

基于ETL的地质环境遗留系统集成及应用研究

郑庭明

(山东省地质环境监测总站, 山东 济南 250014)

摘要: 本文在介绍了遗留系统及相关知识的基础上, 阐述了为什么要对地质环境遗留系统进行集成, 并比较分析了不同系统集成方法的优缺点, 最终选择采用ETL技术进行数据集成的方式。介绍了ETL技术及其应用框架, 并用地质环境一张图进行了示例说明。

关键词: ETL; 地质环境; 遗留系统; 集成

中图分类号: TP39 **文献标识码:** A

Study on Integration and Application of Geological Environment Legacy System Based on ETL

ZHENG Tingming

(Shandong Monitoring Center of Geological Environment, Jinan 250014, China)

Abstract: This paper illustrates why we must integrate geological environment legacy system based on the introduction of the legacy system and related knowledge. In view of the advantages and disadvantages of different system integration methods, this paper chooses ETL technology to integrate data. Then this paper elaborates the ETL technology and its application framework by giving an example of a map of geological environment.

Keywords: ETL; geological environment; legacy system; integration

1 引言(Introduction)

遗留系统指不再适应新的技术或新的业务需求, 但鉴于替换或修改的成本过高而继续投入使用的计算机系统或应用程序^[1]。

地质环境系统的开发思路往往是使用者意识到需要某种功能时就开发具有该功能的系统, 导致地质环境存在较多时期不同、操作系统不同、数据库不同、功能不同的遗留系统。这些遗留系统在运行过程中, 产生和累积了海量的业务数据, 但是各遗留系统的业务数据分布在各自的系统中, 各系统的数据库管理系统、数据标准、操作系统都不相同, 导致这些海量数据之间缺乏关联, 形成“信息孤岛”^[2-6]。利用何种系统集成技术对这些遗留系统进行有效的集成, 在保证不影响现有遗留系统的使用, 又避免重复建设, 节约成本的基础上实现信息共享, 提高地质数据资产使用和管理效率, 这既是地质环境信息化建设中急需解决的问题, 同时也是构建共享开放的数据资源体系的重要一环。

2 系统集成技术选型(Selection of system integration technology)

近年来, 在系统集成、数据集成、工作流程、信息利用等方面出现了很多解决方案^[7], 详见表1。

表1 信息系统集成技术

Tab.1 Information system integration technology

信息系统集成思路	特点
数据集成	当前系统集成技术实际应用过程中最为为基础的内容, 能最为直接的体现集成技术
API集成	一种常见图形信息管理形式, 做好API集成的重要基础是做好数据汇集
方法集成	结合商业逻辑思维设计的一种操作软件

地质环境遗留系统正是拥有大量的历史数据(如灾害点动态监测数据, 地下水动态监测数据), 数据也比较稳定(如地质灾害点, 地质遗迹点), 数据需要长期、系统的保存, 加之基于数据的集成可不影响遗留系统已有的架构、操作, 又可避免重复建设, 还可保持数据的“鲜活度”, 故此本文选用数

据集成的技术。

鉴于ETL可以对各种数据源(如Oracle、Sqlserver、DB2、Mysql、Access、txt、excel、csv、xml等),不同数据库类型间的数据进行处理,数据处理组件丰富,通用性强,可以很好地解决数据抽取、转换、加载的问题等特点,故本文选用ETL技术进行数据集成。

3 关键技术(Key technology)

ETL(Extraction-数据抽取、Transformation-数据转换、Loading-数据加载)其实质就是一种数据转换的工具,在数据源与目标数据库中搭建一座桥梁,并且在桥上设置“检查站”,通过“检查站”去除“脏数据”,确保高质量的新数据源源不断地流向目标数据库。ETL过程其实就是数据流动的过程,将不同来源的数据统一汇入目标数据库^[8]。

ETL具有的功能如图1所示^[9],通过连接、抽取、转换、加载等流程完成多源异构数据的清洗、汇总。



图1 ETL功能

Fig.1 ETL functions

ETL流程图如图2所示,其具有操作简单、支持各类数据源、支持各种软硬件平台、通用性强、提供灵活的定制规则、提供强大的管理功能等优势^[10,11],利用ETL的优势,可在网络相通的前提下实现各级部门之间的数据集成与调用,提高数据利用的广泛性、高效性。

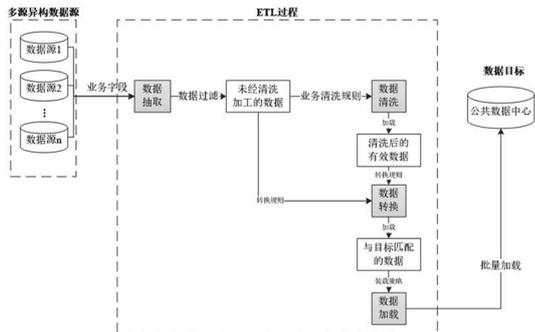


图2 ETL流程图

Fig.2 ETL flow chart

4 系统设计(System design)

4.1 技术框架

利用ETL对遗留系统进行集成的整体框架如图3所示。各遗留系统在各自原有平台进行独立运作,通过数据同步交换中间件(即ETL工具)将遗留系统的数据进行整合,统一汇集到公共地质环境数据中心,在公共地质环境数据中心之上构建信息平台,信息平台中包含各类业务应用需要的数据服务,当新建的业务系统需要使用遗留系统中的数据时,只需通过调用各类服务接口即可,遗留系统及新建的业务系统只需通

过UI集成便可搭建服务于自己需求的应用门户。

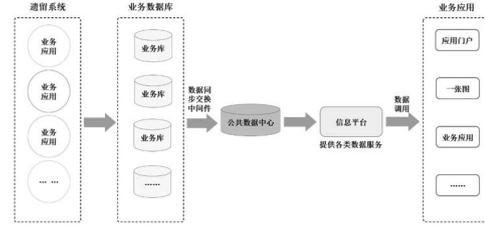


图3 系统框架

Fig.3 System framework

从系统架构层面,集成地质环境遗留系统采用的系统架构是“平台+数据+应用”的模式,遵循平台管理层、数据管理层和应用管理层三层技术架构,其技术架构见图4。

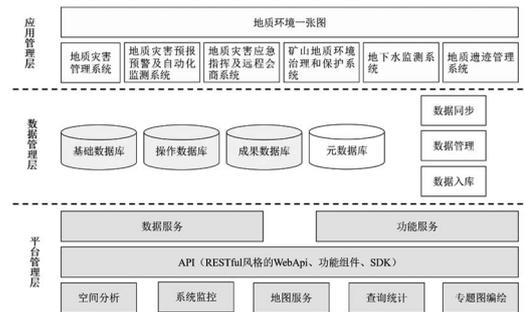


图4 技术架构

Fig.4 Technical framework

平台管理层:基于SOA技术,实现了数据存取、调度、查询、统计、分析、可视化等基础功能,全面支持B/S应用系统和移动端应用的开发和运行。该部分功能实现了面向地质环境业务应用系统的常规数据库访问、Web GIS、Web Service数据服务和数据交换应用。主要组件支持分布式运行、集中管理和一体化服务。以Web服务、REST API和功能组件等服务接口暴露给最终用户的二次开发服务接口。

数据管理层:系统产生的数据纳入公共地质环境数据中心统一管理运维,构建充分共享、适度开放、安全可控的地质环境数据共享开放新机制,实现地质环境全数据共享,推动地质环境数据在系统内外的充分共享和有序开放。

应用管理层:系统需要的各类组件、服务、数据等资源,由地质环境信息平台负责统一管理、配置、调度和监管。

4.2 应用实例

以地质环境一张图为例,传统的地质环境数据都是分散存储在不同的数据库中,业务数据分散,且省市县地质环境数据无法实现上下贯通和共享开放,地质环境一张图各类业务数据的调用更是荆天棘地^[12]。本文先利用ETL将地质环境遗留系统中的数据集成至公共的地质环境数据中心,基于公共地质环境数据中心提供各类数据服务接口,并通过信息平台展示各类数据服务接口,服务接口采用Restful形式,方便调用,地质环境一张图直接调用发布的数据服务接口,直接从公共地质环境数据中心调用数据,大大降低了数据获取和使用成本。下图是使用ETL进行数据集成时的系统界面示例,图5是主作业流程,图6是主转换流程。



图5 主作业流程

Fig.5 Main operation flowchart

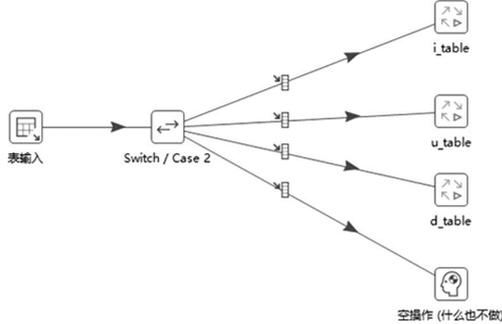


图6 主转换流程

Fig.6 Main conversion process

集成后的地质环境一张图以公共地质环境数据中心为依托，包括不同比例尺的基础地理、基础地质、水文地质、灾害地质、地下水监测、矿山环境保护、地质遗迹、地质公园保护等信息，实现地质环境各类信息的浏览、查询、空间编辑(包括点、线、面元的增、删、改)、空间计算等功能；实现空间信息与地质环境调查、监测、保护等专业属性信息的联动查询，同时可根据新建工程的地理范围对各业务数据进行综合查询。

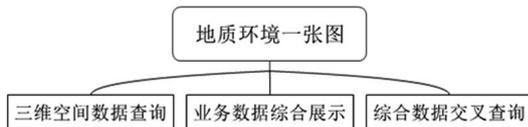


图7 地质环境一张图功能结构

Fig.7 Functional architecture of a map of geological environment

三维空间数据查询——将基础地理、基础地质、水文地质、工程岩组、灾害地质、地质灾害易发性分区等各类专题图层进行组合及叠加显示、浏览。同时，获取图元属性信息的详细说明，以便对地质环境进行分析。

业务数据综合展示——可在三维图上圈定查询范围(地质单元、流域等)，对地质灾害、地下水、矿山地质环境、地质遗迹等信息进行查询和统计分析，帮助快速进行地质环境影响评估。同时提供地质灾害和地下水的监测数据、设备状态的实时查看。

综合数据交叉查询——综合数据交叉查询可通过空间查询及手动录入拐点坐标两种方式圈定空间分析的主线路，并可设置圈定范围往外延查询的缓冲区距离，通过交叉查询，掌握沿线范围及其缓冲区分析范围内所有的灾害点、地下水监测点、矿山、地质遗迹点、地质公园、“三区两线”(重要

风景名胜、自然保护区、城市规划区、铁路和重要公路沿线、海岸线周围)等信息及其相关的空间数据。



图8 地质环境业务数据综合查询

Fig.8 Comprehensive query of geological environment service data

5 结论(Conclusion)

本文阐述了地质环境遗留系统集成方法及流程，并介绍了ETL技术在地质环境遗留系统集成中所具有的优势：它是一种基于数据的集成，无须改动原有系统，可适用于数据文件的集成、不同类型数据库间的数据集成，这样的集成方式不仅使用与地质环境遗留系统，对其他行业的遗留系统集成应用也具有参考价值。

参考文献(References)

- [1] YANG H,WARD M.Successful evolution of software systems[M].Norwood: Artech House,2003.
- [2] 刘传正,张明霞,刘艳辉.区域地质环境可持续利用评价体系初步研究[J].地学前缘,2006,13(1):242-245.
- [3] 邢丽霞,罗跃初,李亚民,等.我国地质环境监测现状及对策研究[J].资源与产业,2011,13(3):110-115.
- [4] 张熠斌,谭文清,徐思瑜,等.吉林省地质环境信息化建设实施方案[R].2014.
- [5] 张熠斌,宋金红,刘传深,等.面向服务的地质环境三维一张图系统研究[J].科学技术创新,2017,(33):1-3.
- [6] 吴红梅,王志强,吴友章,等.山东省地质资料集成与应用研究[J].山东国土资源,2017,33(12):70-74.
- [7] 朱思邈.计算机网络系统集成技术分析与研究[J].黑龙江科学,2017,8(21):122-123.
- [8] 孙安健.通用ETL工具的研究与设计[D].上海:复旦大学,2012:1.
- [9] Matt Casters,Roland Bouman,Jos Van Dongen.初建军,曹雪梅,译.Pentaho Kettle解决方案:使用PDI构建开源ETL解决方案[M].北京:电子工业出版社,2011:13-16.
- [10] 田蕊.保险业经营管理系统中ETL的设计与应用[D].北京:中国科学院大学,2017:6.
- [11] Iteye.为什么要用ETL工具?ETL优点.Available at http://wisgood.iteye.com/blog/1499982.
- [12] 孟铭杰,张文彬,任娟,等.四川省地质环境“一张图”系统设计及基本功能[J].安全与环境工程,2017,24(3):16-19.

作者简介:

郑庭明(1974-),男,硕士,高级工程师.研究领域:计算机应用。