

地理信息系统业务数据同步方法设计与实现

张书华¹

摘要: 中国石油已建统建信息系统与各专业分公司建设的信息系统之间, 存在数据存储格式不相同、系统兼容性差、系统间的业务数据同步效率低等问题。针对该问题, 通过设计“数据源端→中间服务器→地理信息系统目标端”的数据同步模式, 提出了中国石油地理信息系统与其他统建系统、地区公司信息系统之间的数据同步方法。规范了各统建系统的动态数据接口, 实现数据同步的历史可追溯性, 为中国石油地理信息系统建设工作提供了可靠数据支撑, 有效地提升跨系统数据同步的标准化程度及工作效率。

关键词: 地理信息系统; 数据同步; 周期同步; 中间库; ETL扩展服务

Doi:10.3969/j.issn.1006-6896.2016.7.013

Design and Operation of Data Synchronization of GIS System

Zhang Shuhua

Abstract: There exists some problems in the data synchronization process between the unified system and the sub-branch system of CNPC, such as different data storage type, low system compatibility, and the low synchronization efficiency among different systems. Aimed at those problems, through the data synchronization mode's design from the data source to the middle server to the GIS target, the paper put up with a new data synchronization method among the CNPC'S GIS system and other unified system or the sub-branch systems. And this method standardizes the dynamic data interface among different systems, which could reach the historic traceability of data synchronization. It could be expected to provide a solid foundation to the CNPC'S GIS system data construction, and improve the standardization and efficiency of the trans-systems data synchronization effectively.

Key words: GIS system; data synchronization; periodic synchronization; middle data base; ETL extended service

根据中国石油天然气集团公司统一的信息技术总体规划, 中国石油建设了一系列信息系统。共有29个与地理信息相关的统建系统, 系统内包括基础地理数据、业务数据以及其他数据三大类数据。目前各统建系统和专业分公司各自有一套存储数据的系统, 系统间互不兼容, 数据存储格式也不相同, 为了有效地破解地理信息系统建设中数据不能共享、更新不及时、数据重复采集和系统重复建设的难题, 形成规范、统一、高效、安全的空间信息共享平台, 实现各系统间的数据同步成为一种必然的要求。通过对三大类数据的深入分析, 以业务数据同步作为首要突破点, 利用 ETL(Extract-Trans-

form-Load) 扩展服务将统建系统和专业分公司信息系统已有的数据成果进行空间化和迁移。业务数据同步划分为空间表和非空间表, 同步顺序需要遵从“图先行, 属随后”的迁移原则, 由地理信息系统提供空间数据, 统建系统提供实体对象分类的最新实时属性信息, 实现为总部职能部门、八个专业分公司及所属地区公司提供业务导航、空间操作、空间分析、三维展示、专题应用等地理信息服务。

1 数据同步方案设计

数据同步方案的设计主要针对中国石油地理信息系统的两大类业务数据: 油田业务数据和管道业务数据。数据存储提供两种方式: ①中间库的方式

¹大庆油田工程有限公司

(用户程序推送数据到中间库中进行临时数据的存储)；②文件格式的存储(该存储方式仅限于 excel 格式和 Geodatabase-MDB/GDB 格式)。业务数据分类见表 1。

1.1 总体架构

数据同步方案的设计遵循更新的高效性、流程的稳定性、数据的保密性、时间的有限性、业务关系的一致性原则，综合考虑数据量、业务复杂程

表 1 业务数据分类

业务数据分类	包含项	具体说明
油田业务数据	井	采油井、气井、注水井、注入井、水源井
	管线	油、气、水等各类管道
	电力线	35 kV 以上输电线路及 10 kV 以下配电线路
	间	计量间、集油阀组间、气烃计量间、气烃阀组间、配气间、配水间、注配间
	站(库、所)	转油站、转油放水站、脱水站、原稳站、总外输计量站、油库、调压计量站、供气站、集气站、气井集气站、注气站、深冷站、浅冷站、甘醇脱水站、二氧化碳液化站、轻烃储库、注水站、注入站、配制站、配注站、曝氧站、水质站、注水站、深度污水站、排涝站、变电所、配电所、发电站等
	道路	道路、涵洞、桥梁
管道业务数据		管道中心线、侵害、阴极保护、检测、设施、运行、事件支持等七大类

度等因素，以确保用户递交的新数据到系统前端 GIS 服务应用的效率。

数据同步方案采用“数据源端→中间服务器→地理信息系统目标端”的模式。先由数据源端(各统建系统或地区公司)将待更新数据同步到中间服务器(可以是 FTP 地址或数据库方式)中，再由地理信息系统设计的接口 ETL 程序自动同步到系统中。同步流程的运行有自动运行和人工运行两种模式，当自动运行失败时，用户可以通过人工运行的模式重新进行数据同步的操作。

统建系统或地区公司通过不同的方式：Web 发布服务、代理服务器、数据分发服务器、中转数据库及数据文件来提交待同步的数据。针对不同的提交方式，中间服务器提供了三种模式来进行数据的接收：①通过中间库(数据库为 Oracle11g)的方式；②通过 FTP 的模式(主要是针对可交换的数据文件)；③通过 ETL 扩展服务接口(FMESchedule 的方式)来进行数据的提取^[1]。同步总体架构见图 1。

由图 1 可知，各统建系统拥有各自独有的用户

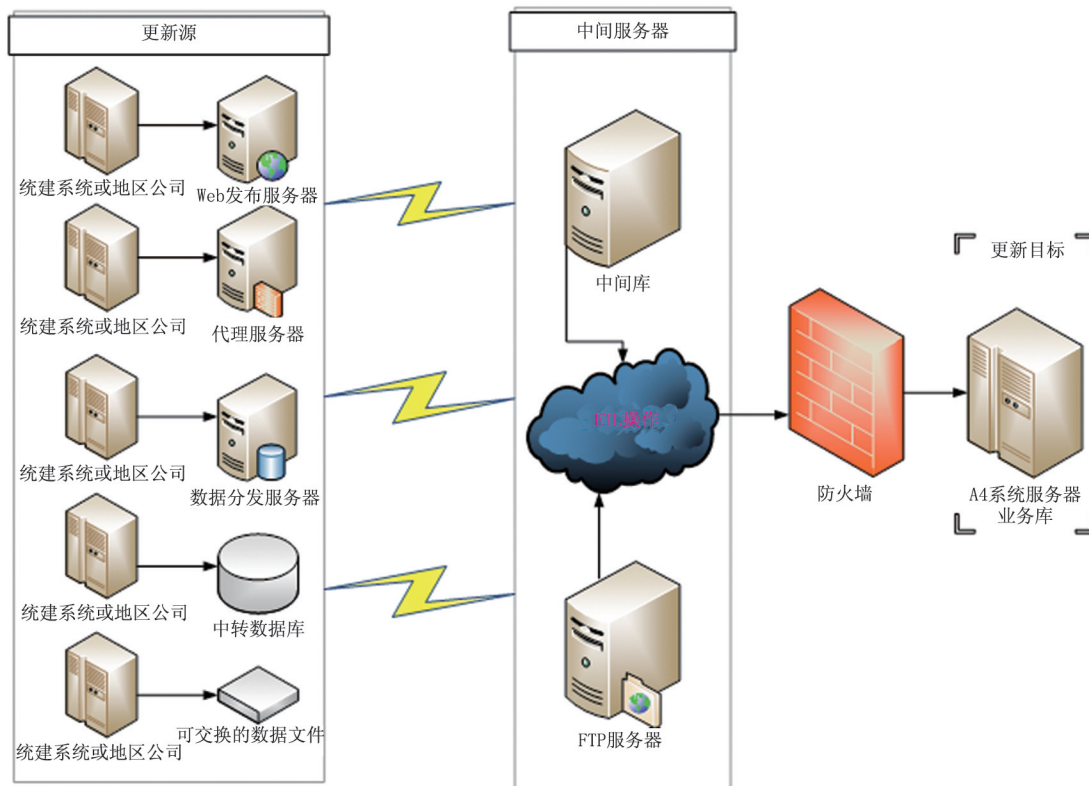
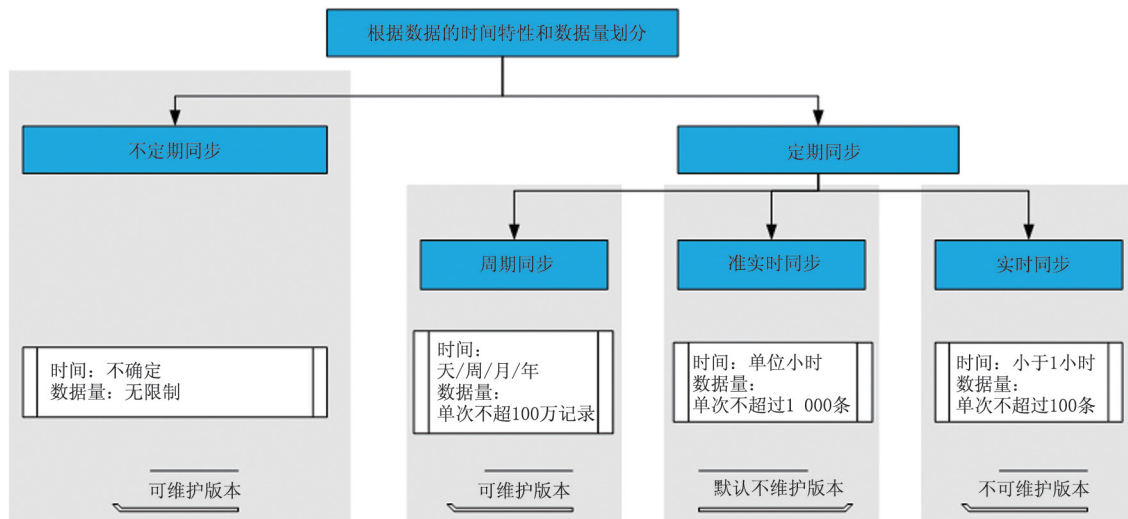


图 1 同步总体架构



来进行访问，并通过中间库中相应的用户来推送各自的数据，但不能更新和删除已推送的数据。

1.2 同步流程

由数据的业务特点可以分为两大类：定期同步和不定期同步。主体结构按照同步的数据量和时间特性又进行了不同级别的分类，具体的结构划分如图2所示。

以地面工程典型业务数据为例，数据同步共分3步，其中1步出现错误，则会提供手动执行的接口。具体实现步骤如下：

(1) 组织机构的同步。同步过程中对数据的主要操作为增加和修改，同步的关键字段为 FULL-

NAME 字段。组织机构同步示意图如图3所示。

(2) 以业务主键为关联字段，实现统建系统到中间库的数据迁移遍历统建系统库中的基础信息属性表，过滤出每张表的“审核通过”的记录，查看“更新时间”和“创建时间”字段，若其中一个为当天时间，则进入同步流程。将过滤结果和中间库中的对应表通过业务主键进行关联，如有关联记录则进行属性记录 Update 操作；若无关联记录，则作为新增记录插入到中间库中，在更新过程中遵循“图先行，属随后”的基本原则。获取更新状态为“删除”，更新日期为当前日期的要素，根据业务主键或者联合主键对中间库进行关联并执行

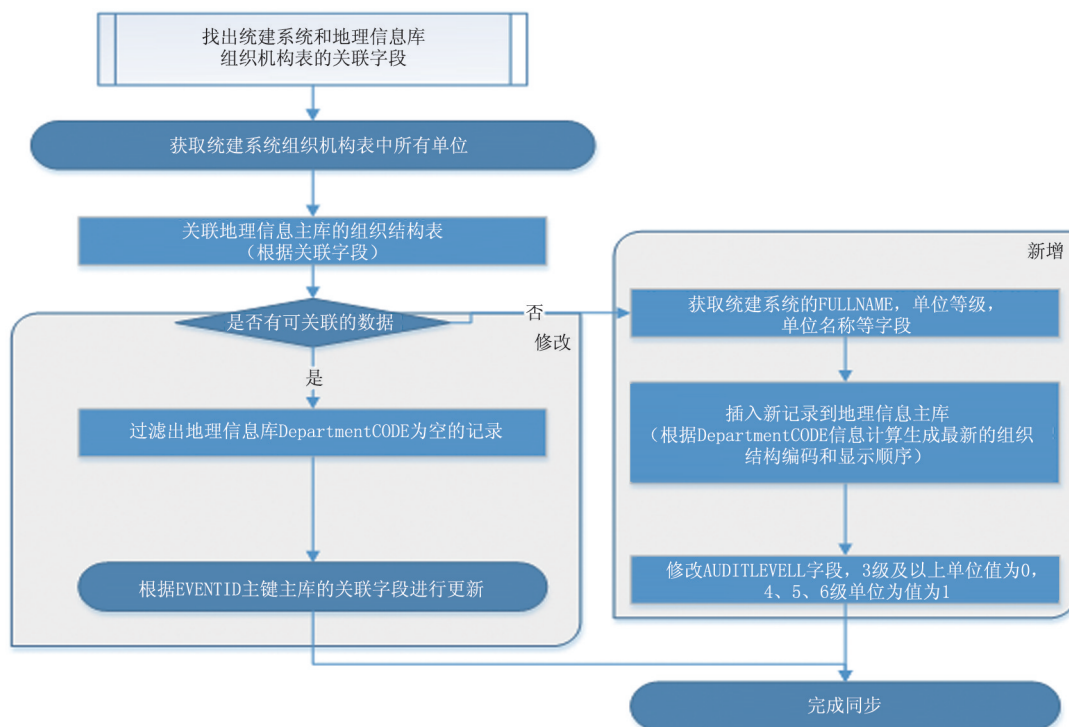


图3 组织机构同步示意图

“标记删除”操作，即将更新状态字段标为“待删除”^[2]。

(3) 实现中间库到地理信息系统库的数据迁移^[3]。遍历中间库中的信息属性表，过滤图形审核级别为通过，且图形审核状态为正常的要素，然后与地理信息系统主库进行关联，关联字段优先采用业务主键，当业务主键为空时，使用联合主键进行关联，若关联到要素，则对地理信息系统库的数据表进行 Update 修改记录的操作；若未关联上，则执行 Insert 插入记录的操作，获取更新状态为“删除”，更新日期为当前日期的要素，根据业务主键或联合主键对地理信息系统主库进行关联并执行“删除”操作。

2 数据同步应用效果

(1) 实时、准确地展示各统建系统对象空间变化情况。保证地理信息系统与各统建系统之间信息的动态同步，为其他统建系统提供及时、准确空间位置服务，为各个统建系统之间的空间信息共享提供了一个公共展示平台。

(2) 辅助完成业务数据的空间分析服务。通过将某生产环境中生成的一些信息进行整合、分析，使其可以借助地理信息系统提供的 GIS 服务获得空间上的直观表达，并在一定程度上可以实现空间分析的功能（例如潜力井、异常井分布情况）。

(3) 数据同步的同时维护数据的历史可溯性。实现现势数据与历史数据的版本自动更替，为地理信息系统重要业务信息的历史回溯功能提供数据支撑。

(4) 保证跨系统数据同步工作流程的标准化管

理。各统建项目或地区公司与地理信息系统进行同步的维护操作包括增删改流程、业务对象结构维护，各系统独立维护的同时，通过同步流程实现不同统建系统之间业务数据同步工作的规范化。

3 结语

业务数据同步方法已经成功应用到中国石油地理信息系统项目建设中，并通过统一的数据标准和软件平台，为总部、各专业分公司及所属地区公司的非统建信息系统提供数据同步服务，实现了总部、各专业公司空间信息资源的整合与共享，为统建系统的数据交换、服务调用和功能开发提供了数据支持。

参考文献

- [1] 杨德勇. 远程数据同步的设计与实现[J]. 信息与电脑(理论版), 2012(3): 112-113.
- [2] 郭哲峰. 浅析分布式数据库同步技术理论[J]. 消费导刊, 2009(11): 49-50.
- [3] 邢晨. Kettle 3.1 数据同步研究[J]. 软件导刊, 2013(6): 67-68.

作者简介

张书华：工程师，2004年毕业于西安科技大学地理信息系统专业，主要从事石油行业地理信息系统建设方面的工作，13936850829, zhangsh_dod@petrochina.com.cn, 黑龙江省大庆市让胡路区西苑街46号大庆油田工程有限公司空间信息技术服务中心，163712。

收稿日期 2015-08-07

(栏目编辑 李娜)

(上接第40页)

- [6] EBERHART R, KENNEDY J. A new optimizer using Particle Swarm theory[C]//Proceeding of the 6th International Symposium on Micro Machine and Human Science, NJ: EEECS, 1995: 39-43.
- [7] 任伟建, 黄晶, 杨有为, 等. 基于改进粒子群算法的油田注水管网优化设计[J]. 科学技术与工程, 2009, 9(11): 2 929-2 933.
- [8] 柯文奇. 基于粒子群算法原油集输管网优化[J]. 石油规划设计, 2010, 21(1): 15-17.
- [9] 蔡昌新. 粒子群算法在油田管网优化中的应用[J]. 长江大学学报(自然科学版), 2010, 7(3): 80-82.

- [10] 乔俊飞, 潘广源, 韩红桂. 基于改进混沌粒子群算法的管网优化[J]. 控制工程, 2013, 20(4): 694-698.

作者简介

孙涛：中国石油大学（华东）储运与建筑工程学院在读博士生，中国石油大学胜利学院讲师，主要从事系统工程优化及优化算法研究，15615065776, 55757910@qq.com, 山东省东营市北二路271号，257097。

收稿日期 2015-11-26

(栏目编辑 李娜)