

何林, 范涛, 曹波. 区域自动气象站数据库整合设计与实现 [J]. 陕西气象, 2014 (4): 44-46.

文章编号: 1006-4354 (2014) 04-0044-03

区域自动气象站数据库整合设计与实现

何 林¹, 范 涛², 曹 波¹

(1. 陕西省气象信息中心, 西安 710014; 2. 渭南市气象局, 陕西渭南 714000)

摘 要: 为了有效解决省内各地市气象局安装多个厂家的区域自动站后, 因各厂家的数据库差异性导致的数据不兼容问题, 从多源异构数据融合的思路出发, 提出一种区域自动气象站数据库整合的设计方案, 并利用 .net 平台编程技术进行实现。根据渭南市气象局区域站的实际情况定制开发软件, 并进行应用测试, 验证了软件的可用性。

关键词: 区域站; 数据库整合; C#; 客户端软件

中图分类号: P409

文献标识码: B

截止 2014 年, 陕西省的区域自动气象站 (以下简称区域站) 共计建站 1 800 余个, 这些遍布全省各地市、县的区域站设备多为华云、长春、无锡三个厂家生产。由于各区域站引进的设备厂家和型号参差不齐, 并且各厂家配备自己的中心站软件, 生成的数据相互独立, 难以兼容, 导致存储数据库的不统一。主要表现在数据库平台和表结构的不同, 这些都加大了对原始观测数据监控和管理的难度^[1]。研究一个既能较好实现不同类型区域站生成数据的融合, 又对当前观测业务和观测数据使用习惯影响较小的数据库整合方案, 并面向地市气象局的实际应用需求开发相关软件产品, 为区域站数据的运维和管理带来方便。

1 软件设计

1.1 设计思路

各种区域站原始数据的存储平台和表结构, 甚至表的数目都有较大差异, 属于典型的多源异构数据。因此在明确常见区域站数据源特点的基础上, 根据数据特点, 选择数据整合的源表和目標表 (源表即各区域站生成的原始数据相关表, 而目标表则需根据实际需求创建一个新表), 特

别的, 当源表的各字段之间具有包含 (子集) 关系时, 则亦可选择一个源表作为目标表。然后, 根据组合主键提取不同类型区域站的源数据存入适当的临时存储结构, 最后将临时存储空间中的数据统一导入目标表^[2], 从而完成数据的整合。

1.2 设计原则与总体架构

遵循简单易操作、可扩展、可维护、整合效率高等原则, 采用 C# 语言结合 VS 集成开发环境将软件设计为 winForm 窗体形式的可视化操作界面, 以客户端的形式呈现给用户。作为第三方软件, 将整合后的数据库作为提供给其他应用的接口, 因此整个软件架构可分为三层^[3-4], 即表现层 (客户端)、业务逻辑层、数据访问层 (数据库服务器端)。
①表现层 提供交互式的操作界面, 主要用于接收用户输入的操作数据并做合法性验证。由于本软件的主要目标是进行后台数据整合, 因此表现层并没有提供数据展示和检索^[5]的功能。
②业务逻辑层 负责处理和转发由表现层获取的用户输入数据, 并选用合适的临时存储结构对数据访问层获取的区域站原始数据进行存储。
③数据访问层 数据库的相关操作集中在该层, 包括从多个源数据表中提取原始数据并

收稿日期: 2014-01-07

作者简介: 何林 (1987—), 男, 陕西武功人, 硕士, 工程师, 从事气象信息软件开发、数据库设计等。

基金项目: 陕西省气象局科技创新基金项目 (2014M-28)

提交给业务逻辑层，以及将业务逻辑层处理加工后的数据导入目标数据表中（见图 1）。

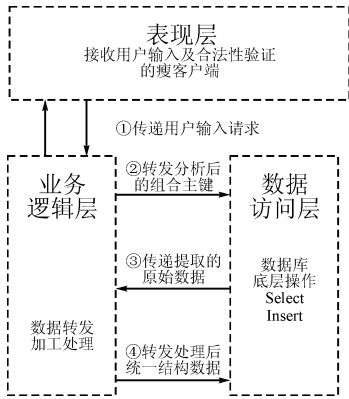


图 1 软件总体架构图

1.3 功能模块设计

综合设计原则和用户需求，将软件分为四大功能模块：对于历史数据和遗漏数据进行操作的手动入库模块，对于日常实时数据进行操作的自动入库模块，便于和其他平台进行交互的多平台数据入库模块以及入库过程中产生的日志查看模块，各模块子功能如图 2 所示。

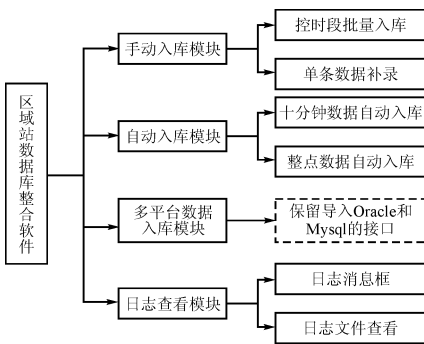


图 2 软件功能模块图

2 关键技术研究 with 实现

2.1 数据库连接与预处理技术

预处理主要解决从不同数据表提取源数据后所做的基本处理，如剔除无用字段、空字段数据的表示等方面，这与 C# 数据库连接操作紧密相关。对于不同的数据库平台，C# 的基本操作流程是一致的，其连接和预处理的伪代码如下：

```
<using 引用对应数据库平台的名字空间>
数据库连接公用方法名
{
```

string dbstr = “数据库服务器 IP，数据库名称，用户名，密码等”；

使用 dbstr 创建一个数据库连接对象 conn；
conn.open ()；打开连接

string sqlstr = “数据表操作语句”；

对数据进行预处理并存入统一的临时存储结构（选用 C# 中的 DataSet）；

conn.close ()；关闭连接

返回值为临时存储结构变量名；

}

2.2 表选取与字段映射

连接操作结束后的首要问题是选择合适的数据库源表和目标表，源表即提取不同区域站原始数据的表，目标表是将预处理后临时存储结构 DataSet 中的数据导入的表。目标表需要综合不同区域站型号和个数、中心站及客户端软件使用情况、源表数目、开发简易度等多重因素进行选取。以渭南市局为例，三类不同厂家的区域站共 161 个，涉及包括正点数据和十分钟数据在内的源表共 26 张，建立 1 张通用站号表，并将目标表选为无锡无线电科学研究所 ZQZ-RG 型区域站的 2 张数据表。

使用 sql 语句从源表中提取原始数据后，将源表和目标表中的对应字段进行映射，同时对不同表的数据类型进行转换。例如，华云区域站正点数据源表中“最大风速出现时间”字段命名为 AFMXT，类型为 datetime，而目标表映射字段名为 WsMaxTime，类型为 char (4)，因此需要对数据进行处理后再导入，主要通过 C# 字符串操作即可实现。

2.3 重复数据验证技术

一般来说，区域站生成的原始数据使用源表的联合主键“站号+观测时间”即可提取，但是当数据导入时，需检验该条记录是否已存在于目标表中，即对导入目标表中的重复数据进行唯一性验证，其基本流程逻辑如图 3 所示。

2.4 日志技术

日志分为操作信息和异常信息两部分，以客户端 textbox 文本框和日志文件两种形式呈现。编写一个专门的日志写入方法，伪代码如下：

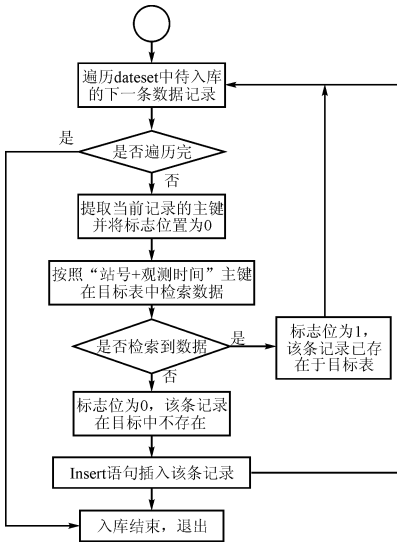


图3 重复数据验证流程图

方法名 write_to_log (消息字符串 str, 标志位 flag)

```

{
while (消息框 text_log 长度在 MAX_LEN 以内)
{
text_log 追加字符串 str;
if 消息框超过最大长度
then 将 text_log 清空;
else break;
end if
}
if 标志位为真
then 将消息追加至日志文件
{
创建字符串 filepath (包含“日志文件名
log.txt 及绝对路径”);
创建文件流对象 sw;
使用文件流写入消息;
关闭文件流;
}
end if
}

```

2.5 批量导入“假死”问题的解决

第一次使用批量导入功能时, 由于历史数据

量非常大, 使用客户端将导致进程未响应的“假死”现象, 使得导入的时效和性能降低。为了解决这个问题, 利用 python 语言编写可执行脚本嵌入到软件中, 从而提高批量导入的效率。

3 应用测试

区域站数据库的整合选取了渭南市气象局进行应用测试。根据实际需求, 对区域站设备使用现状进行分析, 定制研发了符合渭南市气象局站点信息特点的整合软件。经过长期测试, 软件运行状况良好, 已正式投入业务使用。自动入库时, 实时正点与十分钟加密数据共约 900 条记录的导入仅耗时 17 s, 效率较高。同时, 将整合后的数据库与无锡区域站客户端软件结合起来进行数据查询, 对业务人员操作习惯影响较小, 有效解决了区域站数据应用问题。

4 结语

虽然区域站数据库整合的设计并不是一个通用的方案, 需要根据不同地市气象局的实际情况和需求进行个性化定制和处理, 但是整个软件设计与开发的思路却是统一的, 都是从多源异构数据融合的思路出发, 以 C# 语言对技术细节进行实现。因此, 本文提出的软件设计和实现方法是切实可行的, 并且实际测试和试运行也证明了这一点。但是, 该整合方案及软件仅涉及数据入库功能, 未提供区域站数据的展示和检索等具体应用, 还需要在后续开发中逐步加入, 从而真正降低对区域站数据运维、管理及使用的复杂性。

参考文献:

- [1] 曲鹏飞, 刘钧. 区域自动气象站统一数据收集平台的设计 [J]. 成都信息工程学院学报, 2011, 26 (2): 223-227.
- [2] 张靖, 程希来. 多源异构数据整合应用研究 [J]. 成都信息工程学院学报, 2009, 24 (3): 254-258.
- [3] 王垒, 燕东渭. 国家下发疑误数据反馈系统设计与实现 [J]. 陕西气象, 2013 (2): 34-36.
- [4] 高宇, 夏巧丽, 杨银见. 省级上行气象信息传输软件的开发 [J]. 陕西气象, 2006 (4): 23-25.
- [5] 王祝先. 区域自动站统一数据收集平台软件研究与实现 [D]. 哈尔滨: 哈尔滨工程大学, 2012.