International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:10, 2007

Data Transformation Services (DTS): Creating Data Mart by Consolidating Multi-Source Enterprise Operational Data

J. D. D. Daniel, K. N. Goh, and S. M. Yusop

***Abstract*—**Trends in business intelligence, e-commerce and remote access make it necessary and practical to store data in different ways on multiple systems with different operating systems. As business evolve and grow, they require efficient computerized solution to perform data update and to access data from diverse enterprise business applications. The objective of this paper is to demonstrate the capability of DTS [1] as a database solution for automatic data transfer and update in solving business problem. This DTS package is developed for the sales of variety of plants and eventually expanded into commercial supply and landscaping business. Dimension data modeling is used in DTS package to extract, transform and load data from heterogeneous database systems such as MySQL, Microsoft Access and Oracle that consolidates into a Data Mart residing in SQL Server. Hence, the data transfer from various databases is scheduled to run automatically every quarter of the year to review the efficient sales analysis. Therefore, DTS is absolutely an attractive solution for automatic data transfer and update which meeting today's business needs.

***Keywords*—**Data Transformation Services (DTS), Object Linking and Embedding Database (OLEDB), Data Mart, Online Analytical Processing (OLAP), Online Transactional Processing (OLTP).

I. Introduction

INFORMATION is priceless. However, making them useful and presentable is difficult. Many organizations currently face the challenge in trying to integrate mass amount of data to make it accessible [2]. IT professionals face sizable challenges to support those applications as they continue to grow in both their number and sophistication. Typically, much of the data needed by a particular application reside in various database systems running on different platforms. Data collected are always segregated, usually according to departments, teams [3] or even geographical area. Organization needs to migrate the data and access information from diverse enterprise business application within specified time frames [3]. With the increase in an organization size, comes the technology sophistication in consolidating those data. One solution is by using Data Transformation Services (DTS), which allows the integration of data, even from various database platforms.

Authors are with Department of Computer and Information Sciences, University of Technology Petronas, Bandar Seri Iskandar, 31750 Tronoh, Perak Darul Ridzuan, Malaysia (e-mails: {Justin_dineshdevaraj, Gohkimnee}@petronas.com.my, speed_smy@yahoo.com).

II. Related Work

Database is designed to help users examine available data from a variety of perspectives [4]. Data Transformation Services is a collection of utilities and objects that is capable of importing, exporting and converting data from any Object Linking and Embedding Database (OLEBD) compliant data source to any data source, whether to/from another database system (heterogeneous) or to/from another server [5]. Information Technology professionals can write a data movement custom program each time a need arise, however, the problem with custom program is time-consuming, waste valuable development resources and creates a large body of code that becomes increasingly difficult to maintain over time [6] is justified the use of DTS in real enterprise environment as a database solution for automatic transfer and update. A key to understanding the DTS architecture is to understand the DTS Package, which is a complete, self-contained description of all task and steps for completing an import, expert and transformation process [8]. Ideally, end-users should be able to access data from the data warehouse without knowing either where data resides from or also various usages of database applications. Data mart approach is one of the data warehouse categories [10, 11] where it extracts data from a primary data warehouse for data mart application. The extraction has a special purpose. Hence, the repository of the data ware house contains the subject of information, which is in data mart format [12].

III. Methodology

A proposed system framework has been implemented for creating data mart to consolidate multi source enterprise operational data. However, Entity Relationship Diagram, ERD will be used to model data structure for multi-source data and it depicts relationships among system entities. Hence, Dimensional Data Modeling will be used to present useful information in summarized form to identify the trends. Star schema is used to model the new data structure from ERD. Finally, the end product of system prototyping is full-featured, working model of database solution and ready for the real implementation.

IV. System Framework

Technically, Fig. 1 shows that data mart resides in between of Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP). OLTP systems represent the

source systems for a data warehouse. They are transactional in nature and are frequently purge to improve performance. On the other hand, OLAP systems typically take snapshots of production data at regularly spaced intervals and can represent historical data for months or years used for business analytical processing.

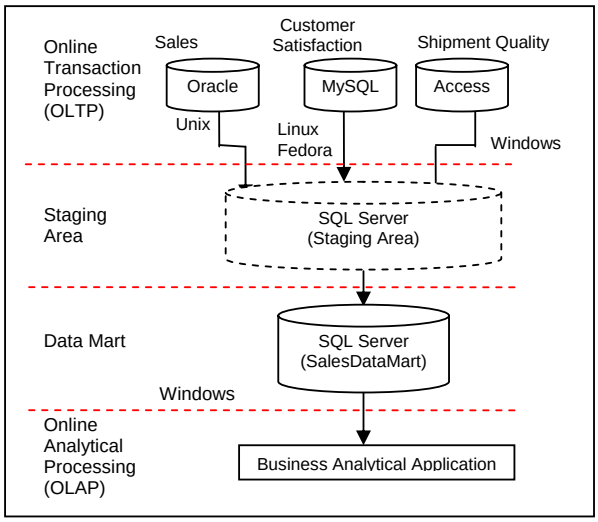

Fig. 1 System Framework for Creating Data Mart by Consolidating Multi-Source Enterprise Operational Data

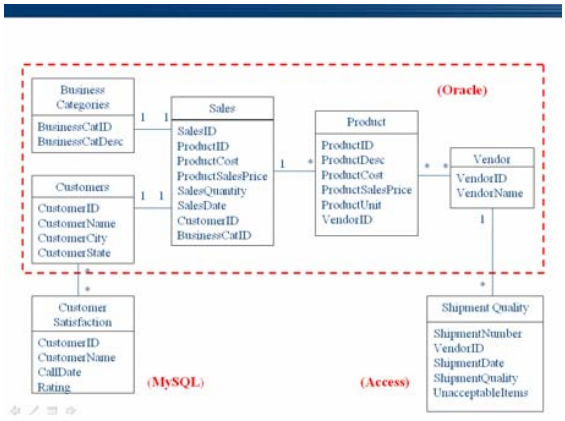

Fig. 2 ER Model

In this case, Sales data in Oracle, Customer Satisfaction in MySQL and Shipment Quality in Microsoft Access are all refer as OLTP. In OLTP, data are structured in ER model as depicted in Fig. 2. These data will be extracted and transformed into a dimensional model i.e. star schema, Fig. 3, in Data Mart through staging area for quarterly sales analysis. Business analytical application which represents the OLAP is used to analyze the business by using the data taken from the Data Mart, which can be used strategically to obtain and analyze data based on their needs [9]. Data Mart is a smaller, more manageable implementation of enterprise data warehouse. When data warehouse or data mart combined with OLAP, It consists of one or small number of fact tables and conformed dimensions. As such, it generates a SalesDataMart by consolidating heterogeneous data sources for efficient quarterly sales analysis.

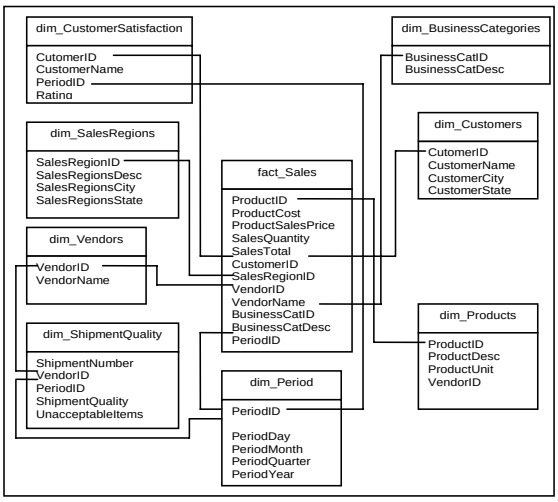

Fig. 3 Star Schema

V. Results And Discussion

Fig. 2 shows that ER model is constituted to remove redundancy and to optimize OLTP performance. In contrast, dimensional model i.e star schema is constituted to optimize decision support query for business analysis through OLAP [7]. In this case, heterogeneous business data such as sales in Oracle, customer satisfaction in MySQL and shipment quality in MS Access are extracted, transformed and loaded into SalesDataMart using DTS. The transformation is performed through staging area because it is useful in fixing source data quality problem and to avoid system degradation on OLTP systems. Extensive process of transforming the source data into appropriate format would only involves between staging area and destination. If direct transformation is performed between source and destination, OLTP system will suffer performance degradation since it has to cater extensive query for warehouse loading on top of the rigorous transactional process such as sales order, shipment and customers data entries.

In addition, SalesDataMart DTS Package, Fig. 1, has two Execute SQL Tasks running in parallel, one to insert new records and another is to modify records. Parallel tasks executions speed up the transformation process and thus improve performance of DTS package.

To illustrate the result of data transformation and update, there are a few assumptions applied. Imagine now is March 31, 2004 (end of Quarter 4). The last running date was December 31, 2003 (Quarter 3). Note that the company operates on a fiscal year as follows:

TABLE I
FISCAL YEAR OF CREATIVE PLANTS

Quarter	Month	DTS running date
Q1	Apr – June	June 30
Q2	Jul – Sept	Sept 30
Q3	Oct – Dec	Dec 31
Q4	Jan – Mar	Mar 31

At the end of Quarter 4, records that need to be transferred and updated should be between Jan 1, 2004 until 31 March 2004 (records of Quarter 4). The following describes data transformation and update for each enterprise operational data: Customer Satisfaction, Shipment Quality and Sales.

A. Customer Satisfaction

Customer Satisfaction data loading demonstrates data transformation between MySQL database system residing on Linux Fedora Server platform and SQL Server 2000 residing on Windows platform. This is possible through the Open Database Connectivity (ODBC). MySQL ODBC driver needs to be installed on the Linux box in order to perform the data transformation. Before the execution of SalesDataMart DTS Package, the data mart contains records only up to Quarter 3. At the end of Quarter 4, SalesDataMart DTS package will automatically transfer customer satisfaction records of Quarter 4 from data source (MySQL) into data mart (SQL Server), Fig. 4.

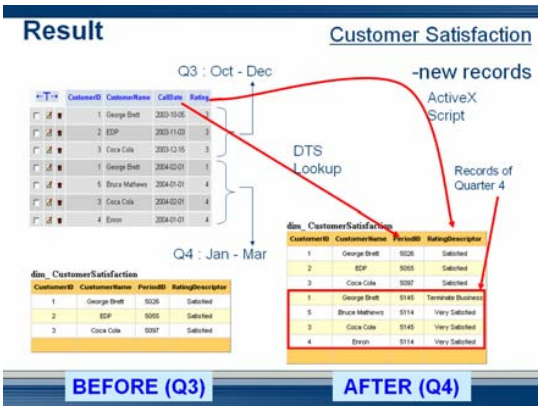

Fig. 4 Result for Customer Satisfaction

During loading process, two columns experience data transformation: Rating and CallDate. Rating value has been transformed into rating descriptor to ease sales analysis as follows:

TABLE II
RATING TRANSFORMATION VALUES

Rating Value	Descriptor
1	Never do business with us again
2	Unsatisfied
3	Satisfied
4	Very satisfied

CallDate also has been transformed into PeriodID. dim_Period represents time dimension. Time dimension is important in optimizing the sales analysis by timeframe. A stored procedure called loadtblPeriod is used to populate the dim_Period accordingly. CallDate is mapped to corresponding PeriodID via DTS Lookup. DTS Lookup allows specifying a query that executes for each value from the source that needs to be looked up in a table that resides in destination. As the data is transferred, the following query is executed for each CallDate value from source:

```
SELECT PeriodID
FROM dim_Period
WHERE PeriodDay =?
```

B. Shipment Quality

Shipment Quality data loading demonstrates data transformation between Microsoft Access and SQL Server 2000. This is done through the Microsoft Jet drivers. The result is shown in as follow:

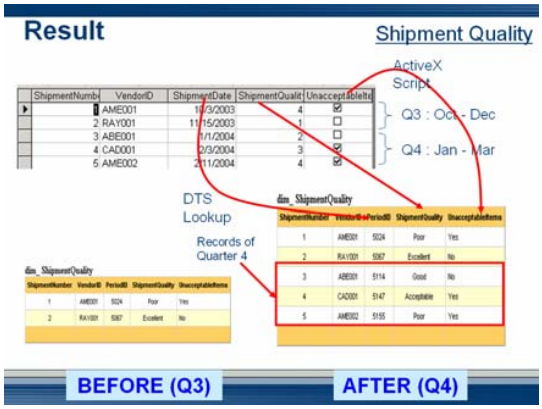

Fig. 5 Result for Shipment Quality

Referring to Fig. 5, at the end of Quarter 4, all shipment records of Quarter 4 will be automatically transferred into the data mart. Based on the figures, three columns have experienced some changes: ShipmentDate, ShipmentQuality and UnacceptableItems. ShipmentDate is transformed into PeriodID by using similar PeriodIDLookup used for Customer Satisfaction. Shipment Quality is transformed into a description-like term instead of numbers using the following ActiveX Script code:

```
Select Case DTSSource("ShipmentQuality")
  Case 1
    DTSDestination("ShipmentQuality") =  "Excellent"
  Case 2
    DTSDestination("ShipmentQuality") =  "Good"
  Case 3
    DTSDestination("ShipmentQuality") =  "Acceptable"
  Case 4
    DTSDestination("ShipmentQuality") =  "Poor"
  Case 5
    DTSDestination("ShipmentQuality") =  "Rejected"
  Case Else
    DTSDestination("ShipmentQuality") =  "BadData"
End Select
```

Unacceptableitems is transformed from checked/unchecked format in Microsoft Access into Yes/No in SQL Server using ActiveX Script as follows:

```
If  DTSSource("UnacceptableItems") = "False"  Then

 DTSDestination("UnacceptableItems") = "No"
Else
 DTSDestination("UnacceptableItems") = "Yes"
End If
```

C. Sales

Sales data loading demonstrates data transformation between Oracle database system residing on Unix Server platform and SQL Server 2000 residing on Windows platform. This is possible through either Open Database Connectivity (ODBC) or Object Linking and Embedding Database (OLEDB). SQL Server box needs to be installed with Oracle client or also known as Net8 or Net9 client in order to enable the box to access remote Oracle database.

For Sales data, not only new Sales records need to be transferred, modified records of Customers, Products, Vendors and Business Categories also should be updated. Fig. 6 shows new Product records that have been transferred from Oracle into the data mart (SQL Server).

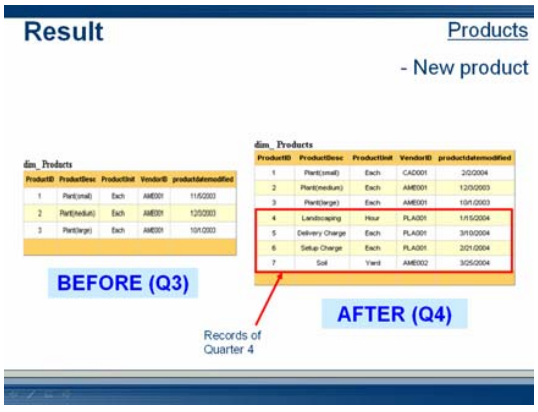

Fig. 6 New Product Record

DTS package will capture this change and update the records so that the new sales entry will follow the new pricing, Fig. 7.

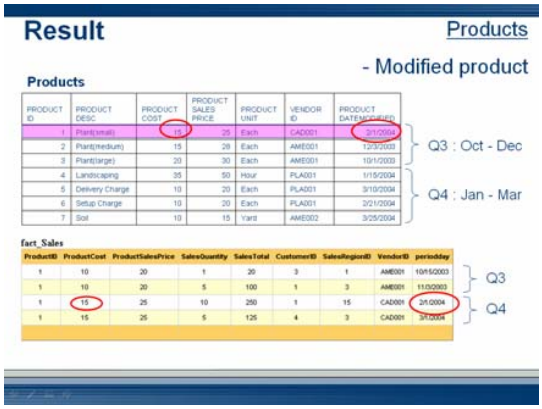

Fig. 7 Modified Product Record

Having all data of Customer Satisfaction, Shipment Quality and Sales been transferred automatically into SalesDataMart, Creative Plants now is able to perform efficient sales analysis.

VI. Conclusion

Data Transformation Services (DTS) is definitely an attractive database solution for automating data transfer and update that suite well with the current trend of business data management. Evolution of legacy system such as mainframe systems demands a solution to integrate it with current technology and business needs. It demonstrates DTS capabilities in automating data transfer between heterogeneous databases such as Microsoft Access, MySQL on Fedora Linux Server and Oracle on Sun Unix Server.

With the advantages of cost effectiveness; excellent performance; powerful and intuitive interface of DTS Designer; wide range of connectivity; bi-directional data transfer and extendable with advanced programming, DTS absolutely is a competitive Extract, Transform and Load (ETL) tool. Due to great success of Microsoft SQL Server 2000 DTS, this gives more value to Database Administrators (DBAs) in performing their job with great productivity.

References

[1] Dan Lampkin,Baya Pavliashvili, 2002, "Introduction to Data Transformation Services (DTS)", http://www.informit.com/isapi/product_id~%7BFA835B2D-C9BD-4349-96D0-A127A5045154%7D/content/index.asp

[2] Steve Hughes, Steve Miller, Jim Samuelson, Marcelino Santos, Brian Sullivan, 2002, *SQL Server DTS*. New Riders.

[3] Timothy Peterson, 2001, *Microsoft SQL Server 2000 DTS*. Sams.

[4] Gerald V. Post, David L. Anderson, 2003, *Management Information Systems*. McGraw-Hill Irwin.

[5] Mark Chaffin, Brian Knight and Todd Robinson., 2000, *Professional SQL Server 2000 DTS*. Wrox Press Ltd.

[6] Joe Guerra, November 2002, "Using Microsoft SQL Server Data Transformation Services with IBM Databases". Andrews Consulting Group.

[7] George Colliat, 1996, "OLAP, Relational and Multidimensional Database Systems", SIGMOD Record.

[8] Don Awalt and Brian Lawton, 1999, "Introduction to Data Transformation Services: DTS takes the movement of data to a new

level”,
http://www.sqlmag.com/Articles/index.cfm?ArticleID=4879&pg=1>.
[9] K.W. Chau, Ying Cao, M. Anson, Jianping Zhang, 2002, “Application of data warehouse and Decision Support System in construction management”, Elsevier.
[10] C.M. Chao, “Incremental maintenance of object-oriented data warehouses”, Information Sciences 160 (2004) 91–110.
[11] Gutie´rrez, A. Marotta, An overview of data warehouse design techniques, Reporte Te´cnico INCO-01-09. InCo-Pedeciba, Facultad deIngenierý´a, Universidad de la Republica, Montevideo, Uruguay. November, 2000.
[12] M. H. Shi, H. C. Tung, L. S. Jia, “Data warehouse enhancement: A semantic cube model approach”, Information Sciences 177 (2007) 2238–2254.