

多源环境监测信息交换模型的设计与实现

何春银, 黎刚

(江苏省环境信息中心、江苏省环境监测中心, 江苏 南京 210036)

摘要: 针对环境监测系统现有的污染源管理、饮用水管理、噪声管理、自动监测管理等信息系统存在异构化、多源化等现象, 造成数据重复录入、数据不一致、资源浪费等问题, 展示了一种数据交换模型的设计思路和实现方法, 介绍模型的数据抽取、数据封装、数据传递、数据解析、数据映射、校验和转换、数据接收等多个环节。该数据交换模型已成功实现与多个外部业务系统数据对接, 现每天采集的数据超过 5 000 条, 系统运行稳定、性能良好。

关键词: 信息管理系统; 数据交换模型; 环境监测信息

中图分类号: X830.3 **文献标识码:** B **文章编号:** 1006-2009(2010)06-0069-04

Design and Implementation of A Multi-source Environmental Monitoring Information Exchange Model

HE Chun-yin, LI Gang

(Jiangsu Province Environmental Information Center, Jiangsu Province Environmental Monitoring Center, Nanjing, Jiangsu 210036, China)

Abstract Environmental monitoring system existed problems such as pollution source management, drinking water management, noise control, and similarly automatic monitoring management information system had problems such as isomerization, multiple source to cause repetitive data entry, inconsistent data and waste of resources. A new design idea and implementing method showed that a data exchange model performed data extraction, data encapsulation, data transfer, data analysis, data mapping, calibration, conversion and data receiving etc. Based on this model, new data processing system had been successfully realized the data exchange and connection with multiple external business system. Now the new processing system excellently operates with stable performance and deals with more than 5 000 messages every day.

Key words Information management system; Data exchange model; Environmental monitoring information

不同阶段环境监测的信息系统由不同建设单位主导设计, 由此产生的数据部门化、异构化问题给决策层造成较大的困扰, 为保护原有系统的投资, 避免“信息孤岛”造成的重复录入、资源浪费、数据不一致性等问题^[1], 充分利用现有资源, 必须将多种异构数据源整合, 实现全面信息管理、分析和共享^[2-5]。

针对项目的实际情况, 考虑业务伸缩性和系统扩展性, 对于每一个需要与中心数据库交换的系统, 在交换服务器为其配备基于数据交换模型的服务, 将数据转换成标准格式后向中心服务器发送。

1 数据交换模型的设计

1.1 模型总体架构

考虑到模型的兼容性和可扩展性, 模型分为多个环节。为便于数据的描述, 在项目中统一采用 XML 文件作为交换格式。业务系统服务器可与交换服务器相连, 也可物理隔离; 交换服务器与中心数据库服务器直接连接。对于与交换服务器直接

收稿日期: 2010-08-01; 修订日期: 2010-10-27

基金项目: 国家“十一五”水体污染控制与治理科技重大专项基金资助项目 (2008ZX07101-014-004, 2009ZX07527-004-004)

作者简介: 何春银 (1962-), 男, 江苏南京人, 高级工程师, 硕士, 从事环境信息技术与管理工。

网络连接的业务系统, 由数据交换服务器将业务数据库抽取、封装好的 XML 文件解析、映射、校验、转换成标准数据库格式后上传至中心服务器。对于与中心服务器物理隔离的业务系统, 可采用手工交换方式, 由专人负责定时抽取临时数据库封装为 XML 文件, 通过移动介质复制到交换服务器, 由交换服务器进行解析处理。数据交换模型的总体架构见图 1。

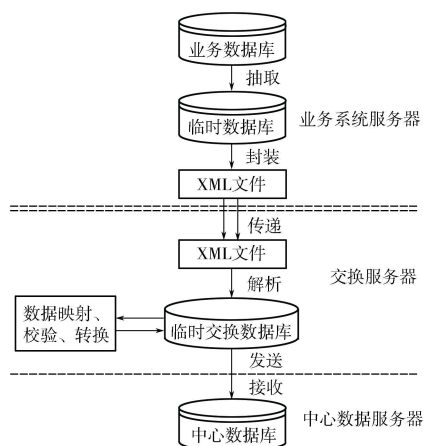


图 1 数据交换模型总体架构

Fig 1 Flow chart of data exchange model

1.2 数据抽取和数据封装

对业务系统的数据库进行分析后选择必要的数据生成临时数据库。从临时数据库中读出数据, 为确保对该类型数据传输交换的规范性, 根据不同的数据选择相应的架构文件封装, 生成 XML 文件。

以太湖水体水环境质量监测数据为例, 定义其数据架构文件 (XML Schema) 格式如下:

```

<? xml version = "1.0" encoding = "UTF-8"? >
< xs:schema xmlns:xs = "http://www.w3.org/2001/XMLSchema" elementFormDefault = "qualified" attributeFormDefault = "unqualified">
< xs:element name = "TailakeMDatas">
< xs:annotation>
< xs:documentation> 太湖水体监测数据根节点</xs:documentation>
</xs:annotation>
< xs:complexType>
< xs:sequence>
< xs:element name = "MDatas" type = "
  
```

```

TailakeMDatasType" minOccurs = "unbounded"/>
</xs:sequence>
</xs:complexType>
</xs:element>
< xs:complexType name = "
TailakeMDatasType">
< xs:annotation>
< xs:documentation> 太湖水体监测数据类型</xs:documentation>
</xs:annotation>
< xs:sequence>
< xs:element name = "测站代码" type = "xs:integer"/>
< xs:element name = "测点代码" type = "xs:string"/>
< xs:element name = "测点名称" type = "xs:string"/>
< xs:element name = "年" type = "xs:integer"/>
< xs:element name = "月" type = "xs:integer"/>
< xs:element name = "高锰酸盐指数" type = "xs:decimal" minOccurs = "0"/>
< xs:element name = "总磷" type = "xs:decimal" minOccurs = "0"/>
< xs:element name = "总氮" type = "xs:decimal" minOccurs = "0"/>
< xs:element name = "叶绿素 a" type = "xs:decimal" minOccurs = "0"/>
< xs:element name = "透明度" type = "xs:decimal" minOccurs = "0"/>
< xs:element name = "富营养化指数" type = "xs:decimal" minOccurs = "0"/>
< xs:element name = "富营养化级别" type = "xs:string" minOccurs = "0"/>
< xs:element name = "水质类别" type = "xs:string" minOccurs = "0"/>
</xs:sequence>
</xs:complexType>
</xs:schema>
  
```

根据以上架构从临时数据库封装, 生成待交换 XML 数据文件^[6]。案例如下:

```

<? xml version = "1.0" encoding = "UTF-
  
```

```
8"? >
< TailakMDatas xmlns xsi= " http //www.
w3. org/2001/XMLSchema- instance">
< MData>
< 测站代码 > 320200< 测站代码 >
< 测点代码 > P1< 测点代码 >
< 测点名称 > 拖山< 测点名称 >
< 年 > 2010< 年 >
< 月 > 3< 月 >
< 高锰酸盐指数 > 4 3< 高锰酸盐指数 >
< 总磷 > 0 080< 总磷 >
< 总氮 > 3 51< 总氮 >
< 叶绿素 a > 0 018< 叶绿素 a >
< 透明度 > 40< 透明度 >
< 富营养化指数 > 58 5< 富营养化指数 >
< 富营养化级别 > 轻富< 富营养化级别 >
< 水质类别 > 劣 V< 水质类别 >
< /MData>
< MData>
< 测站代码 > 320200< 测站代码 >
< 测点代码 > P2< 测点代码 >
< 测点名称 > 椒山< 测点名称 >
< 年 > 2010< 年 >
< 月 > 3< 月 >
< 高锰酸盐指数 > 4 3< 高锰酸盐指数 >
< 总磷 > 0 120< 总磷 >
< 总氮 > 6 10< 总氮 >
< 叶绿素 a > 0 004< 叶绿素 a >
< 透明度 > 20< 透明度 >
< 富营养化指数 > 59 5< 富营养化指数 >
< 富营养化级别 > 轻富< 富营养化级别 >
< 水质类别 > 劣 V< 水质类别 >
< /MData>
.....
< /TailakMDatas>
```

1.3 数据传递和数据解析

将待交换的 XML 文件直接或间接传输到交换服务器指定目录。交换服务器定时从数据接收文件夹中读取文件,选择对应的 XML 架构文件进行解析(可同时进行语法格式的初步校验),将结果写入临时交换数据库。

1.4 数据映射、校验及转换

交换服务器启动数据处理程序,对数据的有效

性、完整性、一致性等内容进行检查,对于其中涉及代码转换的,还需根据代码对照表对数据中的代码进行转换,完成后将结果写入临时交换数据库。

上例中测点代码过于简单,与其他地区湖泊测点编码重复,与中心数据库的标准不一致,可将编码转换为标准格式,转换的对照表见图 2。

测站代码	测点代码	测站代码	测点代码
320200	P1	320200	320200001
320200	P2	320200	320200002
...	...	...	...

图 2 编码转换对照

Fig. 2 Code trans list

1.5 数据发送与接收

通常将交换服务器与中心数据库服务器部署在同一网络并进行连接。交换服务器向中心数据库服务器申请发送数据,经中心数据服务器批准审核后,中心服务器进入等待接收状态,交换服务器进入发送状态,数据交换传递至中心服务器。

在交换服务器与中心数据库服务器不直接连接的复杂情况下,可参考该文已经叙述的方法将交换服务器临时数据库再次封装为 XML 文件后传递,由中心数据库解析导入。

2 数据交换模型的实施

2.1 数据交换服务器及引擎部署

现有环境监测软件大多采用微软的. Net 框架,为保持技术路线上的一致性,选择 BizTalk Server 2009 作为搭建数据交换平台的主要服务引擎。BizTalk Server 是微软公司推出的一种系统间业务协同和数据交换的中间件产品,通过简单配置、开发可形成一个基于标准的、可扩展的交换和共享应用环境。其核心功能包括^[7]: 流程执行引擎,应用业务逻辑管理,复杂流程和事务集合程序调用;业务流程服务设计、执行和管理功能;底层消息传递服务(如发送、接收、分析和跟踪消息文档)和数据映射;跟踪与调试,业务活动监视。

2.2 基于 XML 的数据架构标准定义

数据交换必须采用标准化的数据格式,考虑到 XML 的可扩展性和自描述性及其他特性,选择 XML 作为数据架构标准定义的描述语言。XML 具有的特点^[8]为: ①具可扩展性,XML 是元标记语

言,可以定义其他语言;②结构性强,易于处理,XML对格式的定义非常严格,具有良好层次结构;③纯文本,与平台无关,XML将数据保存在标准的文本文件中,可以使用任何文本编辑工具创建和编辑XML文档,毋须更改直接移植到其他平台上;④数据存储与数据显示分离,XML文档定义仅仅是数据存储的结构和关系,对于数据的显示没有任何约束和限制,为XML最值得称道的地方;⑤强大的链接能力,可以定义双向链接、多目标链接、扩展链接和两个文档间的链接。

3 数据交换模型运行管理

在环境监测数据中心收集的数据中,大部分为各种自动站的实时环境监测数据,数据量较大,对数据交换时效性要求较高。因此,为加强管理,必须要对数据交换模型的运行过程综合监视,及时掌握整个数据传输交换过程的情况。该项目利用BAM API工具开发数据监视程序,对数据交换平台的运行状态和性能进行实时监视,收集运行时流程信息状态,并存储到BAM数据库。

3.1 收集和显示流入状态信息

实时收集流入交换服务器的数据当前流量(条/s)、当天/当月累计流量、处理速率(条/s)、当前尚未处理完成的数据等,并在监视界面上动态更新各类数据状态。

3.2 收集和显示流出状态信息

实时收集流出交换服务器的数据当前流量(条/s)、当天/当月累计流量、当前尚未发送完成的数据量等,并在监视界面上动态更新各类状态。

3.3 收集和显示每类数据的详细交换信息

以列表形式显示每类数据每日交换信息,包括当天上报条数、数据缺失率(与管理要求相比缺失多少数据)、最近上报时间等,并以颜色区分不同交换状态的数据源,使系统维护管理人员能一目了然地了解当前存在的主要问题。

3.4 故障自动报警

系统对数据交换过程中出现的常见故障如网络中断、监控平台数据库连接错误、某类数据超过规定时间未上报、对方服务器无响应等问题记录,并通过短信等方式向指定人员发送报警信息,以便管理人员能快速对故障作出响应。

3.5 统计分析

根据设定的条件,对规定时间范围内数据交换运行情况汇总统计,包括数据缺失率、数据按时上报率和故障率等。

4 结语

数据交换模型自投入运行以来,已成功实现与多个外部业务系统数据对接,发挥了数据集成和共享的作用。当前每天采集的数据超过5000条,系统运行稳定、性能良好,为环境监测中心数据库建设打下良好基础。采用该模型建立的数据交换平台具备的优点为:①可管理性:提供统一的管理和过程监视工具,对所有的数据交换任务综合管理。②可维护性:在管理系统的控制下,提供方便的维护工具和途径,可通过日志快速定位故障原因。③松耦合性:中心数据库和交换系统本身不受外部业务系统影响,在外部业务系统调整时,只需更新修改相应的数据抽取程序或数据映射关系等小型过程模块即可。④安全性:系统在传输过程中可选用各种加密方法,保证数据安全。⑤可扩展性:可快速加入新的外部数据源。⑥经济性:集中进行数据交换,各个分支的节点可以极大地降低管理和维护成本。⑦宽适应性:可实现跨局域网、跨系统、跨数据库、跨平台的数据共享和交换^[9]。

[参考文献]

- [1] 曾小宁,黎明. 基于XML的数据交换中间件的研究与实现[J]. 计算机工程与设计, 2007, 28(12): 2999-3002
- [2] 王合生. 环境监测信息化建设分析——走中国式环境监测信息化建设道路[J]. 环境监测管理与技术, 2006, 18(5): 5-7.
- [3] 王向明,伏晴艳,刘红,等. 环境监测实验室信息管理系统建设——以上海市环境监测中心为例[J]. 环境监测管理与技术, 2007, 19(4): 8-12.
- [4] 沈艺. 环境监测实验室信息管理系统的构建与实施[J]. 环境监测管理与技术, 2006, 18(4): 8-10
- [5] 徐晓宁,廖鸣镝. 多参数环境自动监测数据采集传输装置研究[J]. 环境监测管理与技术, 2007, 19(6): 19-21.
- [6] 马广慧,邵秀丽,李耀芳. 基于XML的异构数据源间数据交换的实现[J]. 天津师范大学学报(自然科学版), 2008, 28(1): 76-80
- [7] 王振宇,罗晓菁. 基于MS BizTalk服务器实现数据集成[J]. 计算机工程与设计, 2007, 28(10): 2435-2442.
- [8] A IKEN P.G. 微软XML技术指南[M]. 北京:中国电力出版社, 2003.
- [9] 刘艳辉,董碧丹,张峰. 数据交换平台的分布式应用研究[J]. 计算机工程与设计, 2009, 30(16): 3780-3782