

全国城市饮用水源地数据库表结构设计

王 维,陆宝宏,谷黄河,王 欢,孟 伟,成丽婷

(河海大学水文与水资源学院 ,江苏 南京 210098)

摘要 表结构设计是数据库逻辑设计的重要组成部分 ,直接影响到数据库的性能。阐述了全国城市饮用水源地数据库表结构的设计步骤 :用户需求调查分析、数据编码、数据库设计约定、表结构规范化设计及表结构优化 ,并重点剖析表结构规范化设计的处理要求。

关键词 城市饮用水源地 ;表结构设计 ;表结构规范化设计

中图分类号 :TV213.4 **文献标识码** :A **文章编号** :1004-693X(2009)S1-0040-03

数据库系统在水利方面应用较广 ,河海大学的张友静等^[1-2]建立了三峡 - 葛洲坝区间水环境决策支持系统及大坝安全管理信息系统 ,全国基础水文数据库已基本建成^[3]。但因全国水源地分布广、信息海量等原因 ,一直未能实现数据库管理。鉴于此 ,在水利部的资助下 ,研制开发了基于 Oracle 的全国城市饮用水源地信息管理及决策支持系统。笔者主要讨论全国城市饮用水源地(以下简称水源地)数据库表结构的设计。

1 表结构设计原则

通常在设计表结构时 ,要考虑满足功能需求、性能需求、扩展需求 ,表结构要尽量标准化 ,还应注意以下几点 :①合理利用字段类型和长度 ;②选取高效的主键和索引 ;③精简表结构 ;④尽可能减少数据冗余 ,确保数据的正确性。

2 水源地库表结构设计步骤

水源地库表结构的设计主要分为 5 个步骤 :用户需求调查分析、数据编码、数据库设计约定、数据结构规范化设计及表结构优化。

2.1 用户需求调查分析

用户需求调查主要是调查用户组织机构情况、明确用户的各种需求及确定新系统的功能和边界 ,以便获得数据库所需数据情况和数据处理要求^[4]。调查了解了用户的需求后 ,需进一步分析和表达用户需求。笔者采用结构分析方法(Structured Analy-

sis ,简称 SA 方法)。SA 方法的基本过程见图 1。

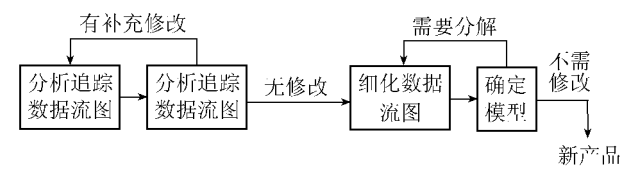


图 1 SA 方法的基本过程

根据用户调查(本次采用专人介绍调查方法)和系统需求分析 ,并参阅文献[5] ,本系统需要水源地管理系统专用的建设基础类数据库、专题类数据库、辅助类数据库、分析类数据库、元数据库等 5 方面数据库 ,为水源地基本信息、水质水量信息等原始数据提供组织、录入、存储、检索、报表、维护等管理功能 ,数据库建设处于初始阶段 ,提供的数据格式为 excel 表结构形式 ,各种数据库结构无相关设计文档 ,需重新设计 ,进行库表结构入库、数据入库检校处理等工作。

2.2 数据编码

本次编码包括水源地编码、水资源分区编码^[6]、河流名称编码^[7]等。对于常用编码 ,采用国家标准 ,如行政区划代码采用国家标准 ,标识符为 ADDVCD ,长度及类型为 (6) ;对于水源地特有编码 ,则参照国家相关规定编制。以水源地编码为例 ,水源地编码参考水利部文件办资源[2005]157 号文中附录 1 城市饮用水水源地编码方法 ,方法如下 :水源地编码标识符为 SRCCD ,由 6 层 14 位定长码组成(14) ,编码原则见图 2。水源地类型通常分为 4 种 :河道型(S)、水库型(R)、湖泊型(L)、地下水(G)。实际流水号在 1 ~ 50 之间 ,若实际流水号超过 9 ,则依次用

A ~ Z 表示 10 ~ 35 ,a ~ z 表示 36 ~ 61 ,见表 1。

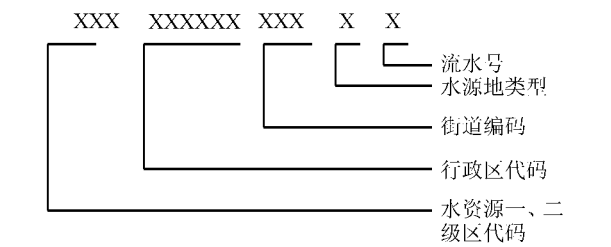


图 2 水源地编码原则

表 1 流水号对照

实际流水号	编码流水号
1-9	1-9
10-35	A-Z
36-61	a-z

2.3 数据库设计约定

每个表结构描述的内容应包括中文表名、表主题、表标识、表编号、表体和字段存储内容规定 6 个部分^[8]。表标识是中文表名英译的缩写 ,在进行数据库建设时用作数据库的表名 ,其编写格式如下 : SRC_α_β。

SRC 为固定前缀(表示全国饮水安全项目);α 为表标识的主体标识 ,一般用英文缩写 ,其长度不宜超过 8 个字符 ;β 为用来标识表的分类 ,表分类按模块来划分 ,其具体划分如表 2 所示。

表 2 标识分类(β)意义

(β) 分类标识	标识意义
B	基本类表后缀
M	业务管理类表后缀
O	专题信息类表后缀
A	分析类数据库表后缀
D	元数据后缀

水源地库表中的字段标识长度不宜超过 10 个字符 ,字段类型包括字符(定长字符 C(d) 和变长字符 VCHAR(d))、数值(N(D[,d]))、时间(T ,采用公元纪年的北京时间)和二进制(B)4 种类型^[9]。水源地库表常用字段标识见表 3。

表 3 水源地库表常用字段标识

字段名	标识符	类型及长度
水源地代码	SRCCD	C(14)
行政区划代码	ADDVCD	C(6)
河流代码	RVCD	C(8)
水资源分区代码	WRRCD	C(7)
投产时间	PRDTM	T

2.4 水源地数据结构的规范化设计

关系数据库中的关系应满足一定的要求 ,而满足不同程度的要求称为不同范式^[10]。第一范式(1NF)是对关系模式的基本要求。在第一范式的基础上进一步满足更多要求的称为第二范式(2NF) ,其余范式以次类推。实践证明以第三范式作为目标

进行规范化设计可以达到比较满意的效果。

第一范式(1NF):关系模式 R 的所有属性都是基本的、不可分的。

第二范式(2NF):关系模式 R∈1NF 且每一个非主属性完全函数依赖于主属性。

第三范式(3NF):关系模式 R∈2NF 且在各非关键字之间不存在函数依赖关系。

根据 1NF 2NF 3NF 定义 ,对水源地数据结构进行规范化处理 :

a. 变换表结构使其满足 1NF。1NF 中的每个属性值都是不可再分的最小数据单位 ,即要求把原始表中的组合数据项化为简单数据项 ,字段重复值化为唯一值^[11]。表 4 为 XX 市地表水饮用水源地水质数据表部分内容 ,该表中基本项目为组合字段 ,水温、PH 记载了重复字段值 ,按照第一范式变换得到表 5。

表 4 XX 市地表水饮用水源地水质数据

城市代码	水源地编号	时间	基本项目	
			水温	pH
XXXXXX	00001	全年	18.2	8
		汛期	24.6	8
		非汛期	11.7	8

表 5 XX 市表水饮用水源地水质数据(1NF)

城市代码	水源地编号	时间	水温	pH
XXXXXX	00001	全年	18.2	8
XXXXXX	00001	汛期	24.6	8
XXXXXX	00001	非汛期	11.7	8

b. 变换表结构满足 2NF。满足 2NF 需满足 :①满足 1NF ;②表中的每个实例或行必须可以被唯一地区分 ;③实体的属性完全依赖于主关键字。现以表 5 为例按 2NF 要求调整表结构。表 5 中 ,城市代码、水源地编号、时间三个字段是查询的基本要素 ,作为联合关键字 ,这样表中的其他非关键字段与关键字段成为完全函数关系。原数据的时间字段值为汉字 ,在计算机中不识别 ,应首先把汉字转换成代码。表 6 为时间代码。表 7 为满足 2NF 的表结构。

表 6 时间代码

代码	时间
1	全年
2	汛期
3	非汛期

表 7 XX 省城市地表水饮用水源地水质数据(2NF)

城市代码 (key)	水源地编号 (key)	时间 (key)	水温	pH
XXXXXX	00001	1	18.2	8
XXXXXX	00002	2	24.6	8
XXXXXX	00003	3	11.7	8

c. 变换表结构满足 3NF。满足 3NF 需满足 :

①满足 2NF ,②非关键字的属性传递不依赖于关键字属性。表 6 为 XX 城市公共水源供水量表 ,该表满足第二范式 ,但非关键字段之间存在函数依赖关系(水库→地表水总量 ,湖泊→地表水总量 ,河道→地表水总量) ,将表 8 分解为表 9、表 10 可消除传递依赖性影响。

表 8 XX 省城市公共水源供水量(2NF)				
城市代码(key)	水库	湖泊	河道	地表水总量
XXXXXX			53 822	53 822
XXXXXX	786			786
XXXXXX	915			915

表 9 XX 省城市公共水源供水量 Ⅰ(3NF)			
城市代码(key)	水库	湖泊	河道
XXXXXX			53 822
XXXXXX	786		
XXXXXX	915		

表 10 XX 省城市公共水源供水量 Ⅱ(3NF)			
水库	湖泊	河道	地表水总量
		53 822	53 822
786			786
915			915

2.5 表结构优化

在水源地表结构设计中 ,表结构优化主要从以下两方面着手。

a. 水源地信息海量 ,内容多样 ,过于偏重较高的规范化等级有时会导致关系表支离破碎、数目过多 ,为了提高查询速度 ,方便查询检索 ,必要时适当地降低某些表的范式等级。水源地库表结构的很多精度都只满足 2NF。

b. 冗余字段可以避免多次关联的查询 ,在缓存不利的情况下 ,冗余字段是提升性能行之有效的方法 ,在水源地表结构设计中适当采用了冗余字段。

表 11 为设计所得的水源地基本信息表(部分内容)。

表 11 水源地基本信息							
表格名称 水源地基本信息表							
表标识 SRC_SRCINF_B							
说明 描述水源地的基本信息							
序号	字段名	标识符	类型及长度	有无空值	计量单位	主键	索引序号
1	水源地代码	SRCD	α(14)	N		Y	1
2	水源地名称	SRCM	α(60)	N			
3	行政区代码	ADDD	α(6)	N		Y	2

3 结 语

作为我国第一个用于全国饮用水源地信息管理

的数据库 ,水源地库表结构实用意义重大 ,设计中可借鉴的资料甚少。该表结构设计以《水质数据库表结构标准》等标准为依据 ,在不使关系表支离破碎的情况下尽量满足范式等级 ,有较好的使用性 ,数据冗余较小 ,便于插入和更新 ,在数据库优化方面采用分离表和索引的方法 ,将表的数据文件和索引的数据文件分开存放 ,以便提高查询速度。

参考文献：

[1] 张友静 ,张元教 ,姚琪 .三峡—葛洲坝区间水环境决策支持系统的开发[J].河海大学学报 :自然科学版 ,2001 ,29 (2) 87-90.

[2] 万定生 ,赵斌 .大坝安全管理信息系统构成初探[J]. 水利水电科技进展 ,1996 ,16(6) 56-61.

[3] 葛维亚 .我国水文数据库建设回顾[J]. 水资源研究 ,2007 ,28(2) 41-43.

[4] 吴兵 ,赵志强 .数据库规范设计[J]. 中国医学装备 ,2006 ,3(10) 43-46.

[5] 朱党生 ,郝芳华 ,张建永 ,等 .中国城市饮用水安全保障方略[M].北京 :科学出版社 ,2008 21-38.

[6] 中国水利水电规划设计总院 .全国水资源分区[S].

[7] SL249—1999 ,中国河流名称代码[S].

[8] SL380—2007 水资源监控管理数据库表结构及标识标准[S].

[9] SL325—2005 水质数据库表结构与标识符规定[S].

[10] 萨师煊 ,王珊 .数据库系统概论[M].北京 :高等教育出版社 ,2000 :171-182.

[11] 葛维亚 .关系型水文数据库表结构设计方法探讨[J].人民长江 ,1991 ,22(6) :15-23.

(收稿日期 2009-04-23 编辑 徐娟)

(上接第 39 页)

参考文献：

[1] 河北省水文水资源勘测局 .唐山市乐亭新区供水工程水资源论证报告[R].石家庄 :河北省水文水资源勘测局 ,2009.

[2] 河北省水文水资源勘测局 .唐山沿海“ 四点一带 ”地区总体规划水资源论证报告[R].石家庄 :河北省水文水资源勘测局 ,2008.

[3] 中国水利水电科学研究院 .唐山市沿海四点一带水资源安全保障战略研究[R].北京 :中国水利水电科学研究院 ,2008.

(收稿日期 2009-07-27 编辑 高渭文)