

大坝安全监测数据仓库的设计与构建 ——以小浪底大坝为例

黄红女¹, 华锡生¹, 宋小刚²

(1. 河海大学土木工程学院, 江苏 南京 210098; 2. 武汉大学遥感信息工程学院, 湖北 武汉 430072)

摘要 针对现有的大坝安全监控系统中存在的历史数据量大、难以集成和访问数据能力不足的问题, 研究了大坝安全监测数据仓库的设计和构建方法. 以小浪底大坝安全监测数据仓库设计与构建为例, 设计了数据仓库系统的框架结构, 详细介绍了大坝安全监测数据仓库的设计方法和具体步骤, 着重研究了主题域及其内容的确定、逻辑模型和物理模型的设计、数据集成、数据的访问设计和维护等内容, 最后还分析了基于 DTS 的小浪底大坝安全监测数据仓库的数据集成方法和步骤.

关键词 大坝安全监控; 数据仓库; 数据集成; 小浪底水利枢纽工程

中图分类号: TV698.1

文献标识码: B

文章编号: 1006-764X(2005)04-0056-03

Design and construction of dam safety monitoring data warehouse : with Xiaolangdi Dam taken as an example // HUANG Hong-nü¹, HUA Xi-sheng¹, SONG Xiao-gang² (1. College of Civil Engineering, Hohai University, Nanjing 210098, China; 2. College of Remote Sensing Information Engineering, Wuhan University, Wuhan 430072, China)

Abstract : In consideration of some limitations in current dam safety monitoring system, such as great amount of historical data, difficulties in data integration, and low capacity in data accessing, the method for design and construction of dam safety monitoring data warehouse was studied. A framework was presented for construction of the dam safety monitoring data warehouse system for the Xiaolangdi Project, and the method and process for its design were introduced with emphasis on the determination of main issues and contents, design of logic model and physical model, data integration, and data accessing design and maintenance. Finally, the method and steps of data integration in the data warehouse system based on data transformation service (DTS) were analyzed as well.

Key words : dam safety monitoring; data warehouse; data integration; Xiaolangdi Hydro Project

随着计算机、电子技术的发展, 大坝安全监测的自动化、信息化水平有了很大提高, 表现在存储容量上, 是数据库的规模越来越大; 表现在媒体形式上, 已不再只是单一的文本或数据, 而越来越趋向于多样化, 涉及到文本、图形、图像等各种媒体. 随着大坝安全监测部门对大坝安全评价信息需求的实时性进一步提高(尤其在洪水时期), 从大量数据中提取(检索、查询等)评价安全和制定安全策略的信息就显得越来越重要. 这种需求既要求联机服务, 又涉及大量用于决策分析的数据, 而传统的数据库系统已无法满足这种需求, 具体体现在 3 个方面: 历史数据量很大, 辅助决策信息涉及许多观测项目的数据, 而这些数据来自不同的数据源, 难以集成, 由于访问数据的能力不足, 它对大量数据库的访问性能明显下降. 因此, 研究、引用并建立新型数据库, 对提高大坝安全监测信息分析和决策的效率, 改善现有监测系统的现状, 具有重大的现实意义. 对此一个比较先进

的技术是建立大坝安全监测数据仓库. 本文主要是以一般数据仓库建立的步骤为线索, 结合小浪底大坝安全监测的特点, 经过总体体系结构设计、数据模型选择和确立、数据抽取、清理和转换、数据载入等一系列步骤, 建立起了适合大坝安全监测信息分析和挖掘的大坝安全监测数据仓库.

1 大坝安全监测数据仓库的框架设计

大坝安全监测数据仓库是一种体系结构, 它不仅仅用来存储监测数据, 还要提供数据进级(从实时数据到导出数据)、分析服务、数据挖掘等各种工具. 根据数据仓库体系结构思想, 结合小浪底工程项目的具体特点以及大坝监测数据分析的特点, 给出大坝安全监测数据仓库系统的框架结构如图 1 所示.

图 1 中原安全监控系统中的内容作为源数据, 用来建立大坝安全监测数据仓库, 包括各种关系数据库, 比如生成库、知识库、现实库、模型库、环境量

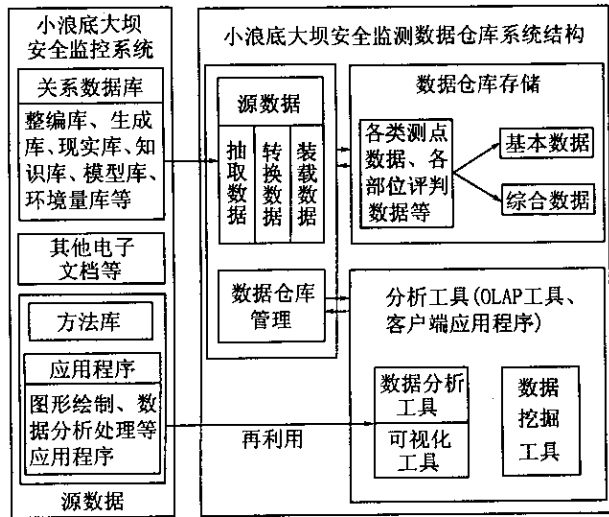


图1 小浪底大坝安全监测数据仓库系统体系结构

库等,还有一些是原系统生成的电子文档形式的结果文件。原系统中编制的一些应用程序也可以为监测数据仓库系统重新使用。

在大坝安全监测数据仓库系统中,主要包括3个部分的内容:数据集成与管理、数据仓库存储和分析工具。数据集成与管理主要是对来自原系统的源数据进行抽取、清理、转换、装载,定义数据的综合、转化过程,使系统自动将数据从源数据中提取出来,转移到监测数据仓库中,并给予维护。数据仓库存储是指各类测点数据和各部位评判数据的存储,数据仓库中存储的监测数据分为基本数据、不同层次综合和分析数据,各部分都按多个决策主题形成多维数据组织形式。

分析工具包括数据发掘工具、数据分析工具、可视化工具等^[1]。数据发掘工具相对于原监控系统来说是一种新的工具,是用户根据监测数据的特点,采用相应的数据挖掘算法编制的一种应用程序,用于发现知识、发现规则、发现新的监测数据关系;数据分析工具用于从源数据中按决策要求的主题形成当前基本数据层,再按综合决策要求形成轻度 and 高度综合决策数据层,然后利用系统分析工具进行各种分析。数据分析工具可以是系统自带的,也可以是用户自己开发的,当然在开发过程中用户还可以直接或间接利用原系统中开发好的一些分析和可视化工具。

2 小浪底大坝安全监测数据仓库的设计

大坝安全监测数据仓库是在现有的监测数据库基础上进行开发的,它着眼于有效地提取、综合、集成和挖掘已有的数据库资源,服务于决策分析的需要。大坝安全监测数据仓库系统开发是一个不断循环、反馈而使系统不断增长与完善的过程,在整个开发过程中自始至终都要求分析人员、开发者以及坝工专家的共同参与和密切合作。结合小浪底大坝现

有的安全监控系统实际情况,大坝安全监测数据仓库的设计包括以下几个步骤:

a. 确定主要的主题域及其内容。这要求系统设计人员经常与监测部门、部门业务人员、分析人员及坝工专家沟通,详细了解分析需求、报表需求、信息需求以及整个大坝的内部结构和如何组织数据、如何反映其运行态势等一系列问题,再归纳成仓库的主题域,并对每个主题域的内容进行明确的描述,如主题域的公共码键、主题域之间的关系、充分代表主题域的属性组。在小浪底大坝的整个运行系统中,数据都是围绕着测点、观测项目、工程部位或整个大坝来组合的。笔者在设计小浪底大坝安全监测数据仓库时,确定了测点和观测项目两个主题。测点是大坝日常运行的基础,大坝或某一工程部位运行的正常与否都是通过测点的监测数据来反映的。观测项目是测点和工程部位之间的桥梁,也就是说测点通过哪种监测量来反映大坝的运行情况。图2是测点和观测项目的主题关系示意图,从图中可见,一方面利用数据仓库对测点进行分析,发现测点之间的隐含关系、测点运行的内在模式以及所反映的观测项目状态如何;另一方面,通过数据仓库对观测项目进行分析,发现同一工程部位观测项目之间的关系以及判断该工程部位的运行情况。同时可以利用发现的模式或关系,对测点、工程部位的运行趋势作出预测,或围绕这两个主题进行其他相关的分析。

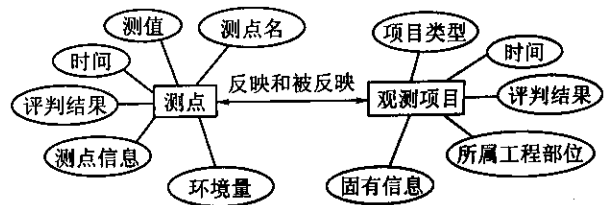


图2 测点和观测项目主题关系示意图

b. 逻辑模型设计。在确定主题后,需要对主题包含的信息进行详细定义,利用维度建模的方法进行数据建模,选择一种数据模型对相关主题的事实表和维表的关系详细定义,使其能够有效地组织数据仓库中的数据。主要工作有:①基于用户的要求,分析主题域,着重于某个主题,设计多维数据结构的事实表和维表。星形结构数据模型已被证实在数据仓库设计中是灵活的、有效的,本文采用星形模式来表达事实表和维表之间的逻辑关系。②确定粒度层次划分。在小浪底大坝安全监测数据仓库中,考虑到大坝安全评判的实时性和评判、报警的级别性,时间维的粒度划分为年、月、日、时;位置维的粒度划分为大坝、工程部位、观测项目、测点,而环境量可以作为一个特殊的维参与数据分析。这里需要注意的一点是,测点本来可

以作为单独的一维去描述测值的特性,但考虑到大坝安全监测数据结构的特点,可以把测点维和位置维合为一维,这样有利于降低数据的冗余度和方便数据的操作。③关系模式的定义。数据仓库的每个主题都是由多个表来实现的,这些表之间依靠主题的公共码键联系在一起,形成一个完整的主题。关系模式的定义就是对选定的当前实施的主题进行模式划分,形成多个表,并确定各个表的关系模式。图3是大坝安全监测数据仓库中用星形模式表示的数据结构示意图。

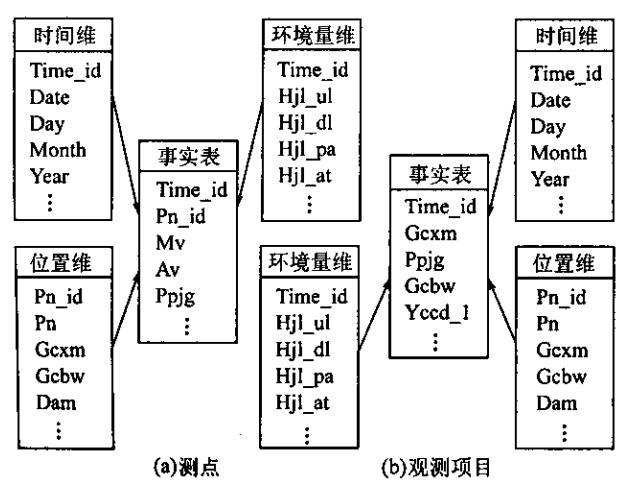


图3 测点和观测项目的事实表和维表的星形结构

c. 物理模型设计。物理模型设计主要考虑数据的存储方式,使得系统有较好的性能。对于记录庞大的事实表,可以考虑分区存放。而记录很少的维表则可以集中存放于某一表空间,甚至可以让数据在首次读取时驻留在系统内存中,以加快数据存取速度。索引的建立也在物理设计中完成。

d. 数据集成。大坝安全监测数据仓库的数据总是来自前台作业系统、已有监测数据库系统以及监测部门以往建立的电子表格、文档等,这些数据并非照搬过来就行,而是要以统一定义的格式从各个系统中抽取出来,经过清洗,再经过数据装载和整理程序集成到数据仓库中。小浪底大坝安全监测数据仓库主要通过 SQL Server 2000 中的数据转换服务(Data Transformation Services, DTS)来帮助管理员和开发人员解决数据集成问题。

e. 数据的访问设计和仓库不间断的维护方案设计。数据的访问设计即数据仓库的使用。根据用户的需求,采用现有数据库产品的 OLAP 和数据挖掘功能,或者是开发设计自己的应用程序,来达到对大坝资料的分析决策和隐含信息的挖掘。另外,数据仓库的运作与传统的作业系统有很大区别,它需要不间断地维护,否则它的性能将越来越差。

f. 编码、测试及实施。在开发过程中,要根据逐步完善的原型法的开发方法,尽快地让系统运行起

来,尽早产生效益,要在系统运行或使用中不断理解需求,改善系统,不断考虑新的需求,完善系统。

经过以上一系列步骤的实施,建立了一个比较完善的大坝安全监测数据仓库系统,为小浪底大坝检测部门提供了一个灵活有效的、实时的、智能的、相对准确的大坝资料分析和信息挖掘平台。

3 基于 DTS 的小浪底大坝安全监测数据仓库的数据集成

SQL Server 2000 中的 DTS 提供了一套图形化工具和可编程对象,以帮助管理员和开发人员解决数据集成问题,其中包括将数据从分散的数据源中提取出来,经过转换,合并到一个或多个目标位置^[2]。另外,管理员还可以将制定的 DTS 解决方案创建为一个或多个 DTS 软件包,安排它定期运行或在某些事件发生后运行,从而减轻监测数据仓库管理员的工作强度。

小浪底大坝的监测资料大部分都通过原有系统的分析和处理,被存储在 Sybase 数据库中,还有一部分结果文件是以文本文件存放的,这为监测数据仓库的数据集成省去了很多必要的处理过程,起到了一定的数据预处理作用。Sybase 数据库和文本文件这两种源数据都可以通过 DTS 方便地转换和合并到监测数据仓库中。在转换和合并的过程中,对于简单的数据转换或数据移动,可以使用 DTS 导入\导出向导,而对于复杂的数据析取过程,可以使用 DTS 设计器创建一个包含复杂工作流的 DTS 包(例如,包含分支、多个步骤、多个连接、复杂的逻辑以及事件驱动代码的包)或编写 DTS 包的应用程序。最终的目标就是利用 DTS 工具,把原有监测系统数据库中的数据以及原系统生成的一些文本文件形式的分析结果装载到监测数据仓库中的基表和一些用于数据挖掘的特定数据表中。

按照数据集成的步骤,首先要进行数据选择,数据选择是根据监测数据仓库中的基表来进行的,所以先要定义和设计基表,然后根据基表中的内容,逐字段地选择源数据并建立连接。有时基表中的字段内容可能是源数据字段内容中的一部分,也可能包括几个源字段的内容,还可以不是源数据,而是需要系统生成的一些数据,这些都需要编写转换程序来实现。

a. 小浪底监测数据仓库基表的设计和定义。根据数据仓库设计时所确定的主题、维度及相应粒度的划分,监测数据仓库的基表定义见图3,包括时间维表、位置维表、环境量维表、测点和观测项目事实表。

b. 监测数据的导入与转换。监测数据仓库中各基表包括的内容不同,其转换的难易程度也各不相同,有的字段可以直接从源数据表中复制过来,而有的要经过一定的转换。所以要根据(下转第62页)

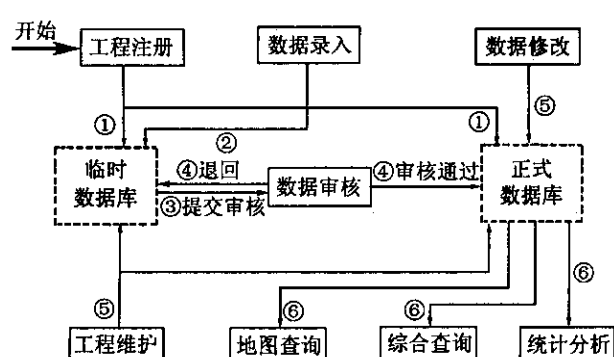


图4 水利工程数据库维护流程

器,会造成负荷过重、响应过慢等问题,同时,病毒攻击、设备硬件故障及火灾等都可能对该服务器中的数据造成不可挽回的损失,为此本文提出采用服务器群的方式实现数据的异地备份和分担网络荷载。

服务器群是同时将安装有相同数据的服务器分别放置于浙江省水利厅和全省11个地市水利局,实现荷载的分解及数据安全同步,具体作用为:①对数据和服务实行异地备份,充分保障水利工程数据的安全性。②实行服务器荷载分担,全省12台水利工程数据库服务器形成一个有机的服务器群,并由系统统一管理调度,省水利厅服务器除提供其他服务器相同的数据服务外,还提供系统单点登录、统一用

户、统一认证、统一授权等管理服务。③实行网络荷载分担,分区域提供用户服务方式减少了省级网络的数据包流量,可大大减轻省级网络的荷载,节约网络带宽资源,降低网络运行成本。④通过专门的数据更新同步代理程序实现省水利厅主服务器对其他各服务器的数据分发,以保证服务器间数据的一致性。

5 结 论

水利工程数据库的维护模式是值得很好研究的一个课题,数据库作用的发挥是与其维护模式紧密相关的。本文根据水利工程数据库的特点,结合我国对水利工程管理的现状,提出了以网络为载体建立省、市、县三级协同维护,临时数据库与正式数据库有效结合以及通过服务器群实现数据异地备份和网络荷载分解等的维护模式,这既能很好地保证数据库维护工作的日常进行,又可以确保工程数据的惟一、完整和及时。

先进的IT技术只有与现行的管理模式很好地结合才能发挥作用,本文提出的数据库维护模式经实践证明,能很好适应水利系统的实际工作。

参考文献:

[1] SL213-98,水利工程基础信息代码编制规定[S].

(收稿日期 2004-07-19 编辑 熊水斌)

(上接第58页)转换的复杂度采取相应的DTS工具进行各表数据的导入和转换。①时间维表数据的导入。使用DTS设计器建立一个数据转换包,在创建新转换时,通过对源和目的字段的日期格式的设置,实现年、月、日、时等时间粒度的提取。②位置维表数据的导入。位置维表中大部分字段的数据都是来自于原工程部位数据库(md_gcbw)中的测点信息表(xldmd_cdx),而且不需要类型及内容的转换,只是简单的数据移动,所以可以使用DTS导入\导出向导来交互地完成位置数据的装载。③其他表数据的导入。其他几张表的数据转换都可以通过以上的两种方法进行数据的转换和导入,只是在处理过程中稍有不同。当碰到复杂的数据转换时,可以在创建新转换时编写转换脚本。另外,以上使用DTS工具进行数据转换和导入的过程,都可以保存为一个DTS包,以备日后对新数据进行转换和导入。

4 结 语

小浪底大坝安全监测数据仓库的建立将更能满足资料分析、信息挖掘和综合评判、安全监控等需求^[3],一些相应技术、方法、工具的应用将在一定程度上解决原系统存在的知识贫乏、分析手段不灵活、缺乏二次分

析能力、数据不一致等问题,为小浪底水利枢纽工程提供一个良好的数据体系化环境,使小浪底大坝安全监控系统展现更强的活力和应用价值,延长其生命周期,使其先进性和实用性达到一个更高的层次。

另外,基于数据仓库的大坝资料分析和安全决策系统是一个庞大、复杂、涉及面广的综合系统,本文虽然初步建立了小浪底大坝安全监测数据仓库,实现了部分功能,但有待于进一步提高和完善。因为一项技术的引入,在解决问题的同时也会带来新的问题,特别是如何灵活地运用数据仓库软件提供的各种接口开发适合分析大坝监测数据的OLAP工具软件,如何开发有关挖掘算法的数据挖掘工具对监测数据进行有效的挖掘以及数据仓库本身一些技术问题的解决,都需要做进一步的研究。

参考文献:

[1] 舒云星,李晶,贺贵明.三峡工程安全监测信息系统决策支持数据仓库技术[J].计算机工程,2002,28(11):108—110.

[2] 袁连海,陶宏才.基于数据仓库的数据转移方法[J].电脑开发与应用,2002,15(2):58—62.

[3] 宋小刚.数据仓库技术在大坝资料分析与安全决策中的应用研究[D].南京:河海大学,2004.1—2.

(收稿日期 2004-06-22 编辑 熊水斌)