维普资讯 http://www.cqvip.com

企业异构数据流集成技术探讨*)

于宝琴[12] 王太勇[1]

(天津大学机械学院 天津 300072)[1] (天津财经大学信息系 天津 300222)[2]

摘 要 本文以某生产制造型企业信息化工作为背景,探讨如何根据企业现状进行异构数据源的集成,并以Mysql和Oracle 9i关系数据库为实例,介绍了低成本的数据移植方法,指出了建立企业级数据仓库是企业长远发展的目标,只有在此基础上,建立生产、工艺和制造部门级数据集市才能更好地为PDM用户服务。最后给出了基于Web的企业数据流模型,以使企业内部、外部数据流得到共享与利用。

关键词 异构数据流,数据移植,数据仓库,Oracle数据库,PDM系统

Discussion of Technology of Heterogeneous Data Flows for Enterprises

YU Bao-Qin[12] WANG Tai-Yong[1]

(School of Mechanical Engineering, Tianjin University, Tianjin 300072)[1]

(Department of Information, Tianjin University of Finance and Economics, Tianjin 300222)[2]

Abstract Based on the application of the information system of a manufacturing enterprise, under the present conditions of the enterprise, this thesis discusses an effective method, a low cost approach of data transplant, about integrating process of heterogeneous data sources. The method is illustrated by the example of the data transplant between Mysql database and Oracle 9i database. It demonstrates the fact that establishment of enterprise data warehouse serves as the foundation of a long term objective for the enterprise. And the set up of enterprise data warehouse can further improve the services of PDM users from data marts of the various department of production, technics and manufacturing. Finally an overall framework of Web-based data flow of an enterprise is developed so that enterprises will be able to share and make full use of both interior and exterior data flows.

Keywords Heterogeneous data flows, Data transplant, Data warehouse, Oracle database, PDM system

1 引言

信息作为现代企业的宝贵资源,占据着越来越重要的地位,已经成为现代企业科学管理的基础、正确决策的前提、有效调控的手段。从供应商管理信息、原材料采购信息、制造工艺过程,分销渠道选择过程,客户关系的管理信息,以及企业流程中物流和信息流整合,每一环节的绩效都将影响整体供应链的绩效。只有对企业信息流实施有效管理和控制,才能在企业竞争中立于不败之地。

基于Web的企业信息管理的最大优势是企业借助于现代信息网络技术,使信息流的流动变得更为通畅。此时的信息流处于一个极为重要的地位,它贯穿产品生产、销售等过程中,在一个更高的位置对产品的整个过程进行控制。在这样的环境下,利用先进的通讯网络技术,建立起通畅的企业信息网络,包括企业内部的信息网和企业外部的信息网,从而大大加快了企业信息流的流动速度,增加了信息的共享程度。

2 企业信息流

Web与数据仓库技术是企业信息化过程中的两项关键技术。Web技术使信息能最大程度地实现共享,为实时访问数据提供了解决方法。而企业数据仓库建立的优劣,直接影响着企业信息系统的运转。只有将Web与数据仓库技术相结合,才能使信息处理与共享变得更加经济和有效[1]。

2.1 PDM中的信息流

企业数据仓库建立的前提是企业异构数据源的集成,它是企业信息化的重要基础。尤其在生产制造型企业中,CAD、CAPP、CAM都是"孤岛式"的单项技术,所产生的数据涉及产品的不同方面,因此产品数据管理(Product Data Management,PDM)技术,建立在网络平台和数据库技术基础之上,将计算机在产品设计、分析、制造、工艺设计等方面应用产生的信息孤岛集成在一起,对产品整个生命周期内的数据进行统一管理,解决了这些技术深化应用的瓶颈问题,真正实现了这些系统的信息集成。另外ERP(Enterprise Resource Planning企业资源规划)作为CIMS的重要组成部分,通过PDM系统可以同CAD、CAPP等系统及时地进行双向信息传递,以使企业全局信息流有效集成与共享[2]。

*)基金项目:国家自然科学基金项目(50175081)。于宝琴 博士生,副教授,主要从事计算机应用、数据仓库、数据挖掘、CIMS等方面的研究。王太勇 教授,博士生导师,主要从事企业信息化、虚拟技术、动态监测与诊断、CIMS等方面的研究。

维普资讯 http://www.cqvip.com

2.2 存在问题

在实施某企业 PDM 系统的过程中，我们遇到了以下问题：

- 对于大量的数据不能提供统一的数据接口；
- 不能采用通用的标准和规范（如使用不同的编码体系），共享通用的数据源；
- 随着业务的增加，管理人员的工作越来越多、操作越来越复杂，用户分散，相互联系程度低，信息相对封闭，共享程度低；
- 大量日趋复杂的办公业务依然或多或少地靠手工处理进行流转；
- 信息加工、处理手段差，影响信息质量，无法直接从各级各类业务信息系统采集数据并加以综合利用；
- 无法对外部信息进行及时、准确地采集、利用，业务系统产生的大量数据无法提炼升华为信息，及时提供给决策部门；
- 已有的业务信息系统平台及开发工具互不兼容，无法在大范围内应用。

2.3 实施策略

根据以上问题，制定出相应的实施方案，并将该企业的信息化过程划分多期进行，上下期之间相互衔接。具体来说，整个系统按照分层（基础技术层、核心业务层、应用层）分模块的策略，基础技术层（基础数据管理模块、数据迁徙模块、系统管理模块等）构造环境基础结构，使用户与操作系统、数据库、网络隔离开来；核心业务层（产品配置管理、零件分类管理等）管理产品业务活动信息；应用层（项目管理、库存管理、销售管理等）在基础技术和核心业务之上，面向特定问题的方案及实现。

本文就企业基础技术层中数据整合问题进行了实践总结，将 mysql 数据库、sql server 数据库、dbase 数据库、excel 电子表格、text 文本文件等原有数据信息成功地集成到 Oracle 数据仓库中，从而实现企业信息流的统一管理，为后期工作奠定了基础。

3 异构数据流集成

数据源的异构问题包括模式异构、数据异构、语义异构等[3]。针对该企业的模式异构和数据异构问题可以采用数据仓库技术和数据抽取工具来实现，但由于建立数据仓库系统投资费用大、项目实施周期长、项目成功率高风险等原因，制约了其在中、小型企业或数据积累少的企业解决异构数据源整合和集成需求的应用；有些数据库管理系统自带的数据转换工具，能低成本地解决异构数据源整合和集成问题。

3.1 低成本的数据迁徙

基础数据的整合是企业信息流的重要成分，其中数据整合过程又离不开数据迁徙工作。数据迁徙主要涉及数据或数据库系统的升级、降级、重组和移植四个方面[4]。数据的升、降级工作主要指数据库新旧版本之间的数据迁徙；而数据的重组则是为了更好地提高性能、有效地管理空间；与前面的工作相比，数据移植具有物理的特征，它是指数据升降级和重组期间的数据传输。通常情况下，数据都是从小系统移植到一个更强大的系统中。

3.1.1 *理论分析* 从理论上讲，数据移植的全过程应被分解成不同的阶段，分别是：

- 移植策略阶段，主要制定目标计划、策略文档等。
- 移植预分析阶段，主要完成移植与测试计划。
- 移植分析阶段，确定要映射的字段，并找出对应字段移植的算法。
- 移植设计阶段，数据映射，程序编制。
- 移植测试阶段，在测试环境调试转换代码。
- 移植修正阶段，修改更正测试结果。
- 移植实现与维护，生成数据并载入新系统，维护数据。

3.1.2 *实践结果* 在实际中，利用现有的低成本的数据转换工具，如：PowerBuilder 的 DataPipeline、SqlServer 的 DTS、Oracle 的 SQL*Loder 等工具可以实现各种异构数据库系统和文本、电子表格等文件系统格式的数据整合和集成。针对具体的系统编写具体的数据转换代码，来一起完成从原始数据采集、错误数据清理、异构数据整和、数据结构转换、数据转储和数据定期刷新的全部过程。图 1 所示为 PDM 系统中的 Mysql 数据库 DATA 转换成 Oracle 数据库 DLDZ 的工作流程图。

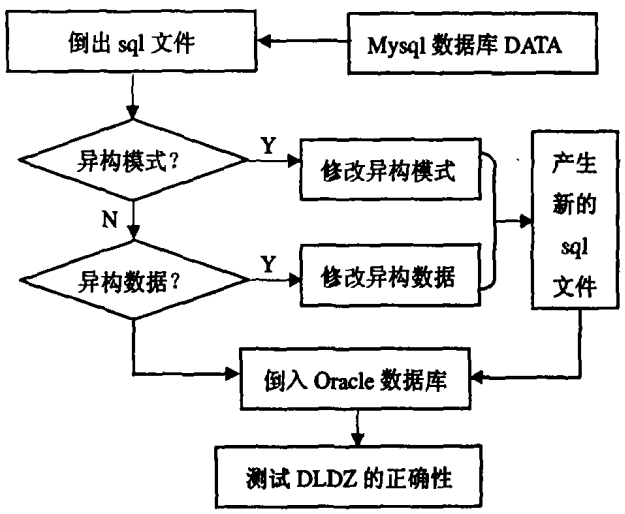

图 1 数据移植的工作流程

维普资讯 http://www.cqvip.com

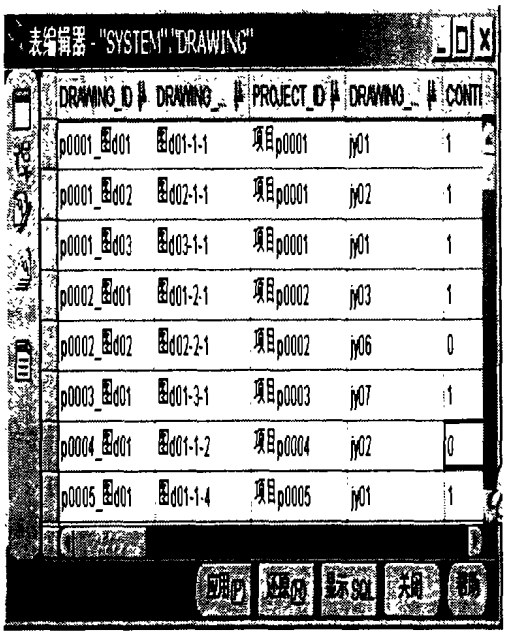

图 2　倒入 Oracle 的数据表

数据倒入 Oracle 数据库之后，用 PowerScript 编写的部分测试数据库连接的程序代码如下：

```
//Profile oracleodbc
SQLCA.DBMS="ODBC"
SQLCA.AutoCommit=False
SQLCA.DBParm="ConnectString='DSN=DLDZ;UID=
    system;PWD=manager"
Connect;
If SQLCA.SQLCODE()0 then
    Messagebox ( " Don'tconnect database ", SQL-
    CA.SQLERRTEXT)
  Return
End if
//dw_1:datawindow,show SQLCA result
Dw_1.setTransObject(SQLCA)
Dw_1.retrieve()
```

实践结果如图3所示，一批数据已由原来 Mysql 数据库成功倒入 Oracle 数据仓库中。

3.2　企业数据仓库--异构数据集成的理想模式

著名的信息系统专家威廉·德雷尔(William Durell)指出：没有卓有成效的数据管理，就没有成功高效的信息处理，更建立不起来整个企业的计算机系统。这种说法同样适用于企业数据仓库的建立。从长远来看，数据仓库应是企业异构数据整合的理想模式[5]。

通常企业数据流涉及业务数据层、数据整合层、信息服务层三个层次。但针对该制造企业需要，系统被划分为4层，分别是业务数据层、数据存储与管理层、信息服务层、应用层。

·业务数据源是数据仓库系统的基础，是整个系统的数据源泉。

·数据的存储与管理：是整个数据仓库系统的核心。按照数据的覆盖范围，数据仓库存储可以分为企业级和部门级(通常称为"数据集市"，Data Mart)。

·OLAP 服务器与 PDM 服务器：均对需要的数据按照多维数据模型[6]进行再次重组。PDM 服务器将与产品相关的结构信息、过程信息、工艺流程信息、零部件信息从部门级数据集市中析取出来，以便 PDM 用户频繁地使用这些信息，提高 PDM 系统的数据传输能力[7]。

·企业各项应用主要包括各种数据分析、报表与查询、数据挖掘和决策支持、PDM 用户的可视化设计与产品开发等，以及各种基于数据仓库或数据集市开发的应用。

4　基于 Web 的企业数据流

现代企业处在一个非常复杂的环境中，企业试图跨越多个部门或跨越一条供应链中的多个组织部署任务，同样，它也允许来自所有的部门、区域和供应链中的组织参与到产品生命周期中的设计、开发和加工过程之中。Web 技术将这种需要变为现实，它不仅打破部门之间空间和时间形成的边界，而且将地理位置分离的组织无缝地连接起来。

采用上述基于 Web 的数据仓库结构来组织管理企业及企业间的数据流，除具有一般 B/S 体系结构共有的特点之外，在 Web 层的设计过程中，以 Servlet 实现控制逻辑，通过 JavaBean 组件实现业务逻辑访问数据库，用 JSP 进行页面展示。数据层采用 Oracle 数据仓库[8]，使 Web 层的 JavaBean 组件通过 JDBC 连接数据库，完成数据访问和存储逻辑。为了提高并行访问数据仓库的性能，本系统设计了数据库连接池 ConnectionPool，预先初始化多个连接提供给其它组件使用，并负责管理所有的数据库连接工作。此外，根据 PDM 用户的需要，本系统实现了将生产、工艺和制造部门的数据源提取到 PDM 服务器中，使 PDM 系统应用于产品开发的整个生命周期中，每个应用可以在不同的设备上运行。例如 CAPP 不再是一个孤立的工艺设计模块，而是产品开发环境下的有机组成部分。它不仅与 CAD 和 ERP 集成在一起，同时与其上层 PDM 服务器从信息到功能上都紧密连接在一起，如图3所示。

总之，建立企业数据仓库是一个大型系统工程，而在此基础上建立生产、工艺和制造部门级数据集市又是重中之重。制造型企业各自特点不同，必须立足全局，在系统实施之前建立一个十全十美的解决方案是不现实的。另外，基于 Web 的 PDM 理论和技术目前还在不断地发展中，有关其应用系统结构、模型、方法和实现的讨论也在不断深入，所以本文提出的内容也会在新的需求和新的研究成果中不断更新和完善。

维普资讯 http://www.cqvip.com

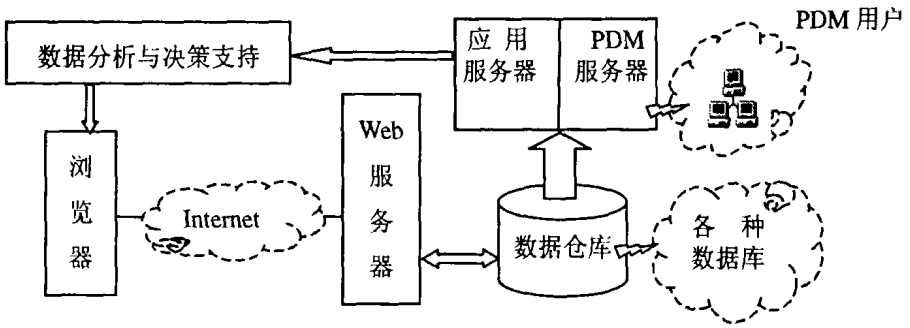

图3 基于Web的企业数据流模型

结束语 本文提出的将企业异构数据流从低成本的数据移植逐步建设成为综合的信息仓库，使得企业能够让多种业务应用系统、多种异构数据源并存，实现了数据动态展现、互访和综合利用，这样既保护了企业的原有信息化投资，又提供了应用系统由旧向新、系统平台由低向高平滑过渡的需要，同时基于Web的企业数据流挖掘不仅能满足阶段性、可扩展性信息系统建设的需要，又能实现所需信息服务目标，还能进一步加强企业PDM系统之间数据访问和产品应用的能力。数据仓库建设没有成熟的模式可循，具体实施时，应根据企业实际状况，不断探索和不断改进。

参考文献

1 陈义，宋执环，李平. 基于Web的流程企业数据仓库体系研究[J]. 计算机集成制造系统，2003，9(6)：493
2 北大高科网站. PDM与企业信息集成[Z]. 中国计算机报. http://www.pku-ht.com. 2004.3
3 熊海灵，伍胜，余建桥. 异构数据源的集成与访问[J]. 计算机科学，2003，30(5)：183
4 林行健. 精通oracle9i[M]. 机械工业出版社，2004.1
5 甘利人. 企业信息化建设与管理[M]. 北京大学出版社，2001.168～170
6 (加)Han Jiawei，Kamber M. Data Mining Concepts and Techniques[M]. 机械工业出版社，2001.44～46
7 朱全敏，蒲工. 基于Web的产品数据管理系统综述[Z]. http://www.e-works.net.cn/ewkArticles/Category33/Article14594.htm. 2003.6
8 oracle公司·oracle express server[Z]. http://www.oracle.com. 2004.4

(上接第8页)

在图4与图5中，横坐标表示相邻时间区段，纵坐标表示相异度，其中的小三角表示在该相邻时间区段间的相异度，图中的平滑曲线表示这些小三角的总体趋势。图4是从29个查询中随机抽取个而给出的相异度曲线及其趋势。为了反映整个系统的情况，我们分别计算这些相邻时间区段间这29个查询相异度的平均值，得到图5中的曲线。从图5中，我们可以看到相异度的总体趋势是渐近于某一值，但不是零。相异度的逐渐减小的趋势说明了FASE不断调整自己的反馈信息库，使得自己更能迎合大多数用户的兴趣。而相异度不会降为零正是FASE搜索引擎不断调整自己的动力所在。在面对众多用户的时候，FASE力求使自己的服务更好地适应大多数用户的兴趣。

结束语 本文根据带反馈的自动控制系统的基本原理，提出了自适应搜索引擎的概念。自适应搜索引擎首先通过一种机制来采集用户对搜索结果的访问序列，在此基础上生成自适应搜索引擎系统的反馈信号，在反馈信号生成过程中，我们采用了粗糙集理论来解决反馈信息库的空间限制和用户兴趣与反馈信息库中的兴趣的模糊匹配问题；然后通过一定的方式将反馈信息综合到搜索结果中去，从而使得系统具有自适应的能力。将来可以进一步探索反馈信号的生成方法和反馈信号与搜索结果集成的方法，以提高搜索引擎对用户的适应能力。

参考文献

1 Levy M R. Web Programming in Guide. Software，1998，28(15)：1581～1604
2 Genesereth M R，Ketch S P. Software Agents. Communication of the ACM，1994，37(7)：48～53
3 Bamshad M，Cooley R，Srivastava J. Data Preparation for Mining World Wide Web Browsing Patterns. Journal of Knowledge and Information Systems，1999，1(1)
4 Budi Y，Dik L. A world wide Web resource database system. IEEE Transactions on Knowledge and Data Engineering，1996，8(4)：548～554
5 张卫丰，徐宝文. Web搜索引擎框架研究. 计算机研究与发展，2000，37(3)
6 邹涛，等. WWW上的信息挖掘技术及实现. 计算机研究与发展，1999，36(8)：1021～1024
7 张卫丰，徐宝文，周晓宇. Web页面中的计数器研究. 小型微型计算机，2000，21(10)：1096～1099
8 张卫丰，徐宝文，周晓宇. Web页面中元素间交互技术研究. 计算机工程，2000，26(8)：62～64
9 Bamshad M，Cooley R，Srivastava J. Creating Adaptive Web Sites Through Usage-Based Clustering of URLs. In：Proc. of the 1999 IEEE Knowledge and Data Engineering Exchange Workshop (KDEX'99)，Nov. 1999
10 Weiyi M. Detection of Heterogeneities in a Multiple Text Database Environment. http://panda.cs.binghamton.edu/~meng/pub.d/coopis99.ps.gz
11 张卫丰，徐宝文，许蕾. Web页面安全性技术初探. 计算机工程与应用，2000，36(11)：158～161
12 张卫丰，徐宝文，许蕾. 利用Agent个性化搜索结果. 小型微型计算机[已录用]
13 Zhang Weifeng，Xu Baowen，Yang Hongji，Chu W C. A Genetic Algorithm Based General Search Engine. In：Proc. of IEEE MSE'2000
14 Xu Baowen，Zhang Weifeng，Chu W C，Yang Hongji. Application of Data Mining in Web Pre-Fetching. In：Proc. of IEEE MSE'2000
15 Zhang Weifeng，Xu Baowen，Chu W C，Yang Hongji. Data Mining Algorithms for Web Pre-Fetching. In：Proc. of WebSem'2000