

Advanced Data Integration Solution for Enterprise Information Systems

Gholamreza Jandaghi, Ph.D.

Faculty of Management, Qom College, University of Tehran, Iran
Email: jandaghi@ut.ac.ir

Abolfazl Gandomi, Ph.D.

Faculty of Basic Sciences, Imam Hossein University, Tehran, Iran
Email: A_gandomi@yahoo.com

Hamid Haji Molla Mirzaee, Ph.D. Candidate

Faculty of Management, Qom College, University of Tehran, Iran
Email: hhmirzaee@gmail.com

Abstract

This article discus about different solutions and approaches to integrating data and information for enterprise information systems. We will explore the methods based on their applications and their advantages and disadvantages.

Keywords: data integration; enterprise; information system.

Introduction

Many enterprises expend a large amount of time and money on “information integration” jointing information from disparate data sources into a united Format. Often quoted as the most expensive issue which information technology structure countenance, actually the market intelligent firm IDC estimates that 2.5 billion USD was paid in the market for data integration in 2007 and it was 40% of their budget(Bernstein and Sergey Melnik 2007; Hass 2007; Halevy et al 2005; Smith 2007; <http://db.cis.upenn.edu>).

It is expected to increase to 3.8 billion USD in 2012 which means that the average of growth rate is 8.7 %(<http://www.slideshare.net>). In this article we try to explain about data integration and clear up any perplexity by; Describing types of data integration solutions.

This article includes six sections which are included; Portal Management, Object to Relational Mappers, Document Management, Virtual Data Integration, Data Warehouse Loading, Message Mapping,

Actually by using the variety of architectural approaches can solve the problems. In this part we explain the summary of these approaches, along with the common types of products utilized. The approaches of data integration are as follow:

1. Portal Management

One of the approaches for integrating data and information is to present all of information on the same screen and in one page. A portal is a complete web site that is created with this kind of integration. For example, the home page of healthcare services web site usually displays patient information , patient operation, medicine consumption and the analysis of recent activities about the medical activities and when the user see them together this in the same screen this become the actual integration of data. To achieve portal management it needs a mixture of content management because the transaction with document and database is necessary and it requires user interaction technology for presenting data and information in helpful intelligent ways. sometimes the product for designing portal is included these technologies together(Firestone 2003). however frequently they are applied one by one according to functionality which is required for designing portal based on the developers and programmers experience who create it.

2. Object-to-Relational Mappers

Nowadays object oriented language is used usually for writing the application programs however data and information typically is stored in the relational database.

While mapping of the applications with databases needs relational integration of the schemas the differences in schema constructs be able to make the mapping rather complicated. For instant we have many ways for mapping the classes which related via inheritance into relational tables so for simplifying the issues an object to relational mapper proposes a high level language to classify mapping(Melnik et al 2007).

And finally the result of mapping will be compiled into programs which translate queries and will update the object oriented interface into queries and updates on the relational database.

3. Document Management

In many huge companies or enterprise a lot of information is included documents, for example text files, spreadsheets and slide shows and etc which include interrelated information pertinent to marketing plans, pricing, and development schedules and etc. To support collaboration and shun duplication in an enterprise definitely all information must be integrated.

In this case integrating data and information might plainly involve making the documents accessible on a single Web page or in a content management system.

Or integration might mean joining information and data from some documents into a new document, for example financial analysis.

Documents can be indexed to facilitate keyword search through the enterprise.

In some applications it is helpful to extract structured data from documents, for example in the healthcare patient name and patient address from email messages received by the patient support team. The ability to extract structured data might also permit healthcare information to integrate unstructured documents with pre-existing structured data. For example the healthcare organization wanted to extract some polls information or any unstructured data related medical report with emails in order to apply integration and enable better analysis of problem reports.

5. Virtual Data Integration

Even though the virtual data integration method integrates data sources but it doesn't materialize the integrated view for them and provides a mediated schema against queries made by the users. In contrast the warehouses materialize the integrated data.

This execution, frequently named a query mediator (Wiederhold 1992) or enterprise information integration (EII) system. Morgenthal et al (2005) translates the user's query into queries on the data sources and integrates the result of those queries so that it appears to have come from a single integrated database. EII is still new technology and now it is not common technology compare with data warehousing.

Even though the databases cover associated subject issue, they are heterogeneous

In that they might utilize diverse database systems and structure the information or data using different schemas.

An EII system may be applied, for instant, by a financial company to organize for each buyer a statement of portfolio positions which consolidates data about his holdings from the local customer database with prices of store that retrieved from an external source.

To manage this heterogeneity of EII, a designer generates a mediated schema which covers the required subject issue in the data sources and maps the data source schemas to the new mediated schema. Data cleansing and reforming problems also become visible in the EII context. But the solutions are diverse in EII because data should be converted as element of query processing rather than by the periodic set process related with loading a data warehouse. Products of EII are different and according to the types of data sources to be integrated the products of EII become different. As instant some products concentrate to integrate Web services, and some to integrate SQL databases, some on to integrate bioinformatics databases.

4. Data Warehouse Loading

A data warehouse is a database which combines data from several data sources(Chaudhuri & Dayal 1997).

For instance, it might consolidate sales information from subsidiaries to provide a sales picture for the whole company. Because subsidiaries have overlap sets of customers and might have inconsistent information about the customer, data must be cleansed to settle such differences. Furthermore, each subsidiary might have a database schema and diagram (that is, data representation) which is different from the warehouse schema.

Consequently every subsidiary's data has to be reform into the ordinary warehouse schema. this problem(Kimball & Caserta 2004) deal with Extract Transform Load (ETL) tools, actually ETL tools solve the problem by simplifying the programming of scripts.

Usually an ETL tool contains a list of cleansing operations and reforming operations for example SQL (Structured Query Language) which has functions to join, sort and select data.

Also this tool might contain some functions such as scheduling to manage the refreshing or periodic loading of the data warehouse.

Some of the ETL tools have ability to customize for master data management which is, to create a data warehouse that holds the master copy of critical enterprise reference information, for example information about customers or products. Master data is first integrated from multiple sources and then itself becomes the definitive source of that data for the enterprise.

Sometimes Master data management tools include domain specific functionality. For example, for vendor or customer data, they might have format for name and address standardization and cleansing functions for correcting and validating postal codes.

5. Message Mapping

Basically the message mapping is the Message oriented middleware that integrate the independent and disparate applications via sending messages among them.

When messages bypass and move through a broker this creation is typically named by an enterprise application integration system or EAI system(Alonso et al 2004).

In case of avoiding of broker and applying same protocol for all applications same as web service the product is named enterprise service bus. And the product is named as workflow system when each application going to invoke as element of whole service.

Moreover these products provide the flow control and protocol translation services.

Conclusion

In fact we have many approaches and solution that has multiple basic technologies that same between all solutions but this article described some of them and basically each method and solution has advantages and disadvantages.

References

- Alon Y. Halevy A. Y. , Ashish N. ,Bitton D. ,Carey M., Draper D., Pollock J., Rosenthal A. and Alonso G., Casati F., Kuno H. A. and Machiraju V.(2004) Web Services.
- Sikka V.(2005). Enterprise Information Integration: Successes, Challenges, and Controversies.
- Bernstein P.A. and Melnik S.(2007) Model management 2.0: Manipulating Richer Mappings.
- Chaudhuri S. and Dayal U.(1997) An overview of data warehousing and OLAP technology.
- Firestone J. M.(2003), Enterprise Information Portals and Knowledge Management.
- Hass L.(2007) Beauty and theBbeast the Theory and Practice of Information Integration.
<http://db.cis.upenn.edu/iiworkshop/postworkshop/index.htm> .
- <http://www.slideshare.net/divjeev/forrester-wave-business-intelligence-platforms-8427563>
- Kimball R. and Caserta J.(2004) The Data Warehouse ETL Toolkit.
- Melnik S. , Adya A. and Bernstein P. A.(2007) Compiling Mappings to Bridge Applications and Databases.
- Morgenthal J. P.(2005) Enterprise Information Integration.Smith M.(2007), Toward Enterprise Information Integration. www.softwaremag.com.
- Wiederhold G.(1992) Mediators in the Architecture of Future Information Systems, Stanford Univ. Press.