

基于XML的数据迁移技术研究

曹大华

(胜利油田临盘采油厂 信息中心,山东 临邑 251507)

摘要:针对临盘采油厂的数据迁移的具体需求,该文提出了一种基于XML的数据迁移模型,以实现数据的提取、转换、写入等功能。它采用JAVA和XML技术,具有很好的跨平台、易扩展、复用性等,能很好的解决异构数据库间数据迁移的问题。

关键词:数据迁移;XML;异构数据库

中图分类号:TP311 **文献标识码:**A **文章编号:**1009-3044(2013)01-0011-04

The Research of Data Migration Technology Based on XML

CAO Da-hua

(The Information Center of Linpan Oil Production Plant of ShengLi Oilfield, Linyi 251507, China)

Abstract: Facing with the statue of data migration of LinPan Oil Production Plant, this paper presents a model of data migration based on XML technology. It includes data extraction, transformation, and data writing functions, and so on. It uses JAVA and XML technology, with a good cross-platform, easy to expand, reusability, etc., It can be a good solution to solve the problem of data migration between heterogeneous databases.

Key words: data migration; XML; heterogeneous database

临盘采油厂综合信息平台始建于2001年,积累了很多重要的油田数据。随着“数字油田”的发展趋势,临盘采油厂的数据需要迁移到一个统一的源头数据库,这样不仅临盘采油厂可以访问数据库中的信息,其他系统也可以共享其数据,以达到信息统一管理、信息共享的目的。这样能很好的解决了“信息孤岛”的现象。针对上述问题,结合现有的数据迁移转换技术,设计一个数据迁移转换接口是十分必要的。它将实现将来源离散且格式多样的临盘采油厂数据进行处理后,再将这些标准化数据迁移到源头数据库中,为临盘采油厂和其他管理部门提供标准统一的数据信息,实现数据的共享。

1 数据迁移相关技术及解决思路

1.1 常用的数据迁移技术

数据迁移是指将源数据库中指定表的数据进行清洗、转换并装载到目的数据库指定表中的过程^[1]。数据迁移可以分为抽取、转换、装载三个步骤。数据抽取是根据实际应用,对旧系统的底层数据进行筛选,选择出需要的待迁移数据。转换是根据新旧数据库的映射关系进行的,首先比较新旧系统中字段数据类型的差异和联系,找出适当的对应关系,然后进行相应的转换。数据装载是通过数据迁移工具或事先编写的SQL程序将抽取、转换后的数据写入到目标数据库对应的表中。数据迁移的基本流程如图1所示。

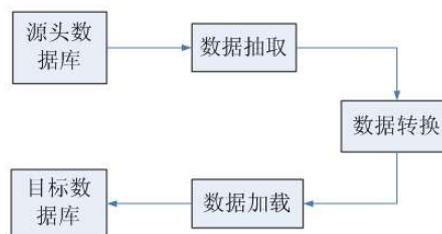


图1 数据迁移基本流程图

常用的几种数据迁移的实现方法有以下几种^[2]:

1.1.1 数据库厂商的专业工具

当前很多数据库厂家都提供专门的数据迁移工具,常见的有Informix的Infomover、Microsoft SQL Server的DTS和Oracle的Oracle Warehouse Builder等。国内的主流数据库厂商如DM、King Base、GBase等也都有着自己的数据库迁移工具。

收稿日期:2012-11-17

为了很好解决上述问题,该文采用了基于xml的中间件,数据迁移的总体流程图如图2。

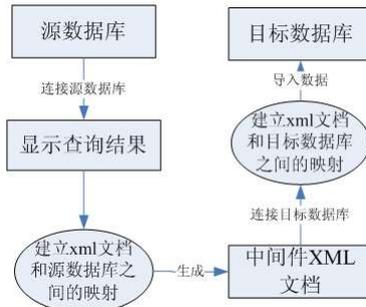


图2 数据迁移总体流程图

用户是在了解源数据库基本结构的情况下,进行数据库中数据表的选择,可以把整个表结构通过Java映射为XML文件,也可以选择其中部分字段进行映射,或者是一些字段组合成一个新的字段。

通过相应的XML格式检测,如果XML文件格式合法,则根据XML文件中的结构进行其与目标数据库中表的结构进行对应。XML中的标签名可以不同于目标数据库中表的列名,这样就解决了源数据库和目标数据库表列名不同的问题。

最后,用Java采用DOM、JDOM或者SAX的解析方式,把XML文件中的数据读取到目标数据库。

2.2 源数据库到XML文件的数据转化

在程序开发过程中,需要连接源数据库和目标数据库,连接数据库采用JAVA语言的JDBC驱动程序。该驱动程序具有效率高的特点,不会增加额外的开销,由纯JAVA语言开发而成,拥有最佳的兼容性。

在源数据库和XML文档之间进行转换和高效的传递数据,必须在XML文件和数据库表结构之间建立映射^[5],建立它们之间的映射有两种常用的映射方法:基于模板驱动的映射和基于模型驱动的映射^[5]。

本文采用了基于模板驱动的映射来实现数据库到XML文档的数据转换。在模板中加入了SQL命令,经过中间件的处理生成相应的XML文件。

流程如下图3。

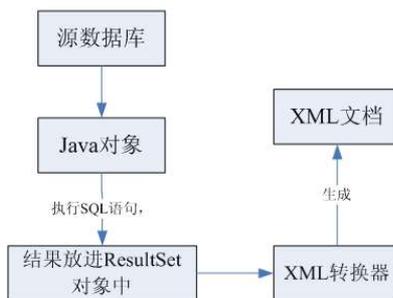


图3 源数据库到XML文档转换

基于模板驱动的映射操作比较灵活,但是它只支持关系型数据库转换到XML文档。

2.3 XML文档数据转移到目标数据库

XML文档中的数据转换到目标数据库中,我们准备采用基于模型驱动的映射模式,利用生成的XML文档中的数据结构显性或隐性的与目标数据库表结构进行映射^[6]。在此过程中采用DOM、JDOM或者SAX接口进行XML文件的解析。

其流程图如图4。

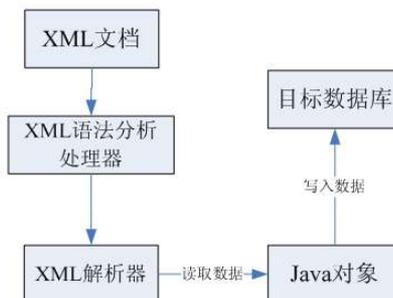


图4 XML文档数据到目标数据库

实现XML文档数据到目标数据库中的转换步骤如下:

- 1)首先通过XML语法分析处理器检测XML文档结构。

