

文章编号: 1673-9469(2012)01-0090-05

## 基于数据仓库的水利工程信息管理系统设计

黄伟建<sup>1</sup>, 徐学钢<sup>1</sup>, 王子轩<sup>2</sup>

(1. 河北工程大学 信息与电气工程学院 河北 邯郸 056038; 2. 北京师范大学 艺术与传媒学院 北京 100875)

**摘要:** 针对我国当前业务型的水利信息管理系统难以获取隐含信息的问题, 提出基于数据仓库技术和 OLAP 技术设计开发水利工程信息管理系统的解决方案。系统使用 Java 作为开发语言, 后台数据库采用 SQL Server 2005 系统, 使用 SQL Server 中的数据仓库工具建立数据仓库。根据需求分析, 系统确定获取渠道、科室和资金 3 个主题的隐含信息, 并采用查询效率较高的星型模型作为概念模型, 通过事实表设计和维度表设计来完成 3 个主题设计的具体实现。

**关键词:** 数据仓库; OLAP; 水利工程; 信息管理系统

中图分类号: TV698

文献标识码: A

### Design of the hydraulic engineering information management system based on the data warehouse

HUANG Wei-jian<sup>1</sup>, XU Xue-gang<sup>1</sup>, WANG Zhi-xuan<sup>2</sup>

(1. School of Information Science and Electrical Engineering, Hebei University of Engineering, Hebei Handan 056038, China; 2. School of the Art and Communication, Beijing Normal University, Beijing 100875, China)

**Abstract:** In view of the problem that current business-oriented water information management system obtain difficultly hidden information problem, the hydraulic engineering information management system was design based on data warehouse and OLAP technology. The Java served as the development language, SQL Server 2005 as back-end database, and the SQL Server data warehouse tools was used to build the data warehouse. According to requirements analysis, the access channels, departments and funds were confirmed as the three implicit information, which were implemented through the fact table and dimension table design with the efficient star query model as a conceptual model.

**Key words:** data warehouse; OLAP; hydraulic engineering; information management system

二十一世纪以来,我国在水利信息化建设的进程中,完成了大量的水利信息管理系统,这对于我国水利工程的投资建设和运行管理都起到了良好的促进作用。随着水利信息化的发展,水利信息系统规模越来越大,对系统运行质量要求越来越高,任务越来越重,管理越来越复杂<sup>[1]</sup>。现有的水利信息管理系统所提供的功能大多是业务型的,只能处理直接的水利数据,难以获取这些数据信息背后的隐含信息,如文献[2]开发的东海县水利工程管理系统,仅仅按照业务需要,把系统划分为农田水利、工程管理、防汛抗旱、自然环境、水利

史和大事记等业务模块;而文献[3]开发的水利工程科技档案管理信息系统,仅能提供用户管理、单位管理、档案管理和管理员管理等管理信息。而这些直接数据信息背后的隐含信息,如各项水利项目的建设趋势、各个承担任务科室的进展趋势、各类项目资金的完成趋势等均无法体现。

本文基于数据仓库技术和联机分析处理(On-Line Analytical Processing, OLAP)技术,设计水利工程信息管理系统。系统按照渠道、科室和资金 3 个主题设计建立数据仓库,以克服传统的水利信息管理系统难以获取这些隐含信息的缺陷,

收稿日期: 2011-10-13

基金项目: 邯郸市科学技术研究与发展计划项目(1134201103)

作者简介: 黄伟建(1964-),男,山西交口人,博士,教授,从事信息管理系统开发与设计方面的研究。

为水利工程决策人员提供高效的信息服务。

### 1 关键技术

#### 1.1 数据仓库

数据仓库技术是二十世纪末在计算机领域中迅速发展起来的一种新技术。数据仓库是面向主题的、集成的、稳定的、随时间变化的数据集合,用以支持经营管理中的决策制定过程。从数据库基础上发展起来的数据仓库技术,从本质上讲是一种新的信息集成技术。数据仓库中存储着海量的当前数据、历史数据和综合数据。这些数据往往是多个数据源的数据集成,并依照主题进行重组。数据仓库中的数据一般不再修改。这些数据虽然也是随时间变化而变化的,但时限比操作型数据的时限要长得多。数据仓库在功能上与操作型的数据库最主要的区别是建立决策支持系统,它可有效地实现对水利工程信息资源的管理和利用,为水利工程管理者的决策分析提供准确一致的量化信息<sup>[4-5]</sup>。

#### 1.2 OLAP 技术

OLAP 技术是基于数据仓库的、针对主题的、能快速进行联机数据访问和分析的软件技术。它对数据信息进行多种不同形式地观察,使用户可以从不同的维度、不同的侧面和不同的数据综合程度来观察数据,从而获得数据背后隐含的新的有价值的信息。OLAP 分析工具是多维数据分析工具的集合<sup>[6]</sup>,而水利工程的数据信息一般是多维的,因此多维数据就是水利工程决策的主要内容。OLAP 技术侧重于对水利局领导的决策支持,以便他们能准确掌握水利工程的状况,制定出正确的方案和措施。OLAP 先将数据仓库当中的数据组织成多维数据立方体后,再通过“数据切片”、“数据切块”和“数据钻取”等操作方法对立方体进行分析操作。

(1) “数据切片”是从多维数据立方体上的某一维度上选定一个值后,把多维数据从  $n$  维降为  $n-1$  维。例如,在“桥梁工程”的多维数据立方体中,时间维度取“2009 年”,则得到关于 2009 年所建设的所有桥梁工程数据信息的一个数据切片。

(2) “数据切块”是从一个完整的数据立方体中切取一部分数据,从而得到一个新的体积较小的数据立方体。例如,在“桥梁工程”的多维数据

立方体中,时间维度取“2007 ~ 2009 年”,项目维度取“以工代赈”,投资额维度取“100 万元以下”,则会得到关于 2007 ~ 2009 年间的建设的投资额在 100 万元以下所有以工代赈项目的桥梁工程数据信息的一个数据切块。

(3) “数据钻取”是通过改变数据立方体维度的层次,来达到变换分析粒度的目的。数据钻取的方向分为下钻和上钻。下钻是指用户从某一个较高的层次出发,去观察较低层次的细节数据信息。上钻跟下钻正好相反,是指用户从某一个较低的层次出发,去观察较高层次的宏观数据信息。例如在时间维度中,从“年”的层次出发,去观察“月”的数据就是下钻;而“月”的层次出发,去观察“年”的数据就是上钻。

#### 1.3 B/S 架构模式

C/S(客户端/服务器)和 B/S(浏览器/服务器)是两种常用的架构模式<sup>[7]</sup>。传统的 C/S(客户端/服务器)架构模式是两层结构,分为客户端和服务器两部分,通过局域网直接连接。而 B/S(浏览器/服务器)架构模式则是三层结构,为浏览器(相当于 C/S 架构模式中的客户端)、Web 服务器和数据库服务器(相当于 C/S 架构模式中的服务器)3 部分。浏览器与数据库服务器不直接连接,而是通过中间层 Web 服务器进行间接连接,见图 1。其中浏览器与 Web 服务器可以通过局域网或广域网进行连接,能充分利用 Internet,延长连接距离。B/S 架构模式将系统功能实现的所有应用软件都安装在 Web 服务器和数据库服务器上,系统的维护工作也集中在这两处,与客户端的浏览器无关,客户端零维护。系统的改进和升级也可以只在服务器端进行,所以系统的可靠性和稳定性比 C/S 架构模式更高。浏览器、Web 服务器、数据

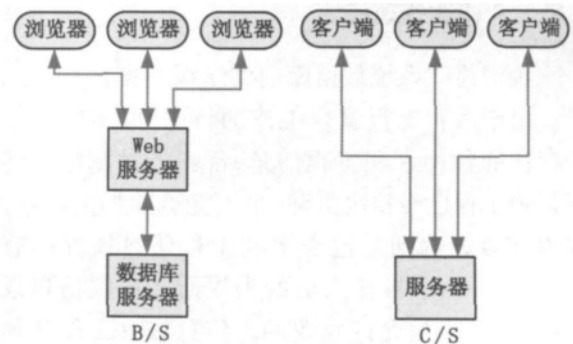


图1 B/S与C/S的区别

Fig.1 The difference between B/S and C/S

库服务器三者为逻辑上的划分<sup>[8]</sup>,这三者也可以只布置在一台计算机上。

## 2 系统的结构设计

水利工程信息管理系统主要包括信息录入管理、信息修改管理、信息查询管理和系统维护管理4个模块(系统结构见图2)。根据具体需求,每个模块都具有各自的功能<sup>[9]</sup>。

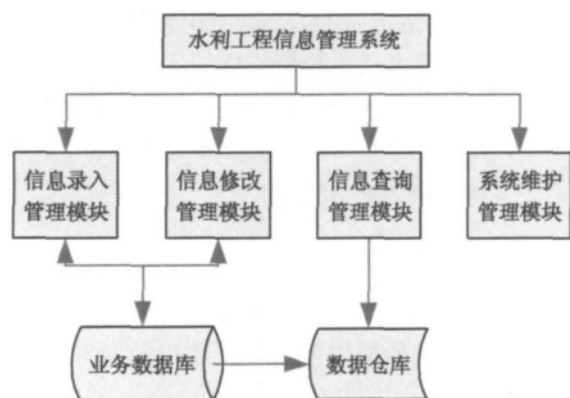


图2 系统结构图

Fig.2 System structure diagram

### 2.1 信息录入管理模块

负责录入水利工程数据信息的工作。由于信息录入人员不一定都精通计算机技术,所以本模块的编制尽量智能化,录入操作尽量简单化。每一次录入时,系统都会通过多次询问来引导录入者输入格式正确、内容全面的数据信息。如果某一项工程不是首次录入,系统还会自动显示前一次录入本工程的数据信息,提示录入者检查本次录入的数据信息是否存在缺项或错误。

### 2.2 信息修改管理模块

负责修改系统数据库内的错误数据信息的工作。由于水利工程管理工作的特殊性,系统内已经存在的各种水利工程信息,都不能被删除。即使某项工程已经报废拆除,也不能删除相关信息,因为这会影响到对过去水利工程建设情况的查询。不过,由于操作人员的失误所造成的信息录入错误,系统是允许修改的。信息修改工作必须谨慎,只能由系统管理员进行操作,其他级别的用户不能进行操作。并且修改工作还不能通过远程网络方式操作。这样才能尽可能保证系统内数据

信息的前后一致性。

### 2.3 信息查询管理模块

负责查询水利工程数据信息的工作。这是管理系统的核心模块。业务数据库中数据信息经过抽取、清洗等处理后,集成在数据仓库内。系统利用 OLAP 技术,将数据组织成多维数据立方体的形式,通过切片、切块、钻取等各种分析操作,使用户能从多角度、多侧面、多层次地查询数据库中的水利工程数据信息。管理系统把一些关键的查询过程和查询结果经过处理后存储下来,并且当系统数据信息有更新时还能自动更新。这样,当以后再进行类似的查询时就不用每次都去搜索整个数据库,而能快速得到正确的查询结果。系统能够根据需要,把每次查询的结果信息自动生成合适的报表,并能实现同上级机关相关软件的接口对接。

### 2.4 系统维护管理模块

负责对用户和数据库的维护进行管理工作。包括对用户进行用户申请、登录申请和用户级别设置等工作,和对数据库进行增加数据库和数据备份等工作。

### 2.5 系统的开发平台

本系统的数据库采用 SQL Server 2005 系统,使用 Java 作为开发语言,基于 Eclipse 开发。SQL Server 是一种关系数据库管理系统,具有使用方便、可伸缩性好、与相关软件集成程度高和可多种平台使用等优点<sup>[10-11]</sup>。在 SQL Server 2005 系统中提供了大量的数据仓库设计、建立和操作等方面的工具。Java 是一种跨平台的、适合于分布式计算机环境的面向对象的编程语言,具有可移植、稳定、简单、高性能和可动态执行等特征。Java 应用程序(Java Applet)存放在 Web 服务器中,应客户端浏览器的请求,传送至客户端,由浏览器去执行。它可以有效地实现 B/S 架构模式的功能。

## 3 数据仓库的设计与实现

数据仓库的构建是一个不断循环、反馈而使系统不断增长与完善的过程<sup>[12]</sup>,它的设计主要包括概念模型设计、逻辑模型设计和物理模型设计3个阶段。

### 3.1 概念模型设计

概念模型设计主要界定系统边界和确定主要的主题域及其内容,它包括需求分析、主题确定和概念模型确定等内容。

#### 3.1.1 需求分析

水利工程的高级管理者,不光需要知道水利工程的具体数据信息,更重要的是需要了解:各个渠道配套工程的建设趋势;各个职能科室承担的工程建设任务的进展趋势;各类项目资金的完成趋势。

#### 3.1.2 主题确定

根据需求分析,确定系统数据仓库的主题为:渠道、科室和资金。

#### 3.1.3 概念模型确定

数据仓库的概念模型常见的有 3 种:星型模型、雪花模型和事实星座模型。星型模型可以采用关系型数据库结构,模型的核心是事实表,围绕事实表的是维度表,通过事实表将各种不同的维度表连接起来,各个维度表都连接到中央事实表<sup>[13]</sup>,见图 3。在该模型中,因为大部分数据都在事实表中,只需搜寻事实表就可进行查询。并且维度表都比较小,能放在高速缓存中,可以快速地与事实表进行连接。星型模型实际上是以“空间”换“时间”,通过适当增加冗余存储空间来显著地提高系统查询速度,所以本系统数据仓库的逻辑模型采用星型模型。

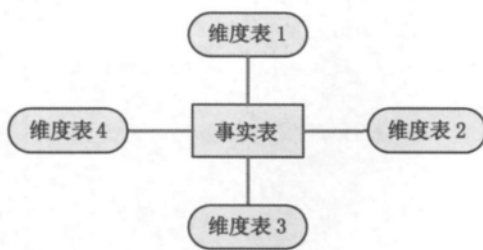


图3 星型模型示意图

Fig. 3 Star type model diagram

### 3.2 逻辑模型设计

逻辑模型是数据仓库从概念模型向物理模型过渡的中间层次模型。逻辑模型设计是把概念模型当中各种概念和实体间的相互关系,进行分解细化,使之能适应计算机系统存储模式。逻辑模型设计主要包括事实表的设计和维度表的设计。

#### 3.2.1 事实表设计

数据仓库中的每个主题都需要通过一个或多个相关的事实表进行具体设计实现。事实表不仅是数据仓库的核心,也是构成数据仓库的所有类型表中体积最大的<sup>[14]</sup>,它包含和决策目标紧密相关的数据信息。事实表的内容一般由两部分组成:键和指标。事实表使用键把各个维度表联系起来,指标用于记录事实表的详细信息。

不同的主题设计的事实表各有不同,以渠道主题为例设计事实表,它包括渠道工程信息表、配套工程信息表和工程历次建设信息表。

#### 3.2.1 维度表设计

事实表中的内容进一步分解,就得到维度表。维度表是事实表的关联表<sup>[15]</sup>,含有事实表中的特征数据,使用关键字同事实表进行关联。维度表中还包含有多个详细属性,用于说明对象的逻辑关系。在设计维度表时,尽可能简化维度表与事实表之间的联系,尽量减少表之间连接数量,以减轻计算机系统的运算负担。

下面列出渠道主题的配套工程信息事实表的维度表:

建设时间维度表:年、月、日;

主要工程量维度表:土方、石方、混凝土方、钢筋用量;

建设投资组成维度表:上级资金、地方配套、群众自筹。

### 3.3 物理模型设计

物理模型设计即为逻辑模型中的数据确定一个符合要求的物理结构<sup>[16]</sup>,也就是将逻辑模型当中的相关内容转变为在计算机系统可以存储的模式<sup>[17]</sup>。在认真分析系统数据仓库的基础上,研究它的存储结构及方法、数据使用环境及使用方式和数据信息规模等主要参数,完成水利工程信息管理系统数据仓库的物理模型设计。

### 3.4 数据仓库的实现

系统使用 SQL Server 2005 中的数据仓库工具建立数据仓库,使用 JSP 实现数据仓库的接口设计,使用 Analysis Services 工具采集和分析数据信息。数据的采集工作就是进行数据的抽取、转换、清洗和装载的过程,将业务数据库当中的水利工程数据信息加载到数据仓库中,从而完成数据信息从数据源向数据仓库的转化。它主要进行:识别数据源信息,管理元数据,集成不同数据源的数

据格式,清理数据集,数据分割,定期更新和维护数据。

#### 4 结语

本文设计开发的水利工程信息管理系统,利用数据仓库技术确定渠道、科室和资金3个主题,利用OLAP技术从不同的维度、侧面和数据综合程度对数据信息进行多种不同形式地观察,从而达到获取这些直接数据信息背后关于3个主题的隐含信息的目的。如果用户对隐含信息的获取有新的或不同的需求,可以增加或更改主题,并按照新主题设计数据仓库。

#### 参考文献:

- [1] 蔡阳,詹全忠,周继续,等. 水利信息系统运行保障平台关键技术研究[J]. 水利水电技术, 2010, 41(2): 91-93.
- [2] 朱俊昌,高亚楠,郑源,等. 中小型水利工程管理系统的设计与实现[J]. 水电能源科学, 2010, 28(8): 140-142.
- [3] 吴杰,王冬梅. 水利工程科技档案管理信息系统设计与实现[J]. 人民长江, 2007, 38(11): 194-195.
- [4] 郭峻峰,倪志伟,高雅卓,等. 一种提高数据仓库查询效率的有效方法[J]. 计算机集成制造系统, 2009, 15(12): 2451-2457.
- [5] 陈步英,马骅,张小志. 高校财务数据仓库研究[J]. 财会通讯, 2010(11): 119-120.
- [6] 杨云,罗艳霞. 基于Web数据仓库构建GSM网络优化系统[J]. 计算机工程与设计, 2010, 31(12): 2894-2900.
- [7] 王春波,杨大兵. 基于ArcSDE技术的地籍管理系统研究[J]. 河北工程大学学报: 自然科学版, 2010, 27(3): 77-80.
- [8] 蒋薇,赖青贵,陈楠. 基于B/S架构的"神龙一号"数据库系统的实现[J]. 强激光与粒子束, 2010, 22(3): 642-646.
- [9] 张颖,刘惠德,侯旭辉,等. 矿井水文地质信息系统的设计与实现[J]. 河北工程大学学报: 自然科学版, 2008, 25(4): 82-85.
- [10] 胡逢恺,赵刚,程旭. 基于SQL Server数据库的三维模型存取研究与实现[J]. 四川理工学院学报: 自然科学版, 2010, 23(2): 212-215.
- [11] 胡开明,陈建华. 用.NET实现对SQL SERVER数据库安全的动态监控[J]. 四川理工学院学报: 自然科学版, 2010, 23(3): 299-302.
- [12] 黄晓颖,李亚芬,王普. 基于数据仓库的学科建设决策支持系统的设计[J]. 计算机工程与设计, 2010, 31(23): 4995-4998.
- [13] 秦永平,王丽萍. 基于数据仓库的突发公共卫生事件预警预报系统[J]. 计算机工程与设计, 2010, 31(13): 3119-3122.
- [14] 赵宝华,阮文惠. 高校财务数据仓库的设计与实现[J]. 计算机工程, 2008, 34(17): 266-268.
- [15] 王志宏,顾新建. 基于数据仓库的产品族供应链管理系统[J]. 浙江大学学报: 工学版, 2009, 43(7): 1197-1202.
- [16] 李志刚,马刚. 数据仓库与数据挖掘的原理及应用[M]. 北京: 高等教育出版社, 2008.
- [17] 刘永立,王海涛,孙维民,等. 基于基础数据库的煤矿应急救援指挥信息系统[J]. 黑龙江科技学院学报, 2010, 20(1): 44-47.

(责任编辑 马立)