

XML 数据仓库技术在电信网络管理系统中的应用

Telecom Network Management System Research Based On XML-based Data Warehouse Technology

陈 康 贾巍巍 赵远东 (江苏南京信息工程大学信息工程系 210044)

摘要:本文描述了当前电信网络管理系统的现状及存在问题,并讨论了 XML 数据仓库的体系结构及相关问题,提出了创建以 XML 数据仓库技术为基础的综合网管系统,较好解决异构数据源之间的交换问题,它的建立对网管现状的改善有积极意义。

关键词:数据仓库 综合网管 XML 联机分析处理 数据挖掘

1 引言

XML 即可扩展标志语言 (Extensible Markup Language),是 W3C 设计的一组规范,是 SGML 的一个子集,是针对 Html 和 internet 设计的标准的、可扩展的、通用的数据格式^[2],它通过文档类型 (DTD) 定义来定义某个领域的 XML 词汇和准则,然后基于 DTD 或 XML Schema 来开发文档,可将文档的内容、展现和操纵分离开来,为不同系统间数据交换提供了客观实现条件。XML 可以自己定义标签,描

细结构,对于数据仓库系统来说,存在多个异构数据源,采用上述方法访问时,必须要熟悉相关数据源,无疑会严重限制数据仓库的实用性、灵活性。由此,考虑到 XML 技术的优点,提出了以上基于 XML 的数据仓库的系统,采用了图 1 的 XML 格式转换器技术,即首先在每个数据源上放置一个查询转换器(图中略),将异构数据源中的数据进行封装变为 XML 格式文件,当 XML 格式文件进入 XML 格式转换器时,转换器对文件内容进行分析判断,并根据 XSLT 样式表

进行名称、格式的转换成为具有统一格式的 XML 文件,最后对格式转换结果进行整理,为不同的应用生成 XML 文档的不同表现形式,从而实现异构数据源模式转换为统一数据源模式,大大简化数据 ETL 等操作。

以上也是一种瘦客户方案(b/s 结构),客户端只需安装支持 XML 的 WEB 浏览器即可,如 IE/NS 等。WEB 服务器用于管理客户机与应用服务器间的信息流,起着桥梁作用,可以用各种流行服务器,如 IIS, APACHE 等,WEB 服务器接

受客户端 HTTP 请求,进行合法性检查、身份验证等后,向应用服务器发出请求,WEB 服务器上可装有 cgi、jsp、asp、nsapi、isapi 等应用程序,这样可与应用服务器交换信息,并将结果以 XML 的形式返回给客户端。应用服务器不仅安装由应用程序软件包,如 OLAP 服务器程序、OLAP 工具、数据挖掘(DM)工具以及装有构造数据仓库的数据抽取、转换和集成(ETL)工具等,而且还有 XML 解析器、文档存取接口 DOM 或 SAX、XSL 和 XSLT 处理器等,当收到来自 web 服务器上的 XML 查询时,应用服务器将查询送至 XML 解析器,并利用 DOM 接口来访问存在 XML 文档中的信息,送至相应处理程序,处理程序对数据仓库进行 SQL 操作,将结果

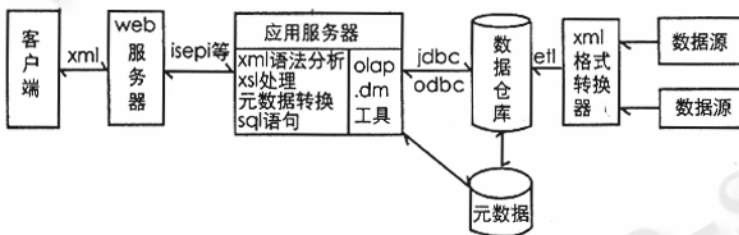


图 1 基于 XML 的数据仓库的体系结构

述不同数据库的数据,相对于 DCOM(依赖于 windows 平台)和 CORBA 等分布式计算技术,它是解决异构数据库之间的数据交换的一种较好的理想方案。将 XML 应用于数据仓库容易实现数据在 web 上的发布,有利于数据集成,解决异构数据源之间的兼容问题,从而大大的简化了数据的处理及查询。

2 基于 XML 的数据仓库的体系结构

基于 XML 的数据仓库系统的体系结构如图 1。

在 b/s 结构中访问数据源时,一般通过 odbc/jdbc 接口使用 SQL 语言来查询访问,这样必须要知道数据库表的详

交由 XSLT 处理器进行封装,并使用 XSL 将结果转换成所需格式的 XML 文档,返回给 web 服务器。

3 综合电信网络管理系统的主要功能

模块及其体系结构

综合网管系统是建立在各个现有专业网管基础上的,它由两个功能模块组成:

(1) 综合管理模块。它是针对各个子专业网管而言的,主要提供资源查询实时处理服务,如交换网更高级的报警等;

(2) 综合分析监控管理模块。它是针对不同的职能部门形成不同的主题进行划分,主要包括综合故障管理模块(全网运行异常状况分析、告警监控、故障定位分析等)、性能管理模块(性能测试、评估性能分析、全网性能控制等)、综合配置管理模块(网络状态综合分析、设备安装监控分析等)、安全管理模块(网络及网络设备安全保护分析、入侵用户分析等)、计费管理模块(业务计费分析、业务关联分析等),它在各个专业网管基础上,为各个部门不同的主题建立分析模型,将数据进行集成,利用 OLAP 工具、DM 工具以及人工智能等方面知识,为不同主题生成决策参考信息,并且还提供了监督和内控系统,加强了风险防范和预警功能,使得网络管理在监督系统的监督下实现决策系统制定的期望目标,实现网管在安全基础上的最优运营。综合网管系统的体系结构如图 2。

图 2 中,XML 格式转换器是将各个专业网管的 XML 格式数据源进行统一,实现异构数据源模式和统一模式的相互转换,ETL 是数据抽取、转换、加载的数据集成工具,即在各专业网管基础上进行 ETL 集成到数据仓库中去,形成企业级的全局数据视图,各部门以其特有的主题在此基础上进一步集成各自的数据集市,OLAP 服务器对要分析的数据信息进一步按多维分析模型集成到多维数据库中,进一步为 OLAP 和 DM 工具提供服务。OLAP 工具以多维数据库中的数据为基础,针对用户的问题进行相关数据分析综合,数据挖掘(DM)工具在此基础上,按照一定算法自动地发现潜在模式和新知识。一方面,数据挖掘出的新知识可以补充到知识库当中;另一方面,知识库响应系统的请求,不仅可将结果提交给数据挖掘模块,用来更好的挖掘新知识,而且还可以使综合管理模块利用知识库中的新知识进行更好的实时处理。

基于数据仓库的综合网管系统以数据仓库为平台,以模型库和方法库为指导,并且利用知识库中的知识对网络数据

进行全方位分析和挖掘,发现潜在的新知识,补充到知识库中,知识库中的知识又能更好的为系统服务,从而使系统在一定程度上具有智能化,较好的协调数据、模型、方法、知识之间的关系,最大限度的使网络运营商获得最大效益。

4 综合网管系统的具体实施及关键技术

4.1 综合网管系统的具体实施

与传统数据库设计不同的是,数据仓库的设计是数据驱动的,根据电信行业的特殊性,开发系统时要以将对环境的影响程度降低到最小值为原则,因此,宜采用“自底向上”结合“自顶向下”方法(快速原型法)进行系统的开发,具体实施如下:进行需求分析,选择实现数据传输的软硬件环境,确定不同的主题,建立数据模型,确定数据仓库的逻辑结构,创建数据仓库的元数据;创建数据仓库,根据制定的元数据,按数据模型或元数据进行 ETL 转换并集成到数据仓库;按不同的主题,设计各自的数据模型,按模型将数据仓库中的数据集成到数据集中,创建多维数据结构的事实表和多维表,为各种信息工具服务;根据主题及解决问题的需要,设计模型库、方法库以及知识库,开发获取信息的工具;管理完善数据仓库环境,定期对数据仓库数据刷新,进一步完善各种工具,尤其是知识库的刷新,使系统逐渐达到完善和智能化,系统的完善是一个不断优化和完善的过程。

4.2 系统实现的关键技术

4.2.1 对各异构数据源的统一访问

系统采用 XML 统一规范不同的数据模式,实现数据之间的相互转换,采用 SQL 语句进行异构数据源的访问,用 XML 来传输、处理、显示数据。由上所述,首先对数据仓库中的数据类型、结构等用 Schema 定义规范化,而后在每个专业网管系统(数据源)上放置一个查询转换器,用来解释 XML 文档,并且在应用服务器上放置 XML 格式转换器,当进行查询访问时,经由 XML 格式上的统一模式传输到专业网管系统时,查询转换器将其转换成本地数据源模式,从而可进行本地访问,查询转换器将结果转换成 XML 格式数据,再经 XML 格式转换器上的 XSLT 转换成统一格式,这样可把多个数据源上的结果放到一个 XML 文档,送给浏览器,再通过 XSLT 转换成所需显示格式。如查询各市 2000 年的用户数,则可在数据仓库上生成一个 XMLSQL 语句:

```
<XMLSQL>
```

```
select < telecome. number /> where < telecom. year /> = 2000
```

```
< /XMLSQL>
```

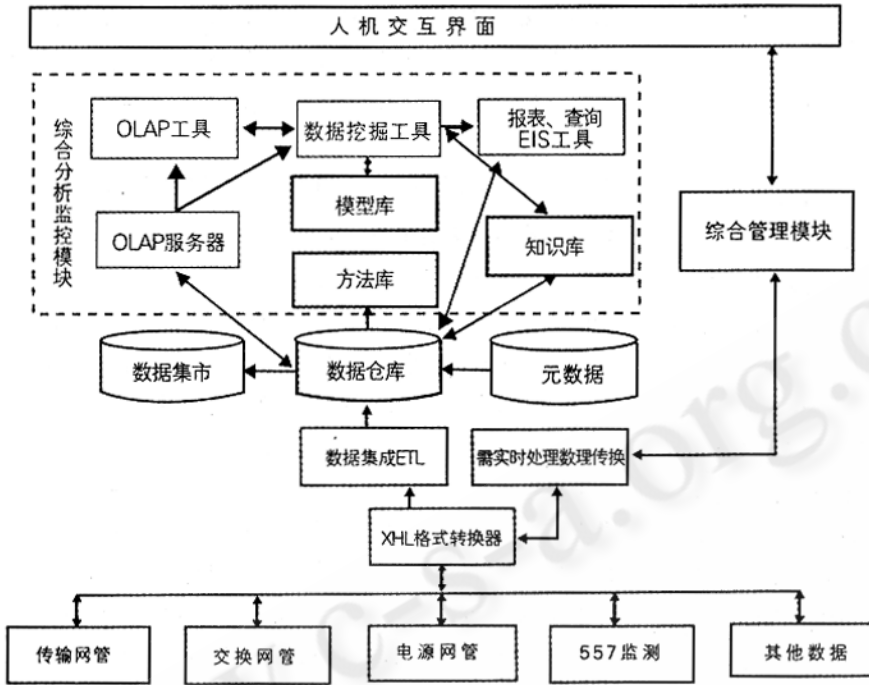


图 2 综合电信网络管理系统的体系结构

```
select citytelecom. people _ number
where citytelecom. year=2000
```

4.2.2 关系数据模式和 XML 模式的相互转换

数据库管理系统中的表和 XML 数据的树型结构难以形成对应关系,因此需要把 XML 数据进行解析后存储,取出数据时又需生成 XML 文件,即要将关系数据库与 XML 模式相互转换,基于 XML 的数据模型是一个带有注释的树状图,XML 元素与节点相应,属性与那些点的注释相应,因此,将关系数据库中的表进行标准化命名,按照某些算法映射为不同的 XML schema 文档和 XML 文档,根据 schema 文档,即可将关系数据库中的某个元级的某字段的值转换成 XML 文档中的内容、属性部分,再将数据从关系数据库中导入并加入相应的标记或将相应的节点根据 schema 文档导出到关系数据库。生成 schema 文档的具

体映射算法,可参考文献^[5,6]等。

```
<! --转换为本地的 SQL 语句-->
<XSL:template match="XMLSQL">
  <XSL:element name="SQL">
    <XSL:copy>
      <XSL:apply-template select="@ * ! node()" />
    </XSL:copy>
  </XSL:element>
</XSL:template>
<! --某一数据源上将统一对象名转换为本地对象名-->
<XSL:template match="telecom. number">
  citytelecom. people_number
</XSL:template>
<XSL:template match="telecom. year">
  citytelecom. year
</XSL:template>
.....
```

经过转换以后,就可在某一局部数据源上,执行 SQL 语句,即:

体映射算法,可参考文献^[5,6]等。

4.2.3 XML 安全访问控制

XML 主要关心的是数据,数据的传输、处理、显示和存储,这必然涉及到数据的安全问题,如加密、数字签名、访问控制等。XML 的几个重要安全标准有 XML Encryption (XML 加密)、XML Signature (XML 签名)、XKMS (XML Key Management Specification, XML 密钥管理规范)、XACL (XML Access Control Language, XML 访问控制语言) 和 SAML (Security Assertion Markup Language, 安全性断言标记语言) 等。为了保护数据仓库及网络的安全,我们不仅采用常规的在服务器上安装防火墙,而且根据 XML 的特点,采用 XACL 根据查询者的身份来决定访问控制权限。XACL 的目标在于为 XML 文档提供一个精细的访问控制模型和访问控制规范语言,利用这一访问控制技术,访问控制策略将控制一个 XML 文档如何显示给用户,从而实现基于角色的访问控制。另外,通过使用 XACL,我们可以根据访问权限对产生的数据进行过滤。增强了安全性,同时也可呈现出具有个性化的操作界面,从而简化了从大量数据中查找需要内容。

(下转第 51 页)

5 结束语

本文采用 XML 技术应用于 web 的数据仓库系统,简化了对数据仓库的操作、挖掘,系统不仅可以解决多个专业网管形成的“信息孤岛”、资源不能共享的问题,而且克服了电信网络中的海量数据却信息缺乏的障碍,引入了 OLAP 技术及知识库等数据挖掘技术在一定程度上智能化的生成业务规则,从全局出发为最优控制网络提供指导,帮助领导和决策者预防风险,提高服务和管理水平,为克服当前各专业网管所带来的不足提供了有效解决途径,对当前网管现状的改变将有现实意义。

参考文献

- 1 W H Inmon Building the Data Warehouse[M] China Machine Press 2003。
- 2 Mark Graves Designing XML Databases[M] China Machine Press 2002。
- 3 匡红阳,初探电信网络管理系统的现状和发展趋势[J],计算机系统应用,2003,3,66—69。
- 4 仇丽青、赵庆祯,基于 XML 的数据仓库系统 [J],计算机系统应用,2004. 2. 12—15。
- 5 吴文辉、殷建平、姚丹霖等,关系模式到 XML 模式的转换研究[J],计算机工程与科学,2004. 2. 94—96。
- 6 邓芳,XML 文档到数据库数据转换研究[J],北京邮电大学学报,2004. 1. 84—88。
- 7 鲍军鹏、张丽霞、刘晓东等,一个基于 XML 数据仓库系统的设计与实现[J],计算机工程与应用,2004. 6. 220—222。