount of money		X	X	X	X										X	X	

Table 1. Bus matrix of the data warehouse

We developed reporting on top of course feedback and the studying data mart and created OLAP-cube on top of the finance data mart. Data warehouse development brought very good results at last. Change of data warehouse architecture to Kimball's was proven to be a very beneficial decision.

5. REFERENCES

Adamson, C. (2010). *Star Schema, The Complete Reference*. McGraw-Hill Osborne Media.

Kimball, R. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition)*. Wiley.

Watson, H. J., Ariyachandra, T. (2005). *Data Warehouse Architectures: Factors in the Selection Decision and the Success of the Architectures*. Retrieved May 2, 2013, from: http://www.terry.uga.edu/~hwatson/DW_Architecture_Report.pdf.

DOI: 10.7250/eunis.2013.027

6. AUTHOR'S BIOGRAPHY

Antti Tikka has completed master of sciences degree in University of Helsinki. His major was computer science and minors were theoretical physics, mathematics and statistics. In computer science he has specialized in databases and data management.

He has worked as a business intelligence developer in Metropolia University of Applied Sciences for three years. His previous job was an application developer in Medicel Company. Before that for a year he worked in KPMG as an application developer and developed SQL based reporting from customer management system. Before KPMG he worked in Sweco Industry as an application developer for three years.

Antti Tikka has eight years work experience in databases. He is a member of SQL Server Central and LinkedIn groups SQL Server User Group Finland, Microsoft Business Intelligence and SQL Server 2012 Business Intelligence. He intends to complete MCSA SQL Server 2012 certificate in summer of 2013.