

# 校园一卡通数据仓库系统<sup>①</sup>

商新娜, 孙连英, 彭涛, 刘畅

(北京联合大学 信息学院, 北京 100101)

**摘要:** 校园一卡通在校园中涉及的应用领域越来越多, 为师生员工带来了极大的方便。同时, 校园一卡通系统中数据量也越来越大, 传统的关系数据库对于如何有效的存储、管理、利用这些数据没有有效的办法。提出了基于数据仓库技术整合、管理校园一卡通的历史数据和实时数据的方案, 建立了数据仓库模型, 并对其中的关键问题进行了探讨, 为校园一卡通系统的进一步发展做出了有益的探索和实践。

**关键词:** 校园一卡通; 数据仓库; 建模

## Campus Card Application System Based on Data Warehouse

SHANG Xin-Na, SUN Lian-Ying, PENG Tao, LIU Chang

(Information College, Beijing Union University, Beijing 100101, China)

**Abstract:** More and more applications are supported by campus Card system. The campus card bring convenience and mass data at the same time. The data center must consider the solution of saving, managing and making use of the mass data. This paper gave a suggestion of Integrating and managing the data in campus card system based on data warehouse technology. The model was constructed and some key issues were investigated. It is a useful explore and practice for the development of campus card system.

**Key words:** campus card; data warehouse; modeling

近年来, 发展和建设数字校园成为许多学校进行科学化、现代化管理的重要举措, 校园一卡通是数字校园建设的重要组成部分, 它的普及使用极大的方便了师生的日常活动和学校的管理。但是, 随着一卡通所涉及的业务越来越多, 大量的流水数据也与日俱增, 对于如何有效的存储、检索、分析、展示这些数据, 传统的关系存储模式已经很难提出有效的解决方案。数据仓库技术为解决海量数据的众多问题找到了一种有效的途径, 因此, 建设校园一卡通数据仓库对于提高数据的使用效率和提供决策支持是一个有益尝试。

## 1 校园一卡通系统与数据仓库

### 1.1 校园一卡通系统

校园一卡通系统是针对目前校园中使用的证件繁多、管理复杂的情况而设计的, 用一张卡代替学校目

前使用的各种证件(包括学生证、工作证、借书证、医疗证、出入证等)。师生员工在学校各处出入、办事、活动和消费均可只凭这张校园卡进行, 实现“一卡走遍校园”, 给师生员工带来极大的方便。

校园一卡通的主要应用领域有:

(1) 消费支付功能: 学校的各个消费场所都提供了刷卡 POS 机, 持卡人可以在这些地方刷卡消费, 如食堂买饭、校园超市购物、校园书店买书、文印处打印复印学习资料、理发馆理发等, 可以在指定网点进行充值、挂失等。

(2) 身份识别功能: 校园卡表面印有持卡人身份标识, 如学号或职工号、姓名、班级和照片, 实现视觉身份识别。校园卡的 IC 芯片和数据库系统中存有持卡人的身份信息, 实现校园内的电子身份识别, 如进出图书馆、宿舍楼、实验楼等的身份验证, 规范了校园

<sup>①</sup> 收稿时间:2011-07-04;收到修改稿时间:2011-07-08

内的秩序,提高了校园的安全性。

(3) 图书借阅功能:校园卡可以与图书管理系统结合,替代原来的借书证。记录持卡人在图书馆的借书、还书情况。

(4) 考勤功能:校园卡与教室、机房的考勤系统结合,学生在上、下课的时候通过从考勤机上刷卡,确定学生的出勤、缺勤情况。教师通过刷卡开启多媒体教学设备,也达到了对教师上课情况的记录。

(5) 其他扩展功能:校园一卡通正在逐步探索与其他系统的融合,如与银行系统的融合,实现上交学费、奖助学金发放等功能;与教务系统逐步融合,实现学籍管理、选课管理、成绩管理等功能;与校园网络不断融合,实现上网统计、计费等功能.....

校园一卡通的建设是一个综合项目,并在不断的发展中扩充和延伸,其体系结构图如图 1 所示。

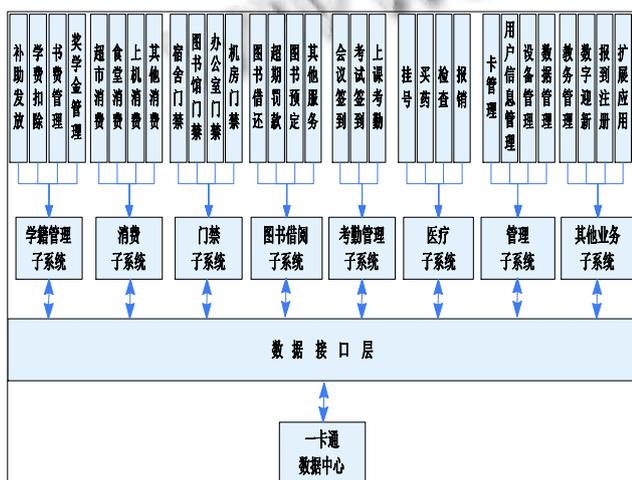


图 1 校园一卡通系统结构图

校园一卡通的发展使得涉及的业务领域越来越多,同时数据量也越来越大,现在校园一卡通的日数据量达到了百万条,年数据量上亿条,并且有不断上涨的趋势。这些信息需要有效管理和组织,并挖掘其中的信息价值。如对学生的消费行为分析、就餐流量分析、食堂经营状况分析、学生对图书的借阅喜好、学生考勤情况分析等,发现学生在校行为的一般规律和特殊情况,为学校的学生管理、困难生考查、经营分析等工作提供了真实准确的依据,便于作出合理正确的决策。

随着数据量的增大,校园一卡通系统逐步表现出对数据有效管理和分析的后劲不足。如何有效管理这

些信息并发挥它们的价值是在一卡通逐步普及过后需要进行的思考和探索。数据仓库和数据挖掘技术是解决这个问题的一种有效途径。

## 1.2 数据仓库

数据仓库(Data Warehouse,简称为 DW)是一个面向主题的、集成的、不可更新的、随时间不断变化的数据集合,它用于支持企业或组织的决策分析处理。数据仓库能将分散的、异构的原始业务操作数据转化为集中统一、即时利用的可用的信息,同时提高访问和处理数据的速度和效率。通过从不同的角度、不同的维度进行数据的抽取、汇总和分析,能灵活的面向用户的各种需求,结合数据挖掘技术,提供决策支持。

## 2 基于数据仓库的系统架构设计

### 2.1 设计目标

基于数据仓库的校园一卡通系统的设计目标就是设计数据模型、有效整合资源、提供分析平台。要在新的数据模型下,实现历史数据和实时数据的自动导入,实现基于维度的数据展示和分析,为深层次的数据挖掘提供数据平台。

### 2.2 校园一卡通数据仓库系统结构

基于数据仓库的校园一卡通系统的建设主要分为三个主要的部分,如图 2 所示。

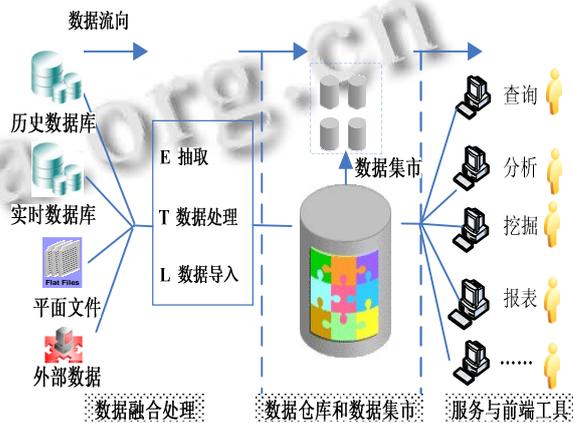


图 2 校园一卡通数据仓库系统结构图

(1) 数据融合处理:校园一卡通系统的数据来源多样,结构各异,因此数据仓库要能将多种数据源进行有效整合,包括数据库信息、平面文件信息、外部文件信息、元数据信息等,既能支持历史数据的集中导入,也支持每日增量信息的自动添加。

(2) 建立数据仓库和数据集市:根据设计的数据仓

库模型，建立数据集市和数据仓库。

(3) 数据挖掘和多维展示:通过数据挖掘技术和统计查询工具，为前端用户提供多维数据展示，为决策提供支持。

### 2.3 数据仓库模型设计

校园一卡通系统涉及的应用范围广泛、信息种类多，主要的信息包括：刷卡消费信息、图书借阅信息、考勤管理信息、医疗及报销信息、学籍管理信息等，这些信息每天都增长很快，是动态变化的信息。通过对系统的主题域分析，基于数据仓库的校园一卡通系统的模型设计为多维数据模型中的星形模式，具有多个事实表和多个维度。其概念模型如图 3 所示。

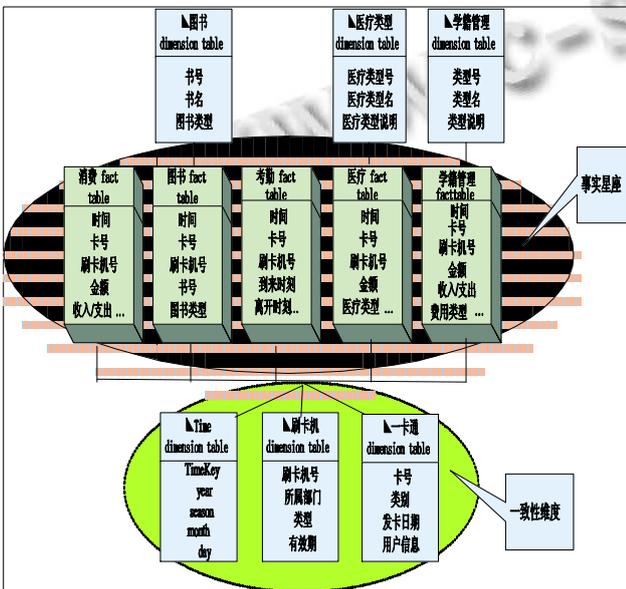


图 3 校园一卡通系统数据仓库模型设计

模型设计的关键问题:

#### (1) 事实星座和一致性维度

复杂的应用一般需要多个事实共享表，这种模式可以看作星形模式集，因此称为星系模式，或事实星座<sup>[1]</sup>。校园一卡通数据仓库系统中，消费信息、图书借阅信息、考勤信息、医疗信息、学籍管理信息等作为星形模式的事实表，它们是不同的主题域。经过分析，针对这些主题的维度既有公共的维度，也有自己特点的维度，如各个主题域都有基于时间、一卡通信息和刷卡机信息的分析维度，图书借阅主题域中又有关于图书信息的维度，医疗主题中有关于医疗类型维度的信息。因此，模型中建立了事实星座和共享维

度，将它们有共同的维度信息：时间、时刻、刷卡机信息、一卡通信息作为共享维度，也称为一致性维度，使得数据仓库中可以实现对一个以上的业务过程进行数据分析，即交叉探查 (drill-cross)<sup>[2]</sup>。

#### (2) 聚合事实表的应用

聚合是在维度数据仓库上提高查询性能的最有效的方法，同原子数据相比，聚合数据表在更高的角度对数据进行了概括<sup>[2]</sup>。在校园一卡通数据仓库系统中，原子事实表都是业务流水信息的记录，其时间粒度为秒，而按照日、月、季度、年的查询和展示需求很多，如果直接在原子事实表上进行聚合运算，将大大减慢系统的响应时间，并且做了很多重复计算，因此进过分析，针对原子事实表，通过聚合计算，产生聚合数据表，作为事实表。如在图 3 所示的事实表基础上，设计产生了各个主题域的聚合表：日表、月表、季度表、年表，如销售主题域的日表的信息包括：时间(具体到日)、刷卡机号、刷卡金额等，其中的刷卡金额就是根据销售原子事实表进行 GROUP BY 和 SUM() 计算得来。

#### (3) 渐变维度问题

在数据仓库的模型中，事实表中存储动态变化的数据，维度表中的信息是相对稳定的。但是在现实世界中，随着时间的变化，有些维度信息也是变化的，虽然这种变化不是经常发生的。如一卡通信息中的用户所属班级或部门，在转专业、留级、聘岗等事件中，这些信息会发生变化，这种变化会影响到针对班级和部门的信息统计分析。处理这种属性值变化的方案是设计渐变维度 (slowly changing dimensions,SCD) 技术，采用增量式添加维度记录行的方法记录维度变化，新行中的属性随时间向前推移一直是有效的，直到再次发生变化而产生下一条记录时，前一条关于变化的记录才失效。使得在进行多维分析时，能区分事实表的历史，准确跟踪维度属性变化。

### 3 系统建设的关键问题

ETL (Extract-Transform-Load)，即填充、更新数据仓库的数据抽取、转换、装载的过程。是数据仓库建设的第一步，是为数据挖掘和数据多维展示提供高质量数据的保证。基于 Microsoft SQL Server 2005 Integration Services (SSIS)进行校园一卡通数据仓库系统的 ETL 实施。SSIS 的 ETL 实施过程是由许多包任

务组建的, 这些独立的包任务能够完成自身的功能。SSIS 工具可实现包任务调用的功能, 即通过执行一个包任务, 有序的依次执行 ETL 规则中涉及的每一个包任务。通过“控制流”的“工具箱”中, 添加“执行包任务”组件进行包任务的封装。图 4 所示为封装包任务。

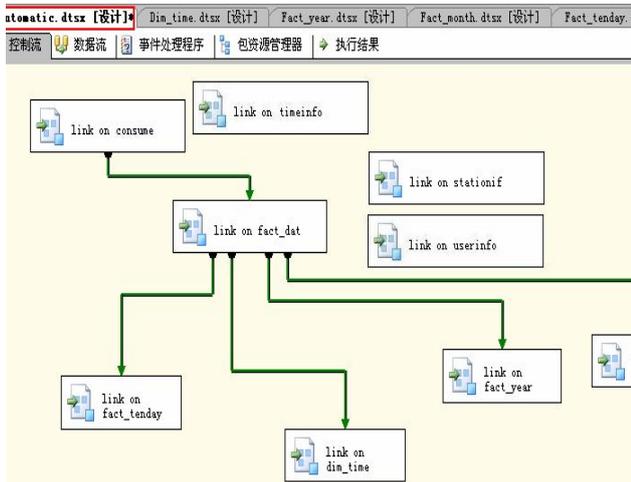


图 4 封装包任务

在 ETL 设计过程中, 数据的装载是一项重要工作, 对于初始数据装载, 可以使用全量抽取的方式, 将原始数据表中的数据全部经过转换和过滤导入到目标表中, 但是对于业务系统来说, 每天产生大量的流水数据, 这些新增数据如何及时的、自动的导入数据仓库, 是 ETL 过程要解决的问题。增量抽取一般有以下几种方式: 触发器方式、时间戳方式、全表删除插入方式以及全表对比方式。增量抽取的时间间隔一般为一天到一周为宜。间隔太长使得数据得不到及时更新, 太短又影响系统响应效率。本系统中, 采用每天执行一次的增量抽取方式。在增量抽取时, 因为只是抽取当日产生的流水记录, 因此可以通过匹配系统当日时间和事实表中的时间进行抽取, 并且设置每日自动抽取。图 5 所示为自动抽取设置。

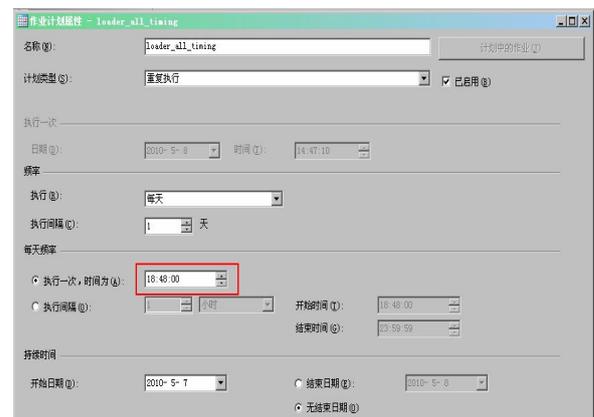


图 5 设置增量抽取自动执行

## 4 结语

数据仓库的建设是解决海量数据存储、管理以及进行数据分析和挖掘的有效方案, 近年来越来越多的信息系统进行数据仓库建设工作, 取得了较好的效果。校园一卡通虽然历史不长, 但是由于其涉及的应用范围不断扩大、用户基数大, 几年的时间里就积累了海量的数据, 因此建设数据仓库对解决校园一卡通系统的数据管理是有价值的, 也为进一步的数据分析和挖掘提供了很好的平台。

## 参考文献

- 1 韩家炜, 等. 数据挖掘概念与技术. 北京: 机械工业出版社, 2006.
- 2 Mundy J. 数据仓库工具箱. 北京: 清华大学出版社, 2007.
- 3 Inmon WH. Building the Data Warehouse. 4th ed. Indiana, USA: Wiley Publications, 2005.
- 4 吕海燕, 傅英亮, 邢翠芳. 数据仓库在银行客户关系管理中的应用. 大连海事大学学报, 2007, 33(3): 183-186.
- 5 袁宝兰. 基于 SOA 的校园一卡通平台设计及其应用. 吉林师范大学学报, 2009(1): 134-137.
- 6 李珊娜. 基于校园一卡通平台的数据挖掘应用研究. 铁路计算机应用, 2010, 159: 55-58.
- 7 吴小勇, 刘磊. 基于数据仓库的装备体系数据建模方法. 计算机工程, 2010, 36(1): 76-80.