

# 水利系统异构数据动态集成的设计和实现<sup>①</sup>

## Design and Application of the Heterogeneous Data Integration Based on Water Conservancy System

任 钢<sup>1</sup> 曹三顺<sup>2</sup> (1.中软公司华南基础构件中心 广东 深圳 518057;

2.飞来峡水利枢纽管理局 广东 清远 511000)

**摘 要:** 在随着水利企业信息化的逐步深入,企业内部之间存在着大量异构信息的交互和共享,开放性和集成性成了企业信息化的重要发展方向。本文就结合飞来峡水利枢纽数据集成系统的实际开发情况,提出了以元数据模型,Java 和 XML 技术建立适用于企业信息化的面向异构系统数据集成的框架结构,并对实现框架结构所涉及的技术设计和实现方法进行了详细的分析和说明。

**关键词:** 数据集成 异构数据集成 java 元数据 XML

### 1 引言

水利企业信息化系统一般都是多个数据源的应用系统。由于各个信息系统之间的应用存在着很大差异,没有统一的数据规范,这造成了各系统相互孤立,信息无法交流和交换。在内部逐渐形成了“信息孤岛”,给信息的共享和集中管理带来了麻烦。

飞来峡水利枢纽管理局就是这样的情况<sup>[1]</sup>。在信息化建设的初期,管理局建立了发电厂计算机监控系统、船闸计算机监控系统、船闸收费调度系统、泄水闸计算机监控系统、水情遥测预报和调度系统、水厂计算机监控系统、大坝安全监测系统、工业电视监视系统、卫星云图、办公自动化系统等多套计算机系统,但是这些系统之间存在的格式、标准、规范没有统一,所以,不能共享各系统的数据,有些专业系统本身也不可能形成统一的标准和规范。

于是,针对众多分布、异构并且自治的信息系统及其产生的数据进行集成,构建统一的数据集成框架平台,使得用户可以透明地访问各个成员数据源,并且该平台具有良好的可扩充性和伸缩性,以适应不断变化应用需求和不断涌现的数据源。本文就是正对上述情况而设计一个异构数据集成平台来实现各种类型异构数据的集成。

### 2 异构数据集成技术分析

异构数据集成技术简单地说是通过一定的技术手段把分散、异构的数据以统一、一致的形式表现出来,从而为用户屏蔽掉各种数据源的来源和结构等差别。一般来讲,一个功能完善的数据集成系统在解决数据的分布性和异构性的同时,还应能保持已有系统的自制性,并具有良好的可扩展性和适应性<sup>[2]</sup>。

建立异构分布式数据集成系统常见的方法有数据复制法和虚拟集中法<sup>[2,3]</sup>。前者是将各局部分散数据源中的数据复制到一个中央数据库中,用户的查询是基于实在的中央数据库中的数据。后者是一种基于“中间模式”(Mediated Schema)的方法,即数据仍然保存在局部数据源中,“中间模式”只记录全局数据模式和局部数据模式的对应关系,即集成规则。用户的查询基于这个虚拟数据库,所以他不必知道每个局部数据源的特点,而由数据集成系统来完成中间模式和局部数据源之间的查询转化。虚拟集中的方法更适合于数据源数目多、各局部数据源的自治性高、异构性大和局部数据经常变化的场合,本文的集成平台模型即是基于虚拟集中法模式<sup>[4]</sup>。

中间模式的集成规则描述可采用元数据方式<sup>[5]</sup>。所谓 Metadata,中文译名“元数据”,通常称为“data

<sup>①</sup> 基金项目:水利部 948 项目(200522)  
收稿时间:2008-10-30

about data”或是“data describes other Data”元数据的格式及其描述方法。元数据是关于数据以及操作数据的应用程序的结构和意义的信息描述，即关于数据的数据。元数据可以描述一个具体的资源对象，并能对这个对象进行定位管理，且有助于它发现与获取数据。一个元数据由许多完成不同功能的具体数据描述项构成，具体的数据描述项又称元数据项元素或元素。

对于不同异构数据源的访问，可以通过 java 来实现。Java 语言<sup>[6]</sup>是由 SUN Microsystems 在 20 世纪 90 年代中期开发的一种流行的面向对象编程语言，同时又具有一系列的面向对象程序设计规范。Java 规范中的 JDBC 是一种用于执行 SQL 语句的 Java API。它由一组用 Java 编程语言编写的类和接口组成，可以用于访问不同异构数据库。java 规范的 JMS 是用于和面向消息的中间件相互通信的应用程序接口(API)。它既支持点对点，也可支持实时数据集成。

对于统一的数据传输和表现，可以通过 XML 来实现。XML 是一种精简的 SGML(Standard Generalized Markup Language)语言，其有助于独立平台、易于提取信息的交换和数据，是数据交换的标准；当然，Java 的 javaBean 可以作为数据传输对象来使用。

Java 和 XML 能很好的结合起来使用<sup>[7]</sup>，国内学者基于 Java 和 XML 结合集成异构系统也做了很多的尝试<sup>[8]</sup>。

### 3 异构数据集成平台模型

异构数据集成平台通过系统内的多个组件共同协作以实现复杂应用逻辑<sup>[9]</sup>。在平台内部，各个功能组件以层次划分，并且不同层次之间通过组件之间的协议相互通信和调用。整个模型可以分为数据接口封装器和查询解释引擎两层，其中查询解释引擎作为外部查询应用的接口，数据接口封装器是查询解释引擎的接口并具体对异构数据进行操作，并将其操作的结果集返回给查询解释引擎组件。其结构模式如下图 1。

在这里，两层都要通过集成元数据规则模型来形成，在这里主要说明一下数据规则元数据模型。

#### 3.1 元数据描述模型

元数据是数据集成的基础，通过元数据模型，可以清晰地描述各个不同应用的数据源情况，由管理员使用手工对元数据进行管理。通过 XML 配置文件来描

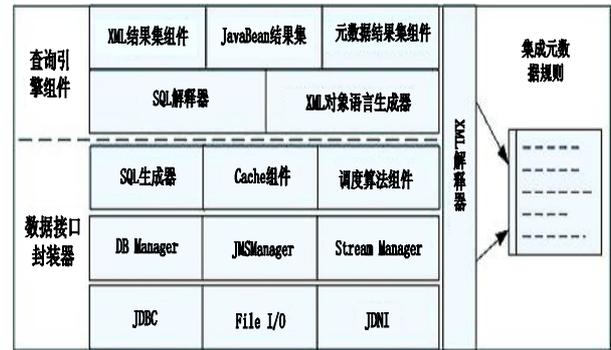


图 1 异构数据集成平台模型

述各个数据源的特性。数据源的特性包括数据源的类型，参数，调度算法，管理模式等。这样把不同数据源的本地模式转化到全局模式，方便统一的管理。

元数据模型采用 XML 文件格式来进行配置。其 XML 的 XSD 如下图 2。

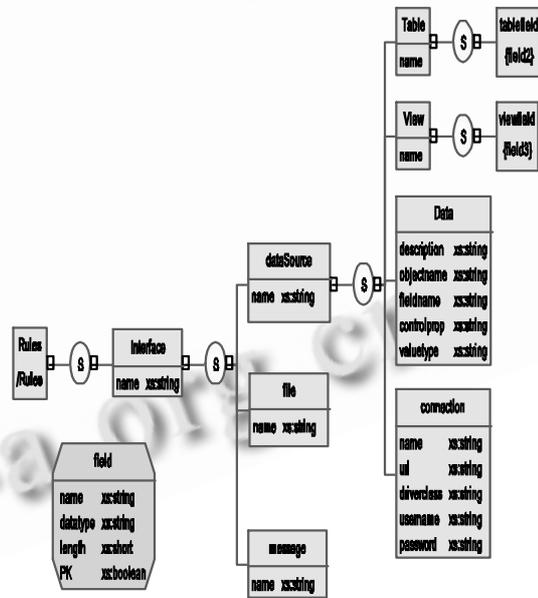


图 2 集成元数据规则的 schema 图

说明：interface 说明要集成的数据源名称。

Datastore 说明数据源的相关连接参数、物理位置和驱动程序，通过该信息，可以访问异构数据库。

File 说明数据源是文件，包含操作文件的相关属性。

Message 表示数据源是消息，包含对消息的操作规则，主要针对实时数据。

Table 说明数据源名称中关系数据库中的表内容。

**View** 说明数据源名称中关系数据库中的视图内容。

**Data** 说明数据源名称中具体要集成的数据,按照一一对应的原则,包括数据对应的对象(JavaBean)和数据传输对象(DTO)。

### 3.2 查询解释引擎组件

用户在表现层的查询要求,通过查询语句提交到查询解释引擎,查询解释引擎按照集成元数据规则,把提交的查询语句进行重新分解和解析,形成封装器支持的对象语句,提交到接口封装器组件上,接口封装器组件把这些对象语句再按照元数据模型的形成 SQL 语句,访问对应数据源的数据库,把返回的数据集形成对象数据组。返回给表现层的查询要求,表现层把对象数据组分解成数据,展现给用户。

查询解释引擎组件包括如下组件:

① **数据集组件**: 获得的结果集有三种类型,一种是 XML 数据结果集,一种是元数据结果集,还有一种是对象结果集(JavaBean 对象)。这三种结果集是遵照集成元数据规则不同的定义而满足上层应用层不同的应用需要。

② **SQL 解释器**: 把用户提交的 SQL 进行分析和解析,按照集成元数据模型规则形成对象查询语句。

③ **XML 对象语言生成器**: 把用户提交的 SQL 进行分析和解析,按照集成元数据模型规则形成 XQuery 查询语句。

### 3.3 接口封装器组件

接口封装器是基于元数据描述模型对应的成员数据源进行封装。它是屏蔽了数据源逻辑上的差异,为系统提供透明化的服务。对于上层应用而言,这样的封装器隐藏了数据源在应用逻辑、查询机制和物理位置的差异,使得上层应用不用直接面对不同的数据源。接口封装器组件对外部只提供接口。对异构数据源来讲,接口封装器就是集成规则的翻译器。接口封装器是由单个或者多个组件组成的复合组件库,其内部组件组成如下:

① **SQL 生成器**: 把查询组件的对象查询操作转化成数据源数据库系统的 SQL 语句。

② **Cache** 组件主要针对实时数据。当用于在本地缓存按照一定时间间隔获得的实时数据,当上层应用要获取实时数据,即从缓存中获取。

③ **调度算法组件**主要各种操作的时间间隔和顺

序。其执行规则也是在集成元数据规则中进行配置。

④ **DBManager** 组件主要针对数据库操作。**DBManager** 通过 SQL 生成器组件生成不同数据库平台的特有 SQL 语句,再通过标准接口 JDBC 访问数据源,取得查询数据并将这些数据按照查询结构组织成以 XML 结构表达的结果文件。

⑤ **StreamManager** 主要针对文件 I/O 操作,可对本地或异地文件进行读取操作。

⑥ **JMSManager** 组件主要针对消息操作,通过对实时数据采用提取操作,并把获得的结果集按照 XML 结构表达或 DTO 返回。

## 4 实现模式

异构数据集成平台在整个系统中的角色如下图 3。

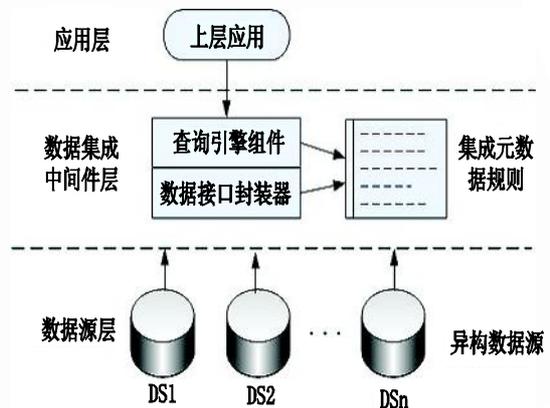


图 3 异构数据集成平台实现模式图

集成平台是核心,它接收应用系统的请求后,由查询解析组件把请求分解封装器组件可以理解的操作并把这些操作提交封装器组,封装器组根据配置文件信息访问具体的数据源获得数据,返回数据给查询解析组件。查询解析组件转化数据后把数据返回应用系统。

封装器组件负责操作数据源,并和对应的查询解析组件进行交互,返回查询解析组件要求的数据。在实际应用中两种情况,一种是通过集成平台获取异构数据库的数据。另一种是通过集成平台去获得实时数据。第一种模式是获取数据进行业务逻辑处理,完成数据集成操作并可以形成各种报表等。第二种是获取数据后,按照一定的时间间隔在表现层显示这些数据。这两种处理的模式也是不一样的。

第一种处理模式如图 4。

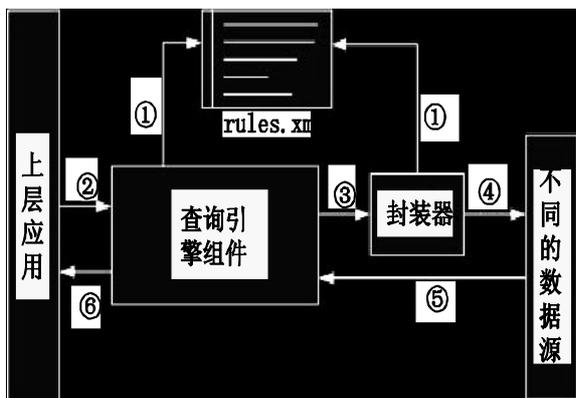


图 4 第一种集成数据模式

实现步骤说明：

- ① 服务启动就加载元数据模型配置文件，并将配置信息一直都保存在应用服务器的内存中。
- ② 上层应用系统层请求的数据查询，提交到经过查询引擎解析。
- ③ 查询引擎组件解析查询语句并提交给接口封装器组件。
- ④ 接口封装器组件按照配置文件的要求，通过 JDBC 方式获取异构数据库数据。返回给上层应用系统层。
- ⑤ 获得的数据返回到查询引擎组件。
- ⑥ 查询引擎组件解析数据集，把数据发送给上层应用系统。

第二种处理模式如图 5。

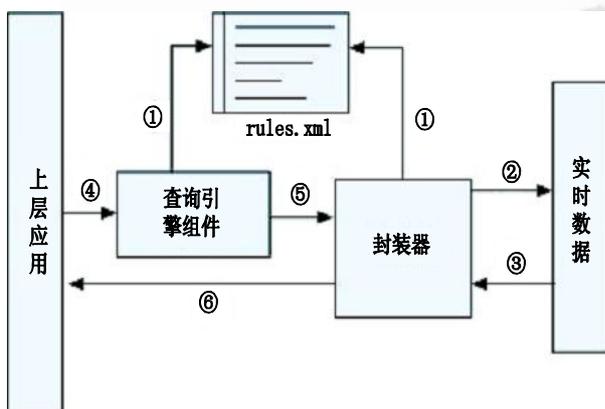


图 5 第二种集成数据模式

实现步骤说明：

- ① 服务启动就加载元数据模型配置文件，并将配

置信息一直都保存在应用服务器的内存中。

② 封装器按照配置文件的要求，定时地通过 JDBC 方式获取异构数据库数据或通过 JMS 方式定时地获取实时数据。

③ 获取的实时数据都保存在应用服务器上的缓存中，并且动态更新。更新的对象只是一些变化的数据。

④ 上层应用系统层请求的数据查询，提交到经过查询引擎解析。只用访问应用服务器上变化的数据。

⑤ 查询引擎组件直接提交给接口封装器组件。

⑥ 接口封装器组件按照配置文件的要求把缓存数据把数据发送给上层应用系统。

## 5 具体应用

针对飞来峡水利枢纽管理局众多分布、异构并且自治的信息系统的现状，为统一共享各个应用系统的数据，飞来峡水利枢纽管理局采用了本集成平台进行数据集成，建设飞来峡水利枢纽综合集成管理系统。

在飞来峡水利枢纽综合集成管理系统中异构集成平台是核心，集成元数据规则又是集成平台的核心，它接收应用系统的请求后，由查询解析组件把请求解析后提交到数据接口封装器组件，再由封装器组件访问具体的数据源获得数据。可访问的数据源的类型分为四种：① 关系数据库系统(如 ORACLE、SQL Server 和 Lotus Domino 等)支撑的数据源，② 以文件形式作为数据存储手段的数据源，③ 非关系数据库(如 Lotus Domino)；④ 实时数据。

在该层次模型下，本系统具备了以下的四点优点：

- ① 采用动态 XML 配置文件，可以实现灵活的数据集成，提高了集成平台的可扩展性和可伸缩性。
- ② 实现对异构数据源的透明访问。
- ③ 利用缓存技术和灵活的信息获取方式有效地提高信息检索的效率。
- ④ 系统具有良好的包容性，可以集成多种形式的数据库源，可以包括各种形式的文件。

异构数据集成平台的引入可以规范各个信息系统产生的各种文件和数据，为用户提供一个统一的检索平台，为开发人员提供一个统一的数据访问接口，这样就保持各个数据源的应用系统自治性，又可以保证数据的一致性。

(下转第 61 页)

## 6 结束语

本文以飞来峡水利枢纽管理局数据集成系统为案例,详细地介绍了基于异构数据集成平台把各个分布式数据源集中管理的方式。该方案通过元数据集成规则的灵活配置及数据缓存区的设置,在实现中采用了目前比较流行的 XML、Java 中间件在数据集成和传输方面的应用技术。实现了异地数据库规模大、数据更新频繁、网络较繁忙情况的数据集成。通过本平台的设计和实现,对解决异构数据集成的问题,提供了一个切实可行的模式和方法。

### 参考文献

- 1 曹三顺.飞来峡水利枢纽计算机集成管理系统.红水河,2006,25(1):100-103.
- 2 高明,陈昕,李炜,宋潮涛.基于 XML 实现异构数据源的联合使用.计算机科学,2002,29(3):83-84,93.

- 3 戴青云,樊沛,李刚,董逸生.Web-based 多数据源集成系统的研究.计算机科学,2002,29(9):119-120,F004.
- 4 张忠庆,赵敬中.基于虚拟集中方法的异构分布式数据集成模型.微机发展,2004,14(6):58-60.
- 5 杨鸿宾,宋明.元数据管理平台总体架构设计研究.计算机系统应用,2007,16(11):17-20.
- 6 张清玉, Schildt H. Java 2 参考大全.北京:清华大学出版社,2002.
- 7 李大成,陈莘萌.Java 与 XML 的结合应用.计算机应用,2002,22(2):59-64.
- 8 陶以政,唐定勇,何铁宁,李秋灵,周南华,雷海红.基于 Java 和 XML 技术的异构信息系统数据集成框架应用研究.计算机应用,2004,21(5):38-40.
- 9 李培峰,朱巧明,钱培德.基于组件的异构数据集成平台的设计与研究.计算机应用与软件,2005,22(9):41-43,107.