

教务决策数据仓库研究与设计

专业名称：计算机应用技术专业

申请者姓名：姜永亮

导师姓名：陈启买 副教授

摘要

随着高校教学管理信息化的日益发展与日趋成熟,基于教务管理信息系统形成的海量历史数据进行深层次分析与应用,对提高学校教学管理与决策的质量与水平具有积极而深远意义。

论文基于高校教务管理与决策需求,将 SDLC (software development life cycle) 方法应用于教务数据仓库的开发与设计,采用有反馈自顶向下方式构建了某高校的教学决策数据仓库系统。基于该数据仓库系统,一方面解决了长期制约高校教学管理改革与发展的高校异构教学数据集成与共享问题;另一方面,借助于 OLAP 分析技术在高校教学决策中的应用,也为高校教学管理与决策的精细化、科学化与自动化管理提供了一种行之有效并具有指导价值的解决方案与途径。

论文采用插件技术设计教务决策数据仓库专用 ETL 工具,将面向对象方法应用于 ETL 工具的实现等,这些技术和方法的研究与应用,对指导当今高校教学数据仓库建设与实施具有重要意义。

关键词：数据仓库 OLAP ETL 教务决策 Brio

The RESEARCH AND DESIGN OF
EDUCATION ADMINISTRATION
DATA WAREHOUSE

Major: the application of computer

Name:JiangYongliang

Supervisor:ChenQimai

ABSTRACT

With the maturity and development of informationization in the higher education management the education management information system has formed massive historical data.If we can analyzed it deeply and used it effectively it will has positive significance to the improvement of management and the quality of decision-making.

This papers based on the management and decision-making needs of higher education, and used the SDLC(software development life cycle) method into the data warehouse development,and used a top-down approach with feedback to construct a decision-making data warehouse system for a university.With the data warehouse system on the one hand the university resolved the heterogeneous data integration and sharing which constrained the reform and development of the education administration for a long time.On the other hand by the OLAP technology the data warehouse provided an effective solution for the university, which can make the management and decision-making fine, scientific and automation.


This paper used plug-in technology to design a special ETL for the data warehouse.The writer used the object-oriented approach into the realization of ETL.It has great significance to research and application of these techniques and

methods. Because they played a great role in guiding the construction of data warehouse.

KEY WORDS: data warehouse, etl, olap, dss of education administration , brio.

华南师范大学学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确的方式标明。本人完全意识到此声明的法律结果由本人承担。


论文作者签名：

日期：08年6月10日


学位论文使用授权声明

本人完全了解华南师范大学有关收集、保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属华南师范大学。学校有权保留并向国家主管部门或其指定机构送交论文的电子版和纸质版，允许学位论文被检索、查阅和借阅。学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印、数字化或其他复制手段保存、汇编学位论文。（保密的论文在解密后遵守此规定）

保密论文注释：本学位论文属于保密范围，在____年后解密适用本授权书。非保密论文注释：本学位论文不属于保密范围，适用本授权书。

论文作者签名：

日期：08年6月10日

导师签名：

日期：08年6月10日

第 1 章 绪论

1.1 研究背景和意义

随着高校教学及管理信息化的建设与发展,学校教学管理系统积累的数据日趋海量,但是,对这些数据的应用,尽管各高校教学管理系统具备一定的数据查询与统计功能,由于技术与手段的滞后,多角度对历史数据进行分析,以支持高校教学管理与决策深层次应用还处于起步阶段,主要表现为:报表功能低下、数据关联度低,对于历史数据缺乏多角度、动态分析能力,不能及时发布信息,分析信息的代价太高等。

如何从海量数据中发现隐藏有用知识和规律,以辅助高校教学管理与决策,已成为当今高校不断强化教学管理意识、提高教学管理效能与水平,从而进一步提高教学质量与水平的核心问题之一。

本文提出的教务管理决策数据仓库的研究与设计课题,旨在通过教学数据仓库创建,对分散于教学业务系统中业务数据,基于决策主题作必要的数据集成与存储,然后对其中数据进行预测、分析,发现其中潜在有用规律与知识,为学校教学管理与决策提供支持。

1.2 国内外研究现状

国内教务管理系统主流产品主要有清华大学综合教务管理系统、应龙教务管理软件、正方教务管理信息系统、AOLONG 高校综合教务管理系统、青果教务管理系统等;国外教学管理主流产品主要有 Centre/SIS、Focus/SIS 等。对比分析这些产品,其主要功能大多侧重于教学管理与教学运行,对教学管理与运行数据的深层次管理与分析应用,还基本停留在个别数据分析指标查询与统计方面。

21 世纪以来,为了基于教学运行数据提供教学管理与决策分析应用,国内许多高校基于数字化校园建设需求,或规划或建设了各自的教学数据中心,例如,清华大学在其原有综合教务管理系统基础上,设计的结构化三层教学决策支持系统数据仓库^[5];中国计量学院将数据仓库技术应用到高校会计分析领域以辅助本校的财务管理决策系统;西南民族大学以本校实验管理系统为基础创建的高校实

验室管理数据仓库等；贵州大学学生成绩分析数据仓库系统；浙江理工大学在原有图书管理系统的基础上进行了有关图书采购决策数据仓库的研究等。这些产品或方案，虽然已经有部分测试性案例或应用，但无论是规模、范围或深度，均还处于起步阶段，其数据仓库模型设计、ETL 设计、数据增量更新策略、OLAP 技术应用深度与广度系统性能优化等方面均显不足，主要表现为：数据集成能力差，尤其在异构数据源的集成方面；数据共享程度不高，很难解决不同部门，不同校区间数据共享问题；对数据分析深度还有待于提高；数据挖掘模式不够丰富，数据挖掘知识发现的应用领域还有待于扩展。

在国外，有关数据仓库建设与实施先于国内，尤其是教务数据仓库方面，欧美等发达国家虽然强于国内，但是由于中、西方教育体制的不同，基于数据仓库技术的分析主题理念与设计应用差别较大，因此，根据自身基于体系与管理决策模式特点研究、设计适合中国高校教务决策应用的数据仓库系统与解决方案不但必要，也是当务之急。

1.3 问题提出

鉴于国内外教学数据仓库系统的建设现状与需求，本文采用 SDLC 方法设计一个适合高校教务决策的教务决策数据仓库其结构如图 1-1 所示。其主要研究技术包括：教务决策数据仓库设计目标与原则，基于信息包图教务决策数据仓库概念模型与逻辑模型设计，基于插件技术教务决策数据仓库专用 ETL 工具设计，面向对象技术在教务决策数据仓库专用 ETL 工具实现中应用。

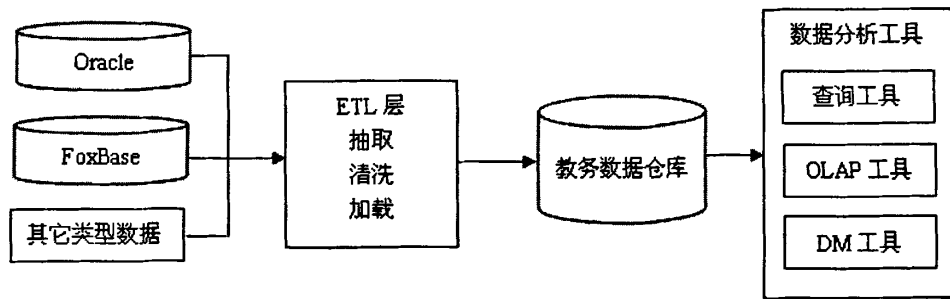


图 1-1 教务决策数据仓库结构示意图

1.4 研究内容

本文主要针对高校教务决策支持领域，以某高校的教务管理系统为基础，开

展教务决策数据仓库设计与实现之研究，主要研究内容有：

1、分析现有教务管理系统数据现状与需求，提出教学数据仓库主题设计原则、目标，构建教务数据仓库概念与逻辑模型。

2、基于教务决策数据仓库主题与决策需求分析，使用信息包图技术完成教务决策数据仓库概念设计与逻辑设计。

3、基于插件技术的教务决策数据仓库专用 ETL 工具研究，面向对象技术在 ETL 工具实现中的应用。

1.5 全文的组织结构

论文的组织结构安排如下：

第 1 章 绪论。论述选题的研究背景和意义、国内外研究现状，给出本文的研究内容以及组织结构。

第 2 章 教务决策数据仓库相关概念与技术。对数据仓库、ETL 进行简单介绍。给出数据仓库的常用数据模型、构建策略。

第 3 章 教务决策数据仓库模型研究。介绍了教务决策数据仓库的设计目标与设计原则，教务决策需求与分析，教务决策数据仓库的体系结构以及教务决策数据仓库模型。

第 4 章 教务决策数据仓库设计。对教务决策的数据源进行理解与分析，给出了教务决策所涉及主题域的分析与提出过程。采用信息包图完成教务决策数据仓库概念模型与逻辑模型设计，对教务决策数据仓库数据粒度进行设计，采用手工编程方式设计一个 ETL 工具。

第 5 章 教务决策数据仓库相关技术实现，采用 Oracle 数据仓库解决方案，实现教务决策数据仓库，采用面向对象技术实现 ETL 工具。

第 6 章 数据仓库典型应用实例分析。给出了教务决策数据仓库中若干主题的典型应用实例分析结果。

第 7 章 总结与展望。总结已做的研究工作，展望未来，提出今后的努力方向。最后是参考文献，致谢。

第 2 章 相关概念和技术综述

2.1 数据仓库

数据仓库概念最早于 1988 年由 W. H. Inmon 在《Building Data Warehouse》一书中提出^[8]。有“数据仓库之父”之称的 W. H. Inmon 给出的数据仓库定义是：数据仓库是一个面向主题的、集成的、不可更新的、随时间变化的数据集合，用于支持管理决策过程。根据这一概念，数据仓库主要具有面向主题、高度集成、相对稳定、随时间变化等四个最重要的特征。

2.1.1 数据仓库的体系结构

数据仓库一词仍然至今仍没有一个统一的定义，通常地，一个完整数据仓库系统由数据获取、数据存储和数据访问三个部分组成，一个典型的数据仓库系统的结构如图 2-1 所示。

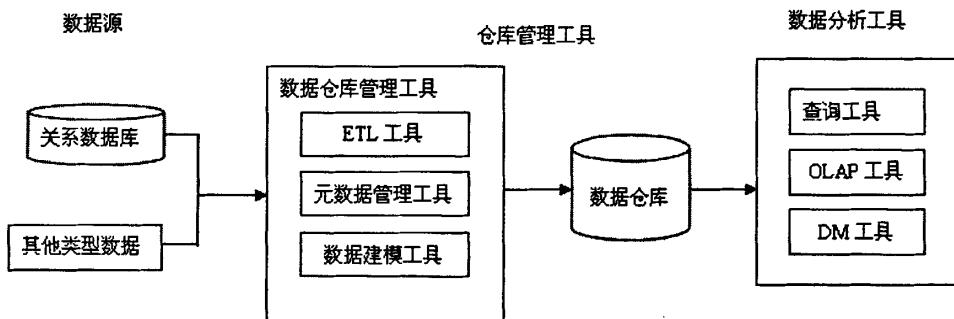


图 2-1 数据仓库系统的结构图

如图 2-1 所示，来源于多个信息源的原始数据，在进入数据仓库之前必须经过抽取、清洗、转换。对数据仓库中数据的访问主要通过访问工具来实现，所以数据仓库访问工具必须能够为用户提供统一、协调和集成的信息环境，以支持企业全局管理与决策应用。

数据源

数据源作为数据仓库系统基础，为整个系统提供原始数据。通常情况下数据源主要包括企业内部信息和外部信息。

仓库管理工具

仓库管理工具实现从源数据到数据仓库数据的抽取、清洗和转换，维度划分

以及数据仓库物理存储结构确定等。

分析工具

分析工具主要指为完成决策问题所需要的各种查询工具、数据分析工具、数据挖掘工具等。

2.1.2 数据仓库的构建策略

随着数据仓库技术的发展,数据仓库在实现策略上已经由原先的自顶向下一种模式发展到:自顶向下,自底向上,有反馈的自顶向下,有反馈的自底向上等模式。

自顶向下

自顶向下模式是首先把 OLTP 数据通过 ETL 汇集到数据仓库中,然后再把数据通过复制方式推进各个数据集中。其创建过程如图 2-2 所示。

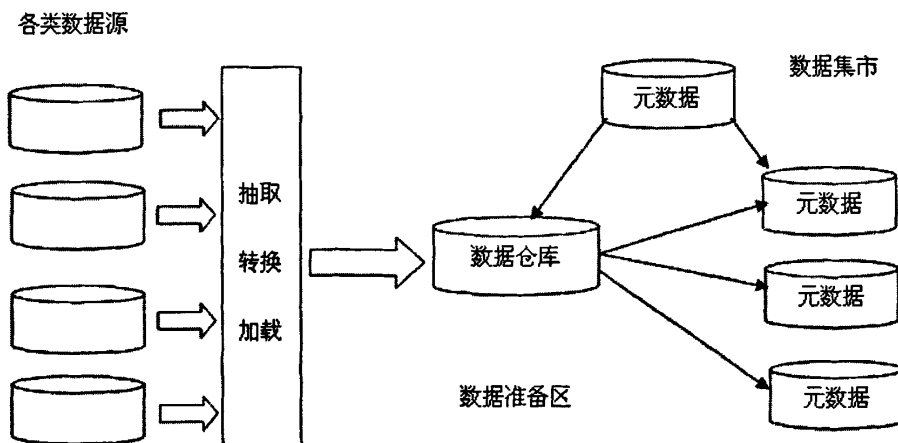


图 2-2 数据仓库自顶向下构建示意图

自底向上

自底向上模式是先将OLTP数据通过ETL汇集到数据集中,然后通过复制方式提升到数据仓库中,其创建过程如图2-3所示。

有反馈自顶向下

有反馈自顶向下模式将用户的新需求反馈分为两个阶段:

第一阶段,用户新需求被反馈给部门数据集市,部门数据集市根据用户新需求产生自身需求变化。

第二阶段,部门数据集市把自身需求变化反馈给全局性数据仓库,全局性数据仓库再做出相应变化,其创建过程如图 2-4 所示。

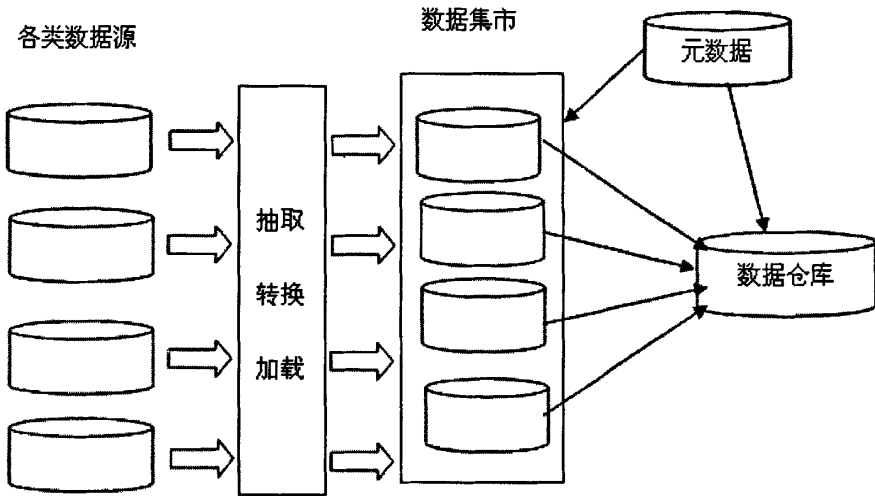


图 2-3 数据仓库自底向上构建示

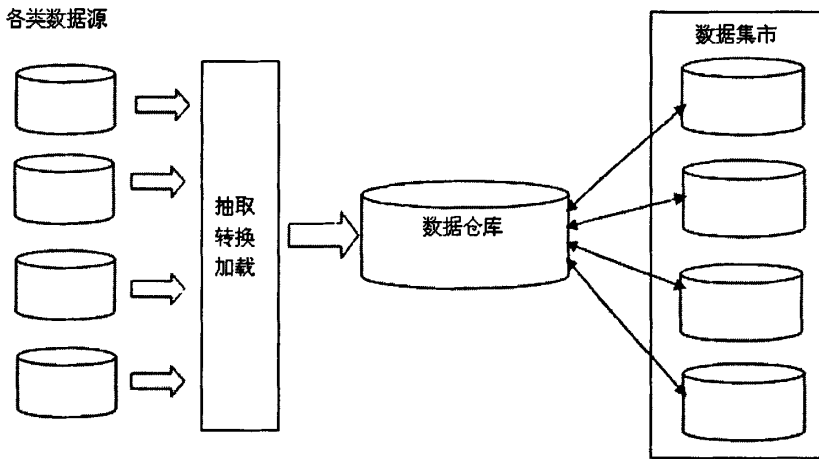


图 2-4 数据仓库有反馈自顶向下构建示意图

有反馈自底向上

在有反馈自底向上模式中，建立好全局性数据库之后，需求变化主要体现在数据集市与数据仓库之间。这样可以在保持各个部门数据集市相互之间数据一致性时，根据用户反馈信息不断调整自己，以这种模式建立的数据仓库在投入使用后能使因用户需求变化所带来的不便程度降低，其创建过程如图 2-5 所示。

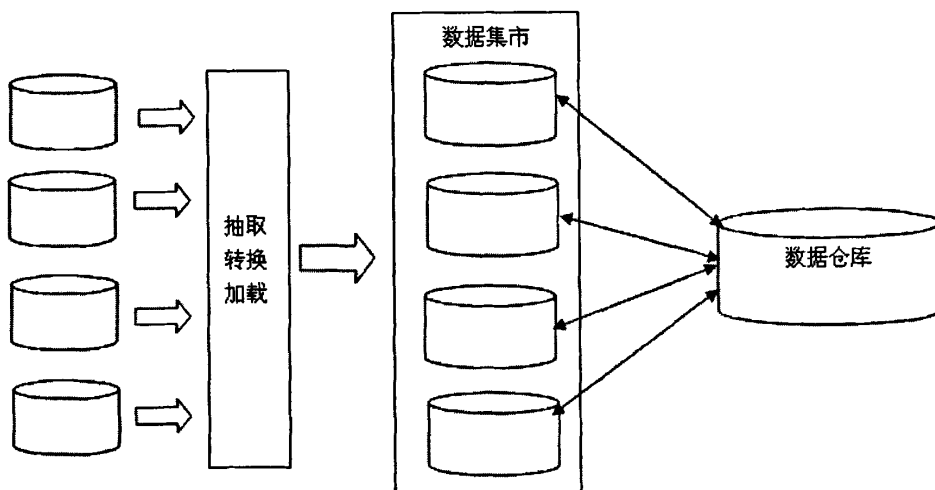


图 2-5 数据仓库有反馈自底向上构建示意图

2.1.3 数据仓库中常用数据模型

数据仓库系统是一个涉及很多业务领域的系统,因此在数据仓库系统开发实施之前,通过建立数据模型,统一规划各种业务数据关系,对于消除不同环节间数据重复非常有利,能有效确保数据在结构和内容上的统一,顺利推进数据仓库系统实施,缩短系统实施工期,降低其建造成本的效果。

常用数据仓库数据模型主要有实体关系模型和多维模型两种数据模型^[11]。实体关系模型在传统数据库开发中被广泛采用,对于复杂系统环境下数据关系抽象和简化非常有效。多维模型是随着数据仓库技术发展而形成,主要用于反映用户需求、表现数据关系以及联机分析处理 OLAP 系统数据建模等。

多维模型是一种用于多维结构数据的建模技术,该技术主要用来描述事物之间的多种关系,并能从多角度、多层次的反映这些关系。通过对数据进行重组与综合,组织和建立数据的多维结构,为决策分析提供数据的多维视图^[12]。与传统关系数据库不同,它是一个多维数据库,它扩展了关系数据库模型,主要以星型架构为结构方式,并在此基础上引申了雪花模型和星座模型。

星型模型

星型模式通过一个包含主题的事实表和多个包含事实非正规化描述的维度表来支持各种决策查询。典型的星型数据模型结构如图 2-6 所示。如图所示,星型模型中各种不同维度表连接是通过事实表实现,各个维度表连接到中央事实表。各个维度表对象之间的联系是通过事实表与另一维度表中对象相关联来建立

的，每一个维度表与事实表的连接通过主键实现。

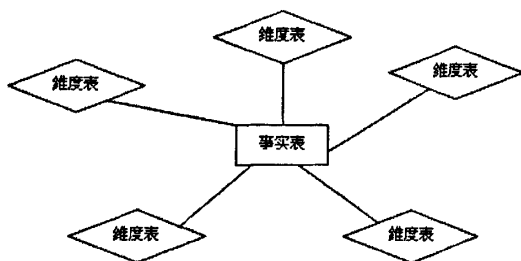


图 2-6 维度与事实的星型模型结构图

如图 2-6 所示，星型模型中各种不同的维度表连接是通过事实表实现，各个维度表连接到中央事实表。各个维度表对象之间的联系是通过事实表与另一维度表中对象相关联来建立，每一个维度表与事实表连接通过主键实现。

事实表主要包含描述特定商业事件数据，在事实表中存放的数据在一般情况下不允许修改，就是修改也只是将新增数据简单地添加到事实表中，事实表中数据各类特征数据存放在各维度表中。

与一般数据库中表的外键允许为空不同，数据仓库中事实表外键不能为空，原因在于维度表利用维度关键字通过事实表中外键约束于事实表中某一行，实现与事实表关联。这种结构使得用户能够从维度表数据分析开始，获得维度关键字，连接到中心事实表，进行查询，而不再对事实表中数据进行扫描，提高了数据查询性能。

尽管星型模式还是一个关系模型，但是它已经不再是一个规范化模型。采用星型模式设计数据仓库，其数据已经过预处理，进行查询时只需扫描事实表即可，故此，采用该模式架构数据仓库具有较高的查询访问效率。维表一般都比较小，与事实表进行连接时其速度较快，对于用户来说理解相对容易点，通过星型模式分析，用户能较容易组合出各种查询。

雪花模型

雪花模型是对星型模型的一种扩展，与星型模型不同，在雪花模型中每一个维度都可以向外连接多个详细类别表。维度表除了具有星型模型中维度表功能外，还连接对事实表进行详细描述的详细类别表，详细类别表通过对事实表在有关维上的详细描述，达到缩小事实表和提高查询效率的目的，数据仓库雪花模型结构如图 2-7 所示。

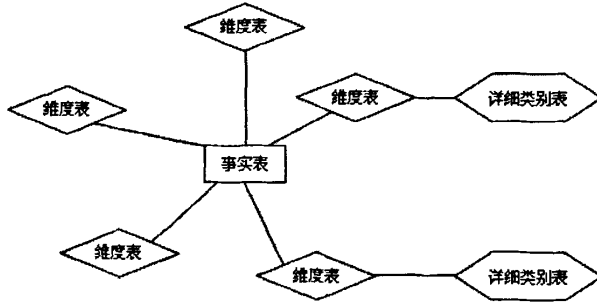


图 2-7 维度与事实的雪花型结构图

在雪花模型中对维度表进行了规范化处理，维度表中存放的是规范化的数据，通过把多个较小标准化表联合在一起来达到改善查询性能的目的。正是由于对维度中数据标准化处理以及维的低粒度，导致采用雪花模型构建的数据仓库应用具有很大灵活性。但在雪花模型中连接花费时间较多，使其查询效率大打折扣。

星座模式

星座模式是为解决数据仓库中多个事实表共享某一个或多个维表问题而扩展的一种数据仓库架构模式，其结构如图 2-8 所示。

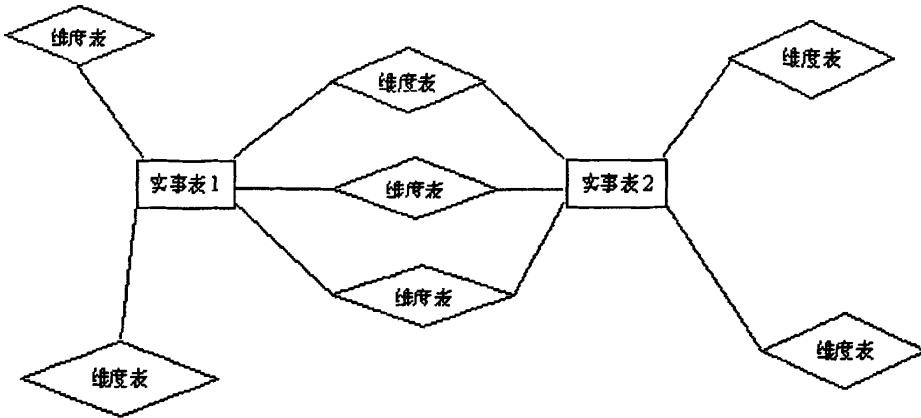


图 2-8 维度与事实的星座结构图

2.2 ETL

通常地，管理信息系统中的数据以操作型数据为主，在进入数据仓库之前必须经过 ETL 工具处理之后才可以进入。在构建数据仓库的过程中 ETL 设计是一个必不可少的环节。

ETL 是一个过程，主要功能就是完成数据的抽取 (Extract)、转换

(Transform)、清洗(Cleaning)、装载(Loading)工作^[16],其体系结构图 2-9 所示。

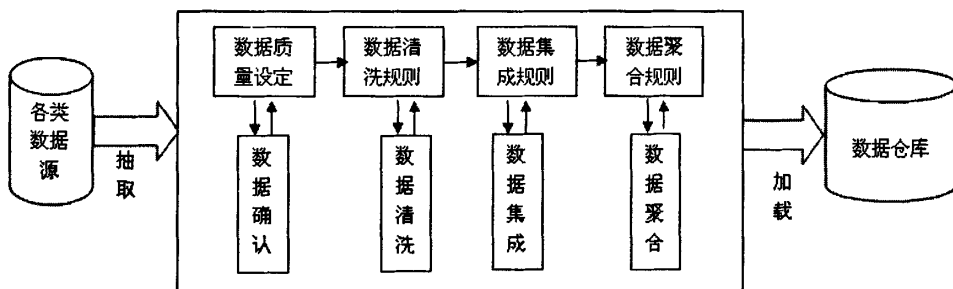


图 2-9 数据仓库中 ETL 体系结构图

数据抽取

数据抽取就是从数据源提取符合要求数据的过程,是数据仓库获取数据的第一步,也是关键的一步。

在进行数据抽取过程中,对数据仓库无用的源数据字段将被过滤掉,对不符合要求的类型与格式还要做相应的转换工作。目前常用的数据抽取方式有 PULL 和 PUSH 两种方式,PUSH 方式是由源系统按照双方定义的数据格式,将符合要求的格式抽取出来,再通过 FTP 或其它文件传送方式拷贝到 ETL 系统中,PULL 则是由 ETL 程序直接访问数据源,获取数据的方式。在数据仓库创建过程中一般有两种抽取模式:增量式装载和全量更新。增量式装载只加载自上次装载以后发生变化或者新增的数据。全量更新,主要解决在数据源中改变记录之后,未留下任何识别标志,在抽取时将需要的整个表提取出来。

数据抽取通常要注意的问题主要有:

- ◆ 源数据库和目标数据库的数据库格式一致性问题;
- ◆ 源数据库中要访问的表和文件;
- ◆ 源数据库中被提取字段、记录的条件问题;
- ◆ 目标数据库中表的结构等。

数据转换

数据转换是在数据仓库元数据指导下,对需要装入数据仓库的源系统中的每个记录进行转换。数据的转换过程不只是简单的数据格式改变,还要进行数据结构和内容转换。

数据进入数据仓库之前大致要经历如下过程:集成、渐变维度维护、参照完整性检查、规范化处理、计算处理、空值处理等,其处理过程如图 2-10 所示。

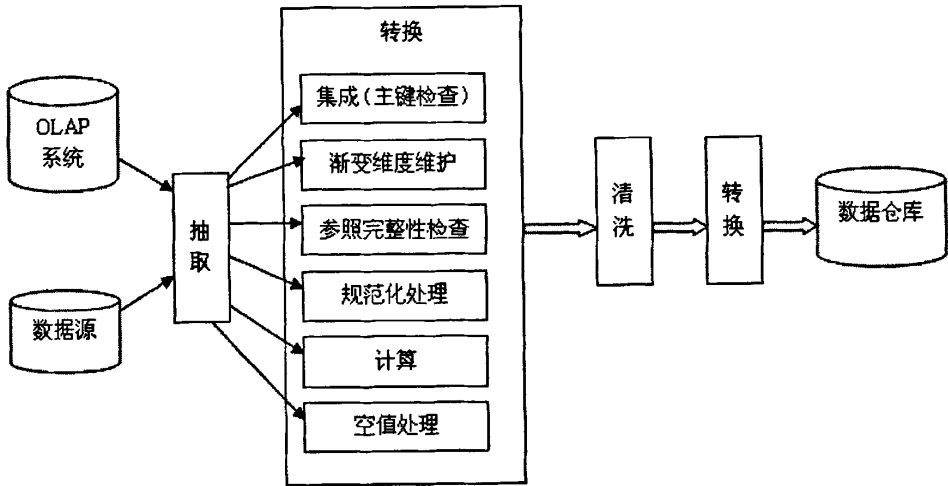


图 2-10 ETL 数据转换图

数据清洗

数据清洗是对进入数据仓库的数据进行全面检查并使其尽可能没有任何错误的过程。其目的在于提高数据仓库中的数据质量。

数据装载

数据经过转换、清洗之后，便可将其装入数据仓库中指定的主题和细节库中。在进行装载时可以使用 SQL 实现，也可以使用批量加载。很多数据库厂商都提供了支持其数据库的加载工具以及编程接口。

在数据加载时，根据情况可以使用数据仓库引擎厂商提供的数据加载工具进行数据加载或通过数据仓库引擎厂商提供的编程 API 进行数据加载。使用数据仓库引擎厂商提供的数据加载工具加载数据，会减少开发人员的负担，使用数据仓库引擎厂商提供的编程 API 进行数据加载会增加编程负担，但是这种方法数据加载性能会好一点，具体选择哪种方式要因具体情况而定。

在选择加载策略时主要考虑的问题主要有：加载周期和数据追加策略。加载周期是指多长时间从业务系统中抽取并向数据仓库中加载一次数据。数据追加策略是指数据每次是如何向数据仓库系统中追加的^[33]。

第3章 教务决策数据仓库模型研究

系统设计的第一步就是需求分析，教务决策数据仓库作为一个决策支持系统，当然在开始设计之前也要进行需求分析。

3.1 教务决策需求分析

3.1.1 教务决策需求特点

决策是一个复杂过程，不同行业有其不同特点。对于企业来说，其决策目的是为了降低成本和增加利润，那么其决策内容主要围绕产品、供应商、客户和销售等。对于高校来说，其决策的目标不是为了利润而是为了培养出更多优秀人才。人才培养与商品买卖不同，它是一个复杂过程，涉及到个人、家庭、社会、学校等诸多因素，这造成高校教学决策需求具有如下特点：

决策需求目标难以量化

高校各项管理决策的最终目标都是为了培养人才，人才培养是一个复杂过程，它受人为因素影响很大，这就导致教务决策需求很难进行定量描述，只能是定性描述。这样就造成教务决策需求分析难度较大，建立决策分析模型困难，难以划定相关数据等。

决策需求不稳定

对于供应链企业来说需要周期性对客户和销售量更改进行决策，但对于高校教学管理来说很少有需要周期性决策内容，在其决策中会出现这样的现象，有时很长时间内不需要决策，有时在很短时间内要对教务管理政策做出很大改动。比如，招生决策需求，一旦高考政策变化，其决策需求马上改变。

决策需求需要多方面数据支持

对于一个供应链企业来说，它的一个决策需求可能只涉及一段时间内所有客户交易信息。但对于高校教学决策来说，它的一个决策需求可能涉及到多方面信息。比如，在要决策是否调整某专业课程安排这一决策方面，要涉及到学生、教师、院系、学校等多方面数据。由此可知，高校决策需求所需数据具有数量大，时间跨度长，来源丰富等特点。

3.1.2 教务决策具体需求

构建教务决策数据仓库旨在帮助教务处决策层领导通过先进计算机信息技术,按照现代教学科学管理的原则、程序和方法,对教务工作的各项活动进行决策计划、组织、指挥、监督和协调,从而减少劳动耗费,并取得科学快速准确决策。

根据实际调研分析以及对高校教学管理制度分析,某高校教务决策需求主要涉及如下几个方面:

学籍管理决策分析

随着招生规模扩大、学分制实施以及综合班、国防生的出现,使得某高校休学、退学,转专业等人数与比例呈上涨趋势,给该校的学籍管理工作带来很大困难。针对这一问题,教务处决策层领导急需一个这样的系统,该系统能够:

①从部门、地区、民族、性别等多个角度对在校学生的基本信息进行动态分析,帮助决策层领导判断本校学生规模是否适当,结构是否合理,在调整来年招生计划,改善学生结构方面为教务处决策层的决策提供帮助。

②从部门、地区、民族、性别、异动原因、异动类别、异动日期角度对学生学籍情况进行动态统计分析,为合理调整专业结构、修订学籍异动制度的决策提供帮助。

③从多个角度对学生学籍异动情况进行实时监控与分析,及时发现学籍异动中出现的问题,为解决学籍异动中出现的问题制定正确指导方针提供帮助。

师资力量决策分析

在师资力量决策分析方面要求系统能够:

①对教师学历、年龄、职称、性别结构做出动态分析,为教师资源组成结构的合理性判断提供帮助。

②从学历、职称、年龄、性别角度对教师分布情况进行统计分析,判断教师分布结构的合理性,为学校制定人才培养、人才引进计划、优化教师队伍结构提供依据。

教学质量评估决策分析

在教学质量评估决策分析方面要求系统能够:

①通过从学历、年龄、职称、性别方面对教师教学质量的统计与分析,发现

影响教学质量的相关因素，为合理安排教学人员，提高教学质量提供决策依据。

②通过多角度的对教学信息系统中积累的教师教学评价数据进行汇总归类，分析教师教学优缺点，发现教师在教学中存在的问题，将信息及时反馈给教师，起到督促教师提高自身教学业务水平的效果。

教学工作量决策分析

在教师教学工作量方面要求系统能够：

①从教师年龄、性别、职称、学历、所授课程、日期角度对教师教学工作量进行统计分析，找出教师教学工作量与上述方面的相关性，为制定教师工作量安排计划提供帮助。

②通过对往年教学工作量的多维分析为制定和完善合理的工作量计算办法提供帮助。

课程选修决策分析

某高校在实行学分制以后，学生有了更多的选择，不再是以前的那种“套餐”形式，学校开设什么课程学生就学什么课程。在学分制下，学生有了更多的自由，可以根据自己的兴趣爱好进行课程、教师、上课时间、上课地点等选择，根据自身情况来定制个性化的学习计划。这样对学生来说在选课方面具有很大自主性，但是对于教务处来说，使得教学管理工作任务加重。在安排课程时必须全面考虑，否则有可能导致大面积课程停开、教师没课上等问题发生，进而影响到学校教学秩序的稳定。

在课程选修决策分析方面，要求系统能够：

①从学生、教师、课程性质、上课地点、上课时间角度对课程选修人数进行统计与分析，找出影响课程选修的相关因素，为制定课程安排计划提供帮助。

②通过对学生选修课程的多维分析，帮助校方了解到学生选课兴趣与选课趋势，结合学校的情况，制定相关课程选修措施，来调动学生学习积极性，促进教师自发的提高自身素质。

③通过学生选修课程多维分析结果的实时监控，能够帮助加强学生与管理人员之间信息交流沟通，及早发现学生选课问题，尽早采取有效措施，避免问题扩散，将不良影响降到最低程度。

成绩决策分析

成绩决策分析分为四六级成绩分析和公修课考试成绩分析。

对四六级成绩分析要求系统能够：

①按部门、地区、性别、来源地等对在校生四六级考试成绩进行多维分析，从中发现影响四六级考试通过率因素，为制定正确的英语学习指导计划提供帮助。

②通过按年级、学院、专业等逐级钻取分析，可以发现四六级一次或者多次通过的人数和具体学生，为全面了解学生的四六级考试情况提供支持。

公修课考试成绩分析方面要求系统能够：

①对公修课的考试成绩按学院、专业、性别、来源地、教师等进行多维分析，从中发现影响学生考试成绩的因素。为教学人员以及管理人员做出针对性解决方法提供帮助。

②通过按年级、学院、专业等逐级钻取分析，发现公修课考试及格人数和不及格人数，为全面了解学生公修课考试情况提供支持。

3.2 教务决策数据仓库系统结构

为解决上述教务决策需求，这里设计一个结构化的四层教务决策数据仓库系统，其体系结构如图 3-1 所示。

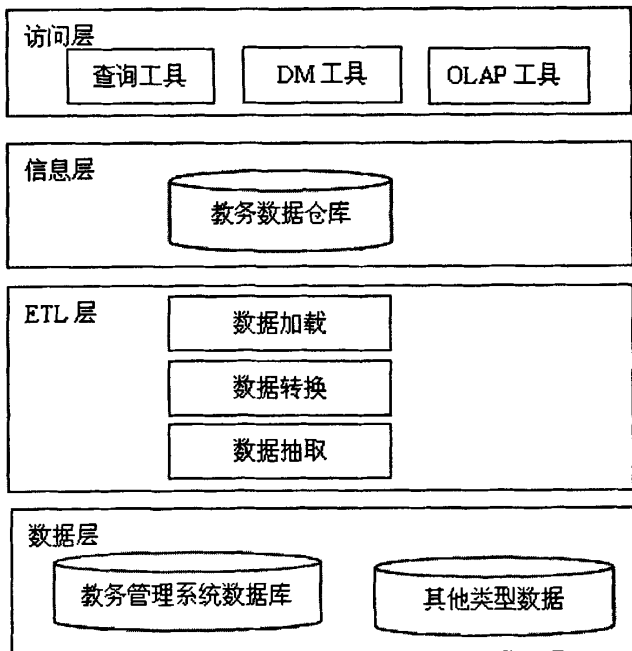


图 3-1 教务决策数据仓库系统结构示意图

数据层

源数据层是教务决策数据仓库系统体系结构最底层，它是构建教务决策数据仓库的基础。数据来源以教学信息系统的数据库为主，教学信息系统数据库包含教学工作日常操作所积累数据，它是决策主题所需数据的主要来源。

ETL 层

负责从教学信息系统中抽取数据，并将抽取的数据进行清理，转换处理，最终将其加载到教学数据仓库中去。

信息层

信息层是连接数据分析层和数据层之间的桥梁，对数据进行处理，形成不同层次信息，为分析层建模提供方便。本系统中信息层的核心是教务数据仓库。教务数据仓库通过对来自数据源的原始数据进行收集、分类、筛选和抽取最终形成能够对教务决策提供支持多重粒度级数据库，实现对数据源统一管理和信息转化。为分析层提供各种应用接口，构造应用于主题分析更高粒度级数据集市，或者是直接为数据访问层提供数据支持。

在信息层构建教务数据仓库为满足决策需要而建，其层次性应该由决策层次来定。教务数据仓库也是一个多粒度集数据集合，不同粒度层次体现了不同细节层次，通过对数据不同层次细节抽象，逐步把原始数据转化为对分析有用的信息。教务数据仓库的结构(如图 3-2 所示)与其它数据仓库类似，也包括早期细节集、基础细节集、轻度综合集和高度综合集。

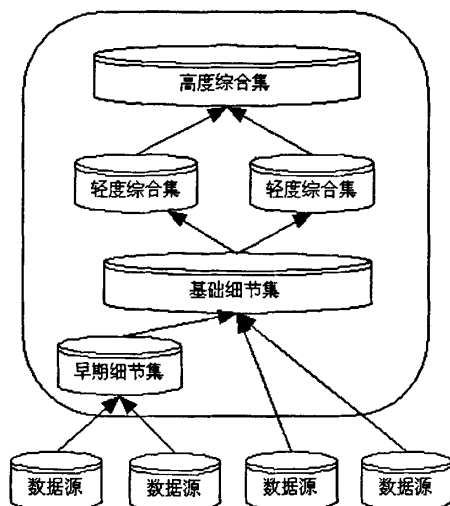


图 3-2 教务数据仓库结构示意图

早期细节集用来存放历史数据。基础细节集是数据仓库的主体,是基于数据驱动原则从数据源抽取、根据数据仓库概念和逻辑模型建立用于分析的数据集。基础细节集包含了全部对决策有用信息,是对数据源的全面信息描述和转化,是决策分析的根本依据。轻度综合集是和某具体主题相关的数据集市的整理和归档。而高度综合集是在轻度综合集基础上进一步抽取,结构化得到能够直接匹配数据挖掘算法的数据,往往是应用于更具体主题的数据集市或挖掘库在数据仓库中的归档。

从教务决策数据仓库系统数据仓库结构可以看出,数据仓库设计的重点在于基础细节级设计。基础细节集只需要一次性设计即可实现,其它细节集通过针对具体主题应用不断扩展。基础细节集设计思想是:首先通过分析数据源确定哪些是对决策有帮助的数据,然后对这些数据重新组织,形成概念和逻辑模型描述。

数据访问层

数据访问层是教务决策数据仓库系统提供给用户的接口,由基于主题的查询、OLAP 工具、数据挖掘工具等组成。针对教务决策需求不稳定问题,提供了一些基本多维分析手段和简单分析模型,这样就不必专门建立应用分析系统,可大大提高教务决策数据仓库系统的普适性和灵活性。对于复杂且相对稳定的决策需求,需要借助数据集市,开发专用分析系统。

3.3 教务决策主题设计

3.3.1 教务决策主题域获取

数据仓库中数据的组织方式是一种面向主题的组织方式,该方式能够让数据仓库在较高层次上对分析对象进行完整一致性描述,能刻画各个分析对象所涉及企业各项数据及它们之间的联系,数据仓库具有更高抽象级别。所以,主题设计是数据仓库设计的开始。

教务决策需求涉及的数据宽泛、决策需求目标难以量化、决策需求多元化以及时间不稳定因素等特点,使得教务决策数据主题呈现多元化特点,进而增加了教务决策主题设计的难度。

教务决策是一个复杂主题,通常情况下主要涉及以下主题:学生、教师、课程、成绩等。但是,如果把学生作为一个主题来设计,在教务数据仓库的后期设

计中会很难确定其属性和维度。为解决该类问题，人们引入了主题域这一概念。

根据教务决策实际需求分析，教务决策数据仓库使用SDLC方法开发，所以目前暂时为教务决策数据仓库设计学籍、课程、教师、成绩四个主题域。

3.3.2 教务决策主题域边界划分

所谓主题域就是对某个主题进行分析后确定的主题边界。对主题域分析就是确定要装载到数据仓库的主题的过程。

教务决策数据仓库构建旨在解决学籍管理决策分析、教学质量评估决策分析、师资力量决策分析、教学工作量决策分析、课程选修决策分析、成绩决策分析问题。

其中，学籍分析可分为基本学籍分析和学籍异动分析，成绩分析可分为四六级成绩分析和公修课成绩分析。根据以上分析，在教务数据仓库中主要包括如下主题：基本学籍主题、异动学籍主题、教学质量主题、师资力量主题、教学工作量主题、课程选修主题、四六级成绩主题、公修课成绩主题。

基本学籍主题、异动学籍主题同属于学籍分析主题；教学质量主题、师资力量主题、教学工作量主题属于教师主题；四六级成绩主题、公修课成绩主题属于成绩主题；课程选修是课程主题的子主题。学籍、课程、教师、成绩这四个主题又同属于教务决策这一大主题。由此，得出教务决策数据仓库各主题域边界划分，如所示 3-3 所示。

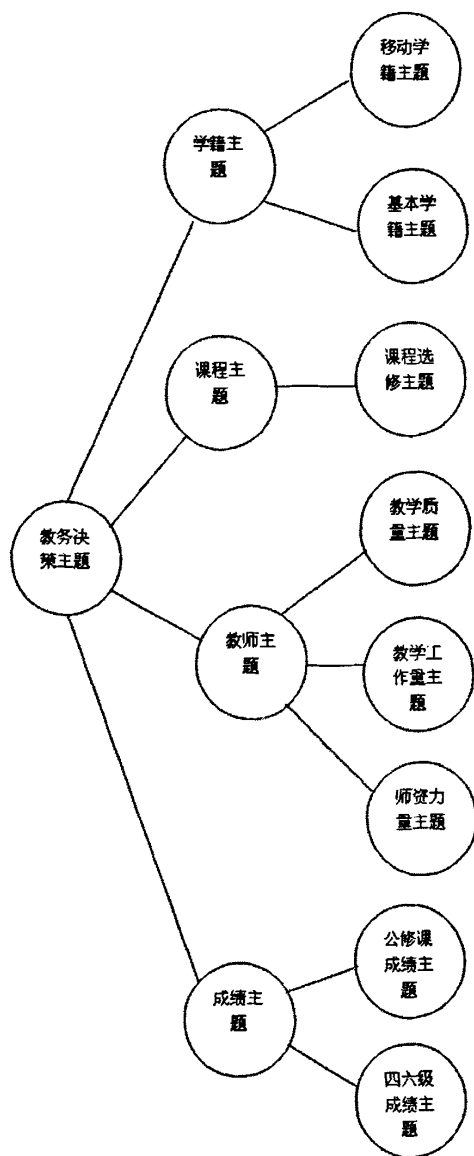


图3-3教务决策数据仓库主题域边界划分示意图

3.3.2 主题内容确定

数据仓库数据组织方式是一种面向主题的组织方式,教务数据仓库中的数据就是如图3-3所示所有叶子节点主题所包含数据的全体。为满足教务决策数据仓库对学籍管理决策分析、教学质量评估决策分析、师资力量决策分析、教学工作量决策分析、课程选修决策分析、成绩决策分析的决策需求,应为各主题确定不同内容,根据对教务决策的分析与对数据源的了解,教务决策数据仓库各主题将涉及的内容如表3-1所示。

表3-1教务决策数据仓库各主题内容表

主题名	主题内容
基本学籍	姓名、学号、院系、专业、来源地等
异动学籍	姓名、学号、异动类别、异动原因等
四六级成绩	学院、专业、状态、学号、考试次数等
公修课成绩	姓名、专业、学院、学号、授课教师等
师资力量	学历、年龄、职称、性别等
教学质量	学历、年龄、职称、性别、学生评分等
教学工作量	所受课程、课程性质、上课学时、教师信息等
课程选修	课程名字、上课教师、上课地点、学生信息等

3.4 教务决策数据仓库设计目标与原则

数据仓库是企业管理和决策中面向主题、集成、与时间相关和不可修改的数据集合。某高校教务决策数据仓库系统作为其数字化校园建设的一部分，其创建的目的在于将分散在不同校区、不同部门异构数据集中起来，达到数据高度共享度，解决教务部门对教学信息联机分析处理(OLAP)、数据挖掘、决策支持等问题。

在解决数据集成方面，数据仓库与虚拟的基于中间件技术不同，它是在物理层上实现异构数据源集成，因此，其设计要区别于其他集成技术的设计，在设计时要将不同来源的数据源分别设计为一个信息源单独处理，对各信息源分别进行ETL，之后形成最终一个统一信息集。教务决策数据仓库中三个独立信息源分别是：Oracle 信息源、Foxbase 信息源、文本信息源。

某高校在进行数字化校园建设中，通过教务决策数据仓库的创建解决异构数据源集成问题，其最终目的是为了满足不同决策需要。所以，教务决策数据仓库无论如何设计都要围绕服务决策、便于决策的原则来设计，但设计时也要注意存储空间与查询效率兼顾的原则。

结合某高校实际情况，某高校正在进行数字化校园建设工作，数字化校园建设是一个复杂过程，在数字化校园建设所需解决异构数据源集成问题，需要一个相对稳定后期维护容易的平台。鉴于此，我们使用有反馈的自顶向下的构建策略（其构建过程如图 3-4 所示）来构建教务决策数据仓库。

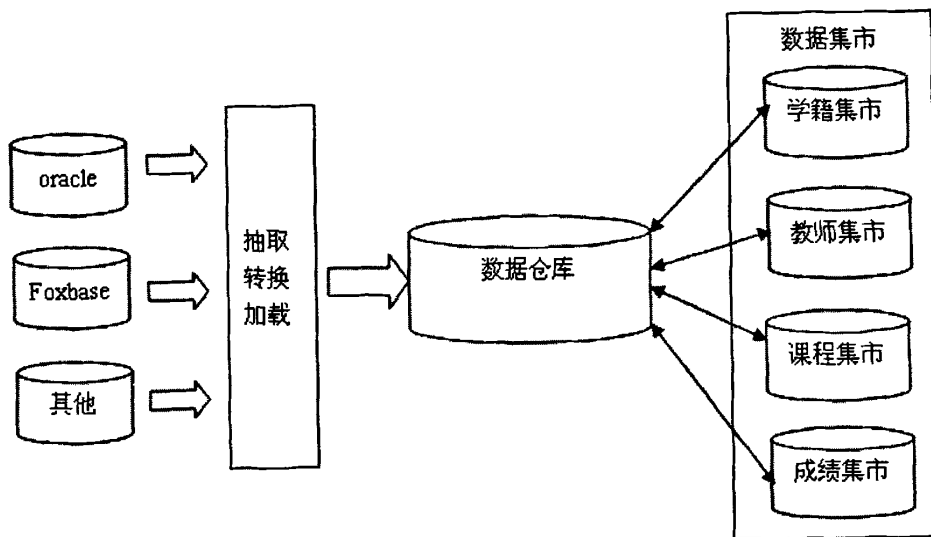


图 3-4 教务决策数据仓库构建过程图

3.5 教务决策数据仓库模型

在数据仓库中维是用来描述决策分析人员分析数据角度的。因此，在数据仓库中维度设计也就是完成决策者在分析数据时所站的分析角度的设计。

3.5.1 维表设计

对基本学籍分析时，从学院、专业、年级、民族、性别、生源地对学生信息分析，所以在数据仓库中应设计部门维表、民族维表、地区维表、地区维表。

对异动学籍分析时，除从学院、专业、年级、民族、生源地角度分析外还从不同时期，引起学籍异动原因，以及异动类别方面对异动学籍进行分析，所以除设计部门维表、民族维表、地区维表外，还应该设日期维表、异动维表。

师资力量分析是从教师学历、年龄、职称、性别、部门、来源地角度分析，设计时应设学历维表、年龄维表、职称维表，性别维表、部门维表、地区地维表。

对教学质量分析时从教师学历、年龄、职称、性别、所授课程分析，应为其设计学历维表、年龄维表、职称维表、性别维表、课程维表。

对教学工作量分析时，从教师、课程、部门、日期角度分析，应为其设计教师维表、课程维表、部门维表、日期维表。

对四六级成绩分析时从学院、专业、性别、来源地、民族角度分析，所以为其设计学院维表、专业维表、性别维表、地区维表、民族维表。

对公修课成绩分析时从学院、专业、性别、来源地、民族、教师年龄、教师职称、教师性别、教师学历、教师所在部门角度分析，所以应为其设计部门维表、性别维表、地区维表、民族维表、年龄维表、职称维表、性别维表、学历维表。

对课程选修分析时从学生所在学院、专业、性别、民族、来源地以及教师职称、学历、性别、上课地点、上课时间，课程性质分析，所以应设计性别维表、民族维表、地区维表、职称维表、学历维表、地点维表、时间维表、课程维表。

结合某高校实际教务决策需求，为各分析主题设计完维表之后，我们发现有些维表重复现象，删除重复维表之后得到目前教务决策数据仓库维表：部门维表、民族维表、性别维表、地区维表、日期维表、异动维表、学历维表、年龄维表、职称维表，地点维表、时间维表、课程维表。各维表所包含信息如表 3-2 所示。

表3-2教务数据仓库各维表信息明细表

维表名称	包含信息
部门维表	部门编号、学院、专业、年级
民族维表	民族编号、民族类别、民族名称
性别维表	性别编号、性别名
地区维表	地区编号、区域、省、市
日期维表	日期编号、学年、学期
异动维表	异动编号、异动原因、异动类别
学历维表	学历编号、学历名称
年龄维表	年龄编号、年龄类别
职称维表	职称编号、职称名
地点维表	地点编号、校区、课室、教室
时间维表	时间编号、学年、学期、星期、天、节次
课程维表	课程编号、课程性质、课程类别

3.5.2 事实表设计

在数据仓库中每一分析主题对应一个事实，教务决策数据仓库为教务决策所设计，所以在教务决策数据仓库中设计如下事实表：基本学籍事实表、异动学籍事实表、教学质量事实表、师资力量事实表、教学工作量事实表、四六级成绩事

表 3-3 教务决策仓库各事实表明细表

事实表名	表所包含内容
基本学籍	学号、姓名、性别编号、地区编号、民族编号、部门编号
异动学籍	学号、异动编号、民族编号、地区编号、部门编号、性别编号、日期编号、异动前学院、异动前专业、异动后学院、异动后专业
教学质量	教学质量编号、教师编号、学生评分数、评价等级
师资力量	教师编号、姓名、职称编号、学历编号、年龄编号、性别编号、地区编号、部门编号
教学工作量	工作量编号、教师编号、所授课程名字、课程编号、课时数
四六级成绩	成绩编号、学号、状态、分数、考试次数
公修课成绩	公修编号、学号、课程名、成绩、教师编号、时间编号、地点编号

3.5.3 教务决策数据仓库模型

为完成基本学籍分析我们为其设计了部门维表、民族维表、地区维表、地区维表，部门维表、民族维表、地区维表，这些维表以基本学籍事实表为中心分布在四周便形成了基本学籍主题星型结构模型，如图 3-5 所示。

在异动学籍分析时要用到部门维表、民族维表、地区维表外、性别维表、日期维表和异动维表，考虑到异动学籍事实表行数不多，可以在学籍异动事实表添加相关字段直接使用部门维表、民族维表、地区维表、性别维表，达到提高数据查询速度效果，这样对于异动学籍主题来说其结构也是星型模型，如图 3-6 所示但是对于由基本学籍主题和异动学籍主题两个主题构成的学籍主题域来说就形成了星座模型，如图 3-7 所示。

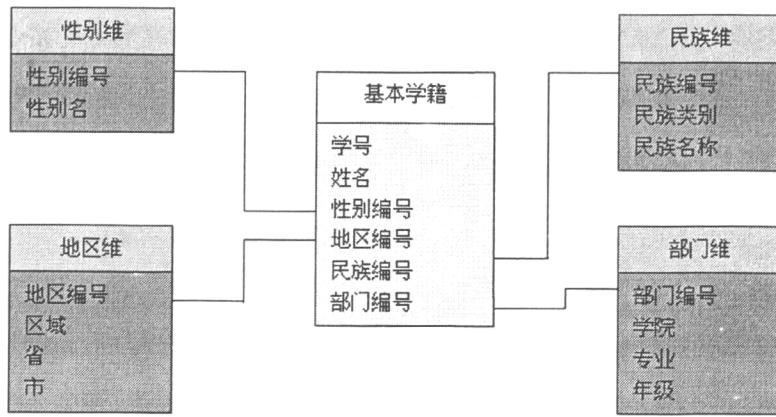


图 3-5 基本学籍主题星型结构模型示意图

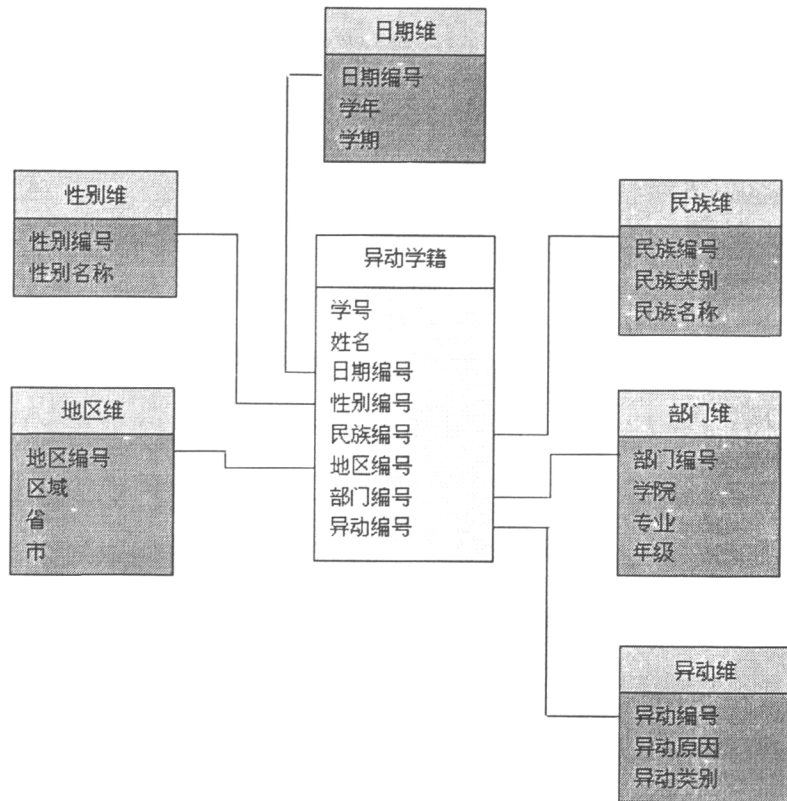


图 3-6 异动学籍主题星型模型示意图

在进行师资力量分析时我们设计了学历维表、年龄维表、职称维表，性别维表、部门维表、地区地维表，对于师资力量主题采用星型模型 3-8。

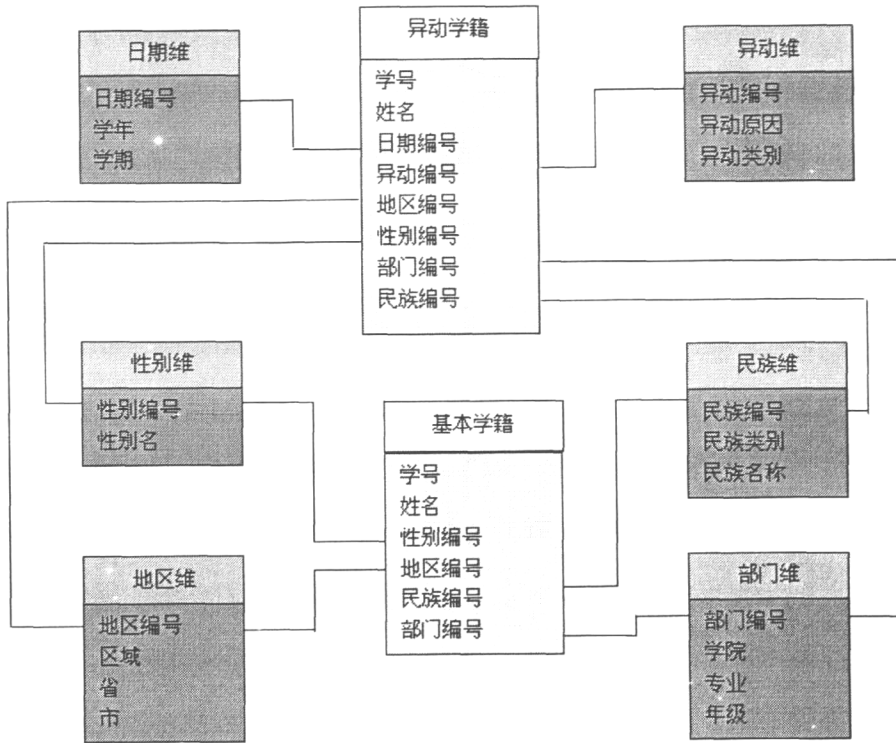


图 3-7 基本学籍主题和异动学籍主题的星座模型示意图

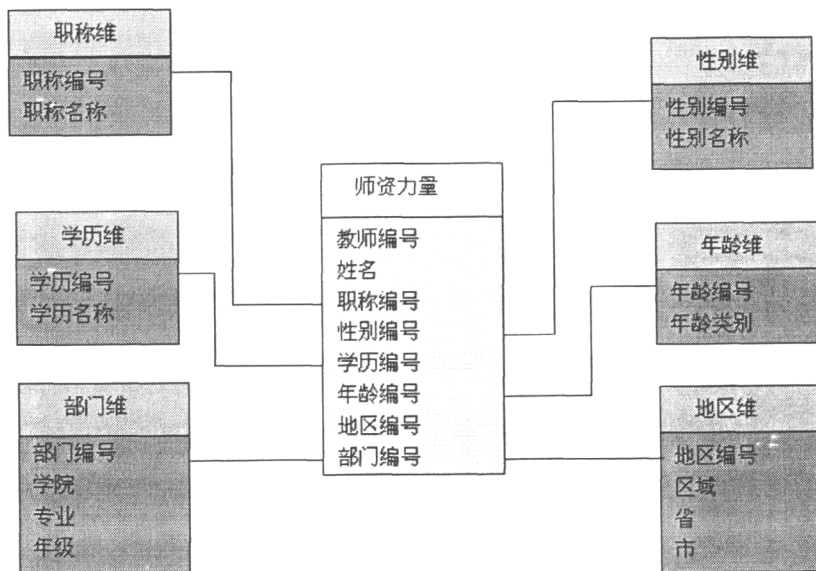
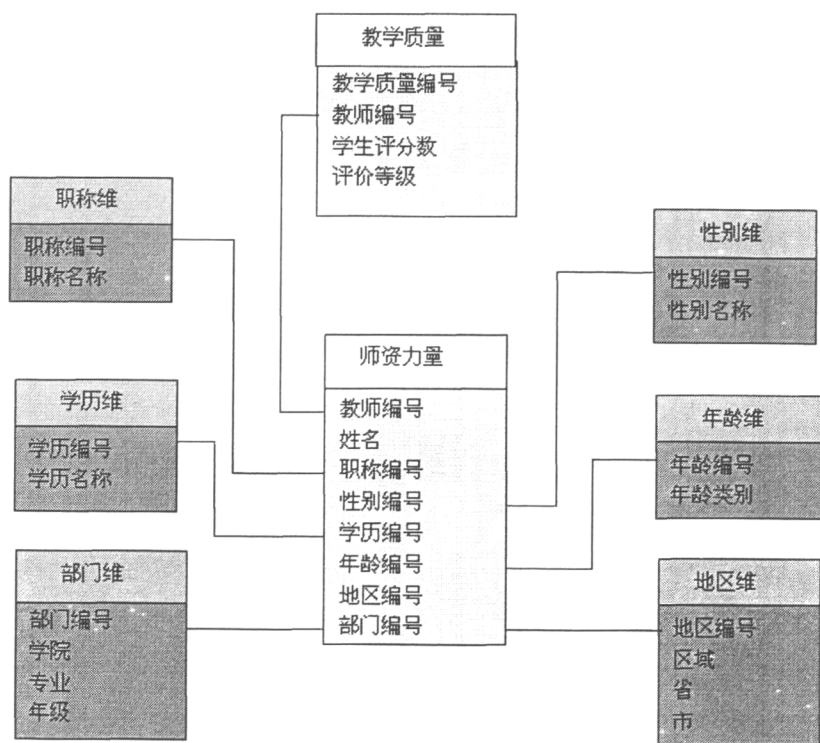
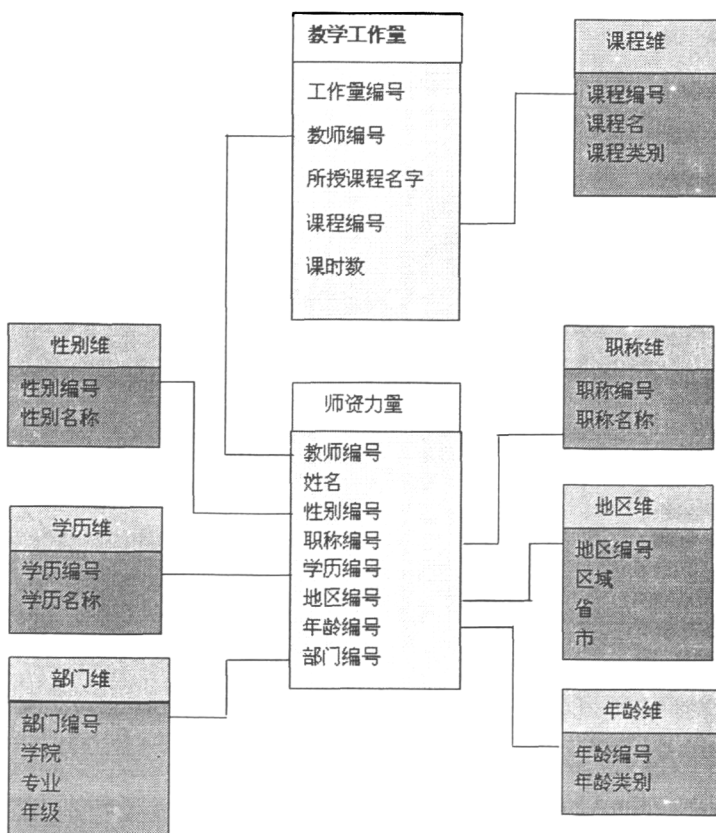


图 3-8 师资力量主题星型模型示意图

在进行教学质量分析时要用到学历维表、年龄维表、职称维表、性别维表、课程维表，教务决策数据仓库中教学质量评价是一个比较大的表，并且每年以将近 20 万行的速度增加，所以，如果教学质量主题仍然采用星型模型，会在教学质量事实表中增加很大的冗余，并且学历维表、年龄维表、职称维表、性别维表均与师资力量事实表直接相连，鉴于此对于教学质量主题最好采用雪花模型，其结构如图 3-9 所示。



对于教学工作量主题的分析主要从教师的学历、职称、性别、年龄、部门、所授课程角度分析，这里将教师作为维表，采用了雪花模型，如图 3-10 所示。



公修课成绩、四六级成绩事实表、课程选修事实表都是很大的数据表。如果对于这几个主题使用星型模型会给整个数据仓库带来巨大的冗余，所以，对于它们我们都采用了雪花模型，公修课成绩主题的雪花模型如图 3-11 所示，四六级成绩主题的雪花模型如图 3-12 所示，课程选修主题的雪花模型如图 3-13 所示。

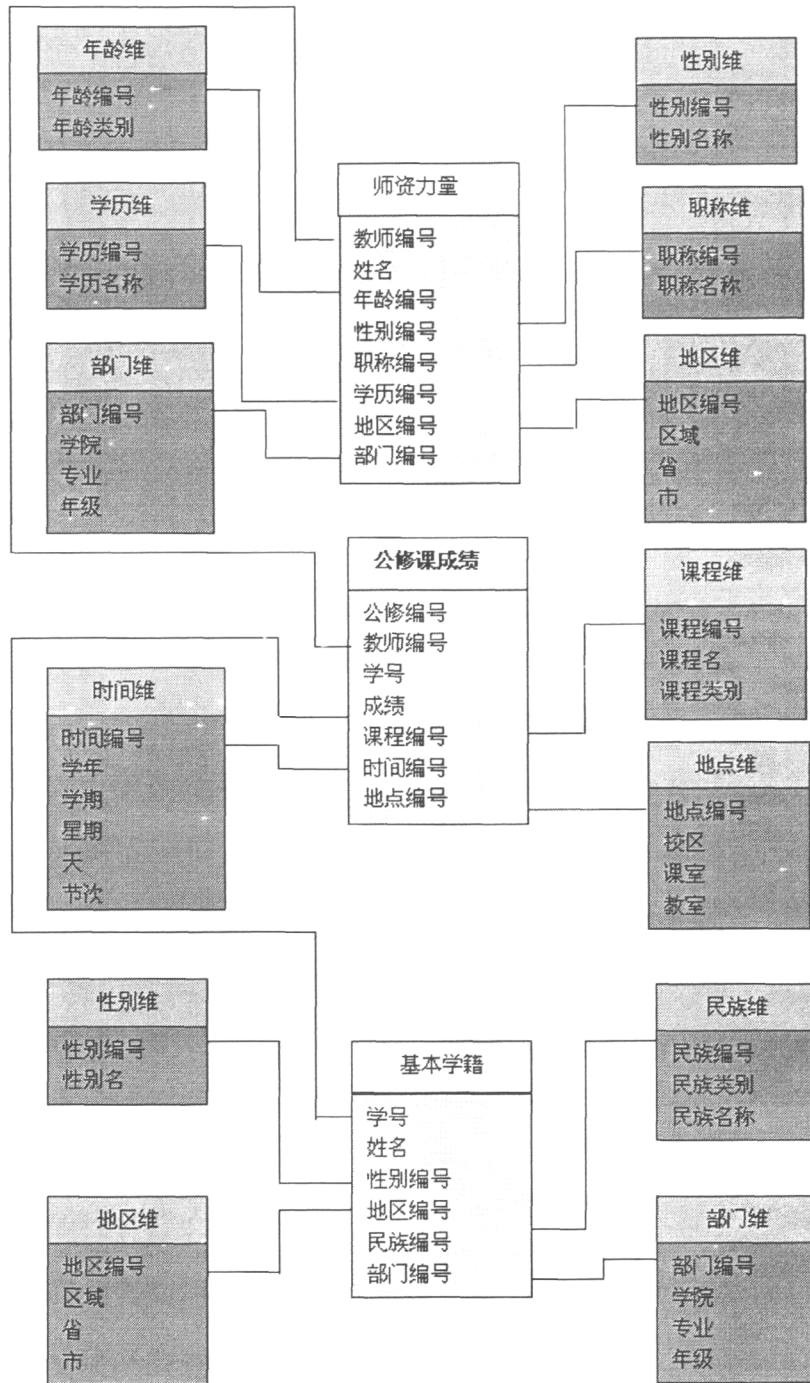


图3-11公修课成绩主题星型模型示意图

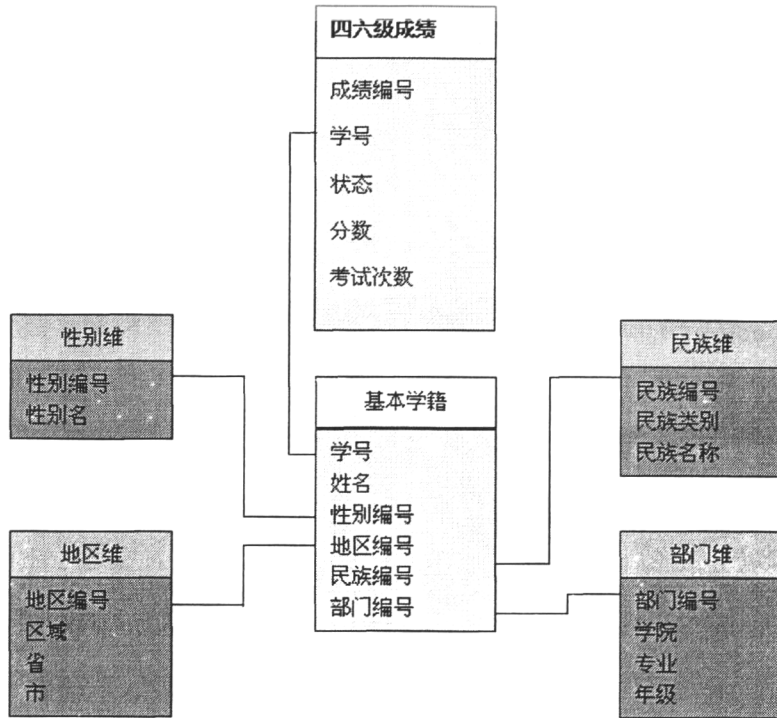


图3-12四六级成绩主题雪花模型示意图

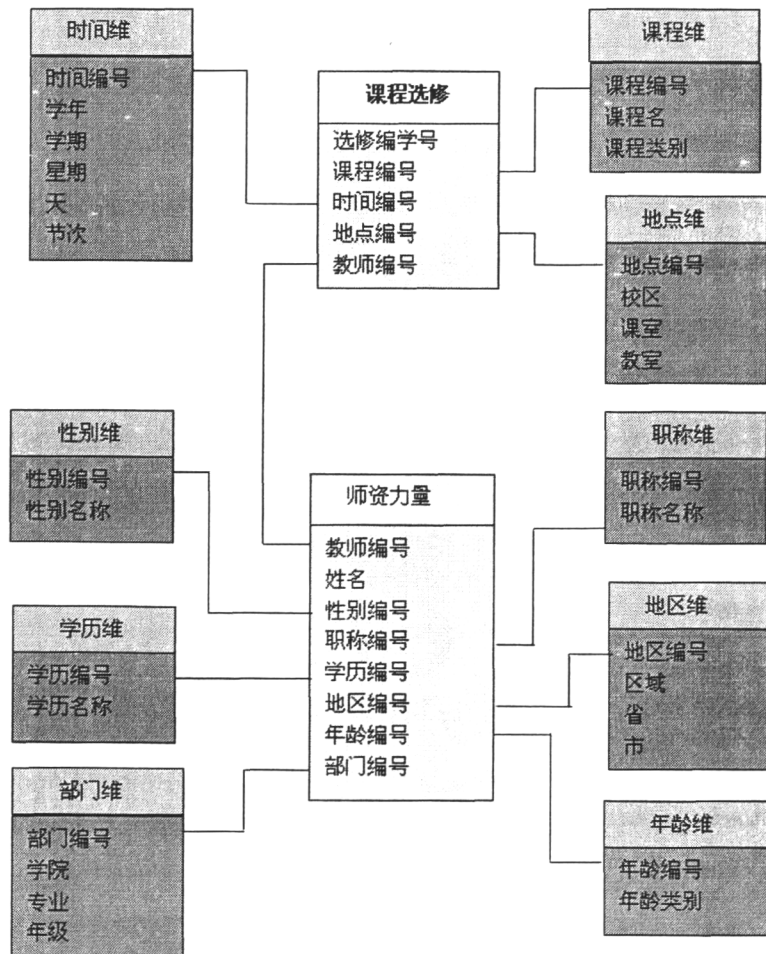


图3-13课程选修主题雪花模型示意图

第4章 教务决策数据仓库设计

在教务决策数据仓库整个设计过程中基础细节集设计可以在一次设计中完成，其它细节集会随着应用主题不断扩展而发生变化，因此，教务决策数据仓库设计重点应放在基础细节集上。教务决策数据仓库基础细节集设计过程是先完成对数据源的分析，再确定对决策有用的数据，最后重新组织这些数据并形成概念模型与逻辑模型。因此，数据源的理解与分析是教务数据仓库设计的开始。

4.1 数据源的理解与分析

某高校构建教务决策数据库，主要以基于Oracle数据库数据、基于Foxbase数据、基于文本格式教务管理信息为数据源。三个数据源格式不同，而且数据仓库从他们那里抽取数据后，数据的更新方式也不同，所以设计时将他们分别设计为独立的信息源。

三个数据源中，基于Oracle数据库的数据源是教务决策数据仓库的主要数据源。根据决策需求分析知，在Oracle数据库中1300多张表中对教务数据仓库有用的数据表主要有：学生基本信息表(XSJBXXB)、成绩表(CJB)、课程代码表(KCDMB)、学生选课表(XSXKB)、学生课程表(XSKCB)、学生学期排名结果表(XSXQPMJGB)、教师评价表(JSPJB)、教师基本信息表(JSJBXXB)、学籍异动表(XJYDB)。限于篇幅这里只给出学生基本信息表和学籍异动表的表结构，学生基本表的表结构如表4-1所示，学籍异动表的表结构如表4-2所示。

表4-1学生基本信息表表结构

字段名称	含义	类型	长度	是否为空	默认值	主键	外键	约束
XH	学号	Varchar2	20	N		P		
XM	姓名	Varchar2	30	Y				
XB	性别	Varchar2	2	Y				
CSRQ	出生日期	Varchar2	10	Y				
ZZMM	政治面貌	Varchar2	12	Y				
MZ	民族	Varchar2	15	Y				
JG	籍贯	Varchar2	20	Y				
LYDQ	来源地区	Varchar2	30	Y				
XY	学院	Varchar2	30	Y				

X1	系	Varchar2	30	Y				
ZYMC	专业名称	Varchar2	40	Y				
XZB	行政班	Varchar2	50	Y				
XZ	学制	numeric	1	Y				
XXNX	学习年限	numeric	2	Y				
XJZT	学籍状态	Varchar2	20	Y				
DQSZJ	当前所在级	numeric	4	Y				
PYFX	培养方向	Varchar2	20	Y				
ZYFX	专业方向	Varchar2	20	Y				
ZYLB	专业类别	Varchar2	20	Y				
RXRQ	入学日期	Varchar2	10	Y				
BYZX	毕业中学	Varchar2	30	Y				
SSH	宿舍号	Varchar2	50	Y				
DZYXDZ	电子邮箱地址	Varchar2	40	Y				
LXDH	联系电话	Varchar2	30	Y				
ZKZH	准考证号	Varchar2	20	Y				
SFZH	身份证号	Varchar2	20	Y				
GATM	港澳台码	Varchar2	10	Y				
JKZK	健康状况	Varchar2	10	Y				
YWXW	英文姓名	Varchar2	2	Y				
BZ	备注	Varchar2	250	Y				
MM	密码	Varchar2	20	Y				
BDH	报到号	Varchar2	8	Y				
YYDJ	英语等级	Varchar2	1	Y				
KH	卡号	Varchar2	30	Y				
ZSSJ	招生时间	Varchar2	6	Y				
RXZF	入学总分	Varchar2	8	Y				
XSLB	考生类别	Varchar2	2	Y				
XMPY	姓名拼音	Varchar2	10	Y				
YHZH	银行帐号	Varchar2	50	Y				
DJ	等级	Varchar2	20	Y				
JXBMC	教学班名称	Varchar2	50	Y				
RDSJ	入党时间	Varchar2	20	Y				
ZYDM	专业代码	Varchar2	4	Y				
KJBYM	录检表页码	Varchar2	20	Y				
SFKZC	是否可注册	Varchar2		Y				
CC	层次	Varchar2		Y				
YZBM	邮政编码	Varchar2		Y				
SFGSPYDY	是否高水平运动员	Varchar2		Y				
XSZH	学生证号	Varchar2		Y				
USERID	WEB 登录时限	Varchar2		Y				

	制用户名							
KSLB	考生类别	Varchar2		Y				
SFZC	是否注册	Varchar2		Y				
YDLB	异动类别	Varchar2		Y				
XYM	曾用名	Varchar2		Y				
BYRQ	毕业日期	Varchar2		Y				
CSD	出生地	Varchar2		Y				
LJB	录检表	Varchar2		Y				
LJBC	录检表册	Varchar2		Y				
LJBH	录检表行	Varchar2		Y				
DZZCH	电子注册号	Varchar2		Y				
KSH	考生号	Varchar2		Y				
CCQJ	乘车区间	Varchar2		Y				
DLMC	大类名称	Varchar2		Y				
KSTZ	考生特征	Varchar2		Y				
TC	特长	Varchar2		Y				
RXFS	入学方式	Varchar2		Y				
SFZDS	是否走读生	Varchar2		Y				
BXXS	办学形式	Varchar2		Y				
BXLX	办学类型	Varchar2		Y				
XXS	学习形式	Varchar2		Y				
ZSJJ	招生季节	Varchar2		Y				
SHBJ	审核标记	Varchar2		Y				
Y SHBJ	预审核标记	Varchar2		Y				
ZXWYYZ	主修外语语种	Varchar2		Y				

表4-2 学籍异动表的表结构

字段名称	含义	类型	长度	是否为空	默认值	主键	外键	约束
YDXH	异动序号	number	10	NOT		P		
XH	学号	VARCHAR2	20					
XM	姓名	VARCHAR2	30					
XB	性别	VARCHAR2	2					
YDLB	异动类别	VARCHAR2	20					
YDYY	异动原因	VARCHAR2	50					
YDSJ	异动时间	VARCHAR2	10					
YDSM	异动声明	VARCHAR2	200					
YDQXY	异动前学院	VARCHAR2	30					
YDQXI	异动前系	VARCHAR2	30					
YDQZY	异动前专业	VARCHAR2	40					
YDQXZ	异动前学制	NUMBER	1					

YDQZYFX	异动前专业方向	VARCHAR2	20					
YDQPYFX	异动前培养方向	VARCHAR2	20					
YDQXZB	异动前教学班	VARCHAR2	50					
YDQXJZT	异动前学籍状态	VARCHAR2	10					
YDHXY	异动后学院	VARCHAR2	30					
XQ	异动学年	NUMBER	1					
XN	异动学期	VARCHAR2	10					
YDHSZNJ	异动前所在年级	NUMBER	4					
YDQSZNJ	异动后所在年级	NUMBER	4					
YDHXJZT	异动后学籍状态	VARCHAR2	10					
YDHXZB	异动后教学班	VARCHAR2	50					
YDHPYFX	异动后培养方向	VARCHAR2	20					
YDHZYFX	异动后专业方向	VARCHAR2	20					
YDHXZ	异动后学制	NUMBER	1					
YDHZY	异动后专业	VARCHAR2	40					
YDHXI	异动后系	VARCHAR2	30					
YDQZYDM	异动前专业代码	VARCHAR2	10					
YDHZYDM	异动后专业代码	VARCHAR2	10					

4.2 教务决策数据仓库概念模型设计

在多维建模中常用的建模工具有UML和信息包图。信息包图是一种简单直观自上而下数据建模方法，具有简单、高效特点。教务决策数据仓库采用自顶向下策略构建，所以这里选择信息包图为教务决策数据仓库建模。

采用信息包图为教务决策数据仓库进行概念模型设计需要如下步骤：定义名称、定义度量、定义维度、定义层次类别。

从教务决策数据仓库的概念模型设计阶段所获得的是信息包图，信息包图建立工作结束标志着教务数据仓库概念模型完成。

限于本数据仓库设计主题较多，这里以学籍主题域中基本学籍主题的信息包图的设计为例，给出基本学籍主题的信息包图的设计过程。

为表示方便，这里将基本学籍主题信息包图名称设为基本学籍（其它主题的信息包图的名称定义也是使用本主题的名字命名），将人数定义为度量。

根据3.2.1需求分析可知，在对基本学籍主题进行分析时，主要从性别、地区、民族、部门角度分析，所以在基本学籍信息包图中应该确定性别、地区、民族、部门四个维度。

根据对数据源的分析与教务决策的需求分析，应分别为上述四个维度定义如表4-3所示的维成员。

表4-3学籍主题相关维度信息表

维度名称	维成员
性别	性别名
地区	区域、省、市
民族	民族类别、民族名称
部门	学院、专业、年级

对于性别、地区、民族、部门四个维度，其中性别、民族都只有一个维成员不存在层次关系，对于地区、部门两个维度维成员之间存在类别层次关系，这里要为它们定义层次类别关系。其中地区维度的维成员中区域层次最高，其下一层是省、市，其类别层次定义如图4-1所示。按照类似方法，部门维度类别层次定义如图4-2所示，完成类别层次定义之后，按照各维度维成员的层次关系排列各维度维成员，即可形成基本学籍信息包图，如图4-3所示，按照同样方法可以为其他主题创建信息包图，其他主题的信息包图请参看附录一。

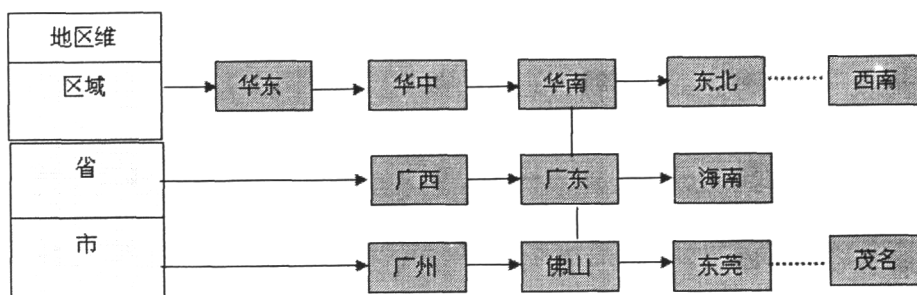


图4-1地区维度类比层次定义图

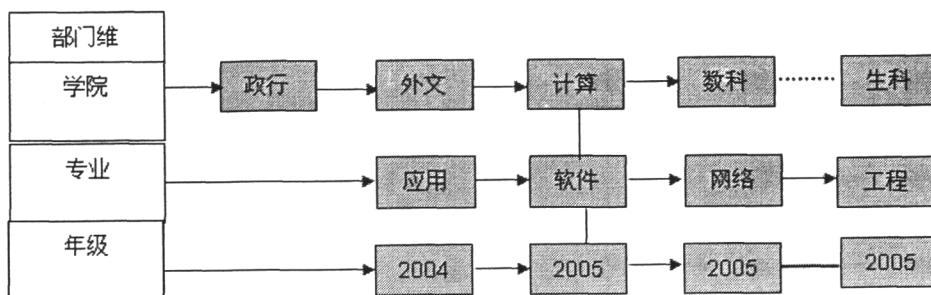


图4-2部门维度类别层次定义图

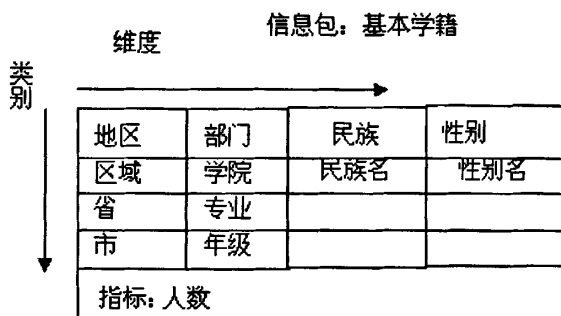


图4-3基本学籍信息包图

4.3 教务决策数据仓库逻辑模型设计

主题信息包图建立完成了教务决策数据仓库中该主题的概念模型设计，根据数据模型的精练过程，还需要将概念模型细化为逻辑模型，即需要对所得的信息包图进行转换。

信息包图可以为星型图设计提供完备概念基础，而星型图的逻辑实体又与信息包图的对象相互对应，可以很好地表述信息包图；星型图支持教务决策者的观点，将各个逻辑实体组织成适于处理的形式，便于不具备计算机专业知识的业务专家理解和使用教务决策数据仓库。所以，在教务决策数据仓库的逻辑设计中，应先将信息包图转化为星型包图，之后再规范化处理得到雪花模型。

与信息包图相对应，星型图由五类逻辑实体组成：指标逻辑实体、维度实体、层次实体、详细信息实体和类别实体。其中，度量逻辑实体位于星的中心，维度逻辑实体位于星的各个角上，维度逻辑实体向外延伸，与层次逻辑实体、详细信息逻辑实体和类别逻辑实体相连。

由于教务决策数据仓库涉及的主题较多，这里只以基本学籍主题为例介绍有关基本学籍主题由信息包图到星型图的转换过程。

4.3.1 指标实体设计

由基于信息包图的数据仓库概念模型向星型图转化的过程是：①将信息包图中的度量对象转换为其星型图中的度量逻辑实体，并将度量逻辑实体用矩形框来表示，定位于星型图中央；②将信息包图中各个维度对象转换为其星型图中相应的维度逻辑实体，并将维度逻辑实体用菱形框表示，同时定位于星型图的角上；③考察信息包图中每一维度对象所包含的层次和类别。

信息包图的指标对象对应着星型图中的数据指标，而信息包图中每个维度的最低级类别可以纳入逻辑指标中，这样得到的数据指标和逻辑指标共同构成了星型图中的指标实体，该指标实体位于星型图的中央。

在基本学籍信息包图中地区维度、部门维度、民族维度、性别维度的指标实体分别为地区编号、部门编号、民族编号、性别编号，将上述四个指标实体放入一个矩形内，基本学籍主题将人数定义为度量，根据信息包图到星型图的转换过程，这里的人数应该对应将来转换得到基本学籍主题星型图的数据指标，将数据指标放到另外一个矩形内，之后将两个矩形内的指标实体再统一放到一个矩形内，这样就形成了基本学籍主题的指标实体，如图4-4所示，用类似的方法可以为其它主题建立指标实体，其他主题的指标实体参看附录二。

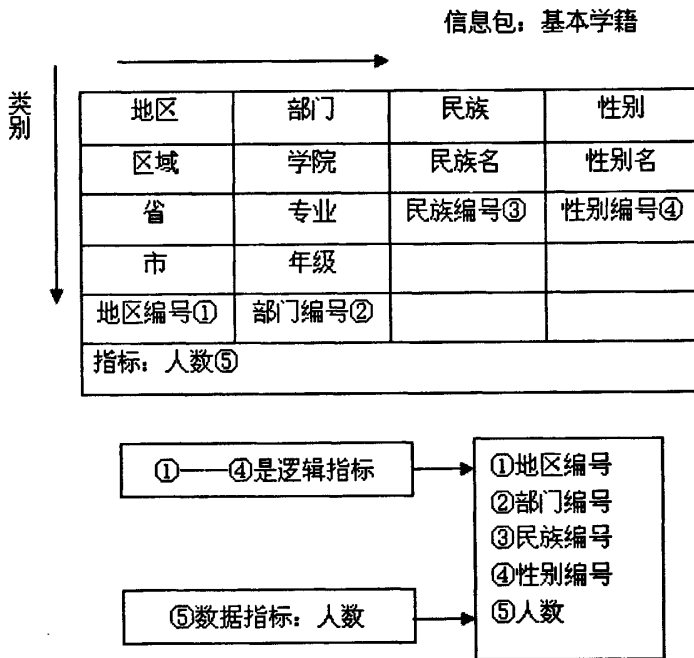


图4-4基本学籍主题指标实体示意图

4.3.2 维度实体设计

维度实体设计本质的就是将信息包图各个维度对象转化为星型图维度实体，以菱形形式表示，将来要将其置于星型图角上，在基本学籍主题中主要有地区、部门、民族、性别四个维度，分别将四个维度的维成员依次放到菱形中去，便完成了基本学籍主题各维度实体的设计，基本学籍各维度设计设计结果如图4-5所

示，其他主题的维度实体见附录三。

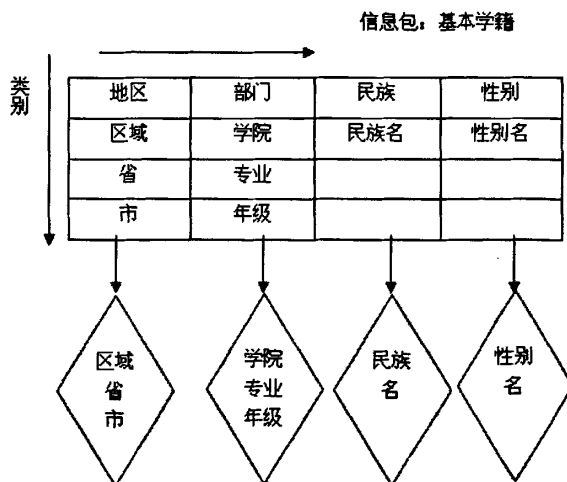


图4-5基本学籍主题维度实体设计结果图

4.3.3 指标实体与维度实体关系确定

维度实体总是要在指标实体中表示，它们之间的关系是通过逻辑模型中的详细类别定义，是一种典型的一对多关系。在星型图中，每个维度实体通过最底层的详细类别实体和指标实体中对应的逻辑指标相连接。在基本学籍主题中地区维度的详细类别实体是地区编号，部门主题的详细类别实体是部门编号，民族实体的详细类别是民族编号，性别实体的详细类别是性别编号。由此，可以确定基本学籍主题的指标实体与地区维度实体、部门维度实体、民族维度实体、性别维度实体之间的连接关系分别是地区编号、部门编号、民族编号、性别编号，在星型图中用结束符号表示指标实体与维度实体之间关系，这样，基本学籍主题指标实体与维度实体之间关系确定结果如图4-6所示，其他主题的指标实体与维度实体之间的确定结果参见附录四。

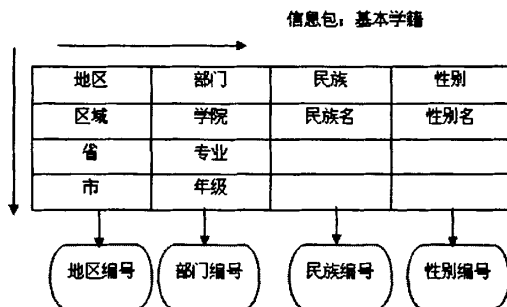


图4-6基本学籍主题指标实体与维度实体之间关系确定结果图

4.3.4 信息包图到星型图的转换

确定完指标实体与维度实体之间的关系后，将地区维度实体、部门维度实体、民族维度实体、性别维度实体分别通过他们之间的关系连接到指标实体，就得到了基本学籍主题的星型结构图，如图4-7所示，其他主题的星型结构图参见附录五。

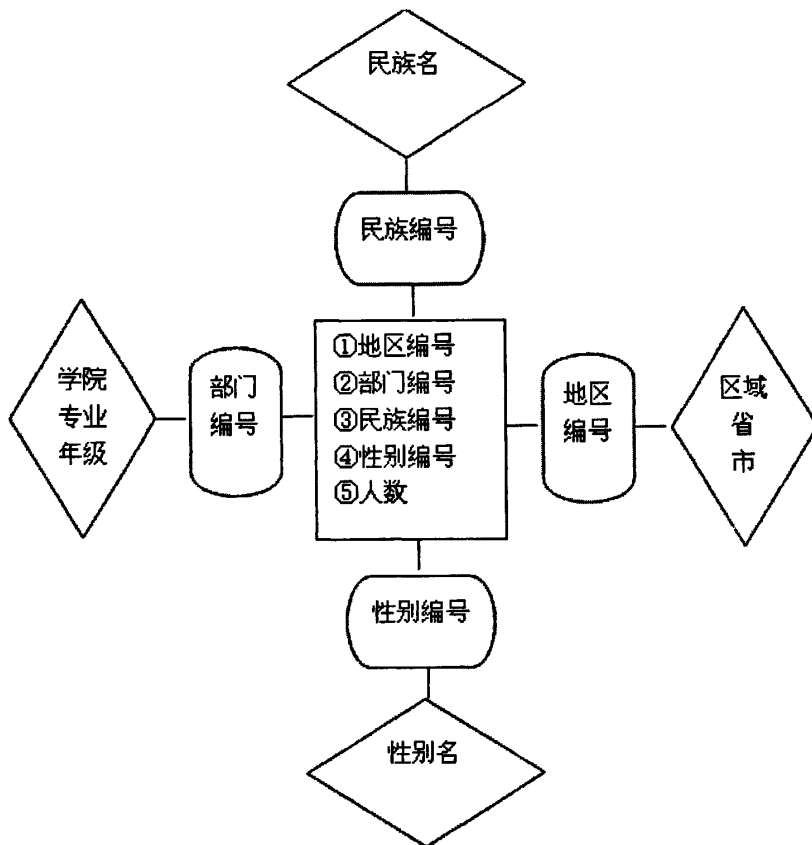


图4-7基本学籍主题星型结构图

完成主题由信息包图到星型图转化之后，我们根据需要进行相应规范化处理得到教务决策数据仓库的雪花模型。

4.4 教务决策数据仓库数据粒度设计

在数据仓库中粒度描述了数据单元的详细程度和级别。数据粒度越小，其级别也就越低，详细程度就越高；相反数据粒度越大，其级别也就越高，综合度就越高，详细程度就越低。在数据仓库中粒度设计至关重要，它将影响到数据仓库的逻辑设计、数据存储以及将来的分析效果。

目前教务决策数据仓库中主要涉及到早期细节集、当前细节集、轻度综合集

三个粒度级别。来源于正方教务管理系统中的数据以及来自于Foxbase的教务管理系统中的数据和其他类型的数据经过抽取、清洗、转换后先进入到当前细节集，之后根据决策分析需求再进入到轻度综合集。当前细节集中老化的数据，可将其转异到早期细节集中。

教务决策数据仓库设计主要用途在于教务决策分析，因此，其数据查询大多是在一定程度综合数据上进行，很少涉及到细节数据查询。此外，教务决策数据仓库与教务管理信息系统数据库相比，其数据量非常庞大。因此，对于教务决策数据仓库中数据必须认真分析，为其设计合适粒度。

长期以来在数据仓库中性能和存储空间一直是一个不可调和的矛盾，又是所有数据仓库设计者所必须要解决的问题。

在数据仓库设计中，如果把数据粒度设计的过小，将会使数据仓库的储存空间很大；若把数据粒度设计过大，将会导致数据决策无法观察细节数据。在进行粒度设计时应认真把握，可遵循如下原则：①权衡粒度级别，对于业务量大，分析要求比较高的情况下，应采用多重粒度形式；②对于某个具体事实来说，应采用“最小粒度原则”，将量度的粒度设置到最小。

教务决策数据仓库主题域主要有学籍、教师、课程、成绩四大主题域。结合对数据源的分析，在教务决策数据仓库中主要设计数据表有：学生基本表(xsjbb)、学籍变动表(xjdbd)，教师基本表(jsjbb)、教师工作信息表(jsgzb)，学生评分表(xspfb)，学生选课表(xsxkb)、课程成绩表(cjb)、四六级成绩表(sljcb)等。

在学籍主题域中学生基本表是最大表，该表主要用来保存所有学生的基本信息，该表与数据源中学生基本信息表对应，目前学生基本信息表中数据总行数为19285行，数据仓库中的学生基本表与该表对应，数据仓库中学生基本表的数据增长速度主要取决于每年的招生规模，根据对近5年来每个年度招生人数分析，某校每年全日制本科生招生规模基本稳定在5000人左右，那么学生基本表数据增长速度大概就以每年5000行速度增加，这样一来五年之后在教务决策数据仓库中该表数据20万行左右，对于该表一年内数据可以采用单粒度，五年数据可以使用多重粒度设计，每学年一个粒度。学籍变动表与学生基本表相比其行数要少很多，所以在设计中也不用考虑多重粒度设计问题。

在教师主题域中，教师基本表主要保存学校所有任课教师的职称、学历，教

师编号等信息，该表的总行数等于所有任课教师的人数，该表行数不太多，不到3000行，并且每年的增长速度等于当年引进新教师的人数，该表每年数据增长行数不会超过300行，对于该表只需设计单粒度即可，此外，教师工作信息表数据行数与教师基本表行数基本相同，所以教师工作信息表也是采用单粒度策略。教师评价表记录学生对该学期各任课教师的上课情况的打分情况。假设某高校全日制本科生24654人，按每人每个学期平均选修课程门数位4门来算，一年两个学期，这样一年内该表将产生20万行左右的数据，五年之后该表数据行数的增加数将会超过100万行，对于如此大的数据表，我们在设计时应该为其设计多重粒度，其粒度设计如图4-8所示，其中左边的学生评分记录为数据仓库的当前细节集，右边分别为一个学期、一个学年的学生评分汇总，真实资料描述了学生评分的详细信息。

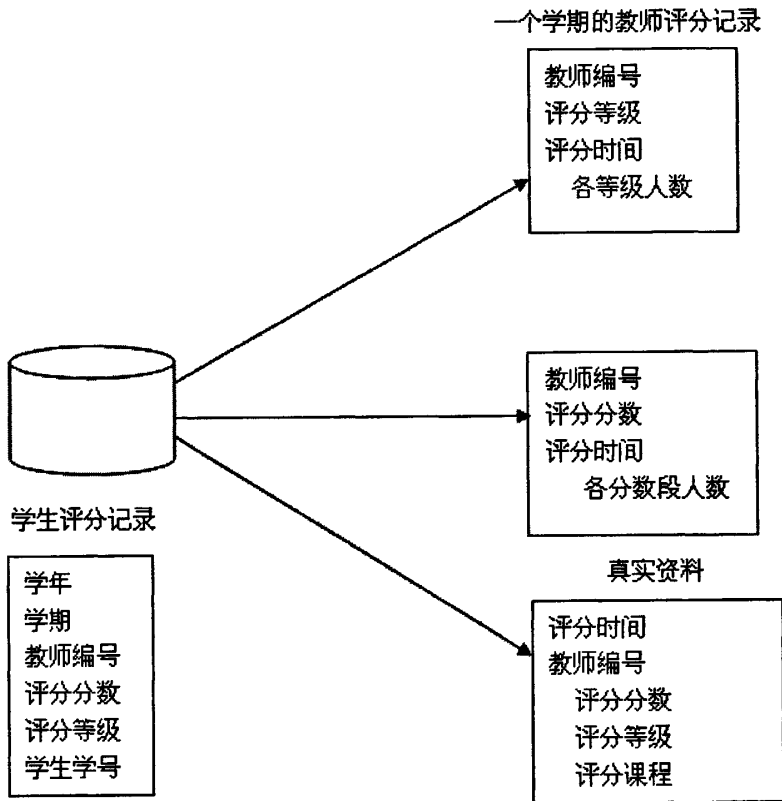


图 4-8 教学质量主题数据多粒度设计图

对于课程主题域的学生选课表、学生成绩表、学生课程表它们的信息量与教师评价表的信息量差不多，因此在设计时也采用多重粒度设计，限于篇幅这里不再一一列出。

4.5 ETL 设计

4.5.1 设计思想

尽管有很多ETL工具可供选择，但是考虑到系统适用性、灵活性、扩展性与成本等因素，以及对教务决策数据源的理解与分析，教务决策数据仓库ETL采用手工编程方法设计，该ETL采用插件技术开发。

该ETL在使用时，在被抽取数据主机上安装数据采集代理，数据分析主机上安装数据采集服务器即可实现数据抽取。设计时将基于Oracle和基于Foxbase的数据源归为数据库数据，其他教务管理信息数据归为文件数据，我们会在ETL工具中分别为它们设计程序。

4.5.2 结构设计

教务决策专用ETL工具采用如图4-9所示模型架构，与以往ETL不同，在ETL工具中除了元数据管理工具外还专门建立一个元数据库。

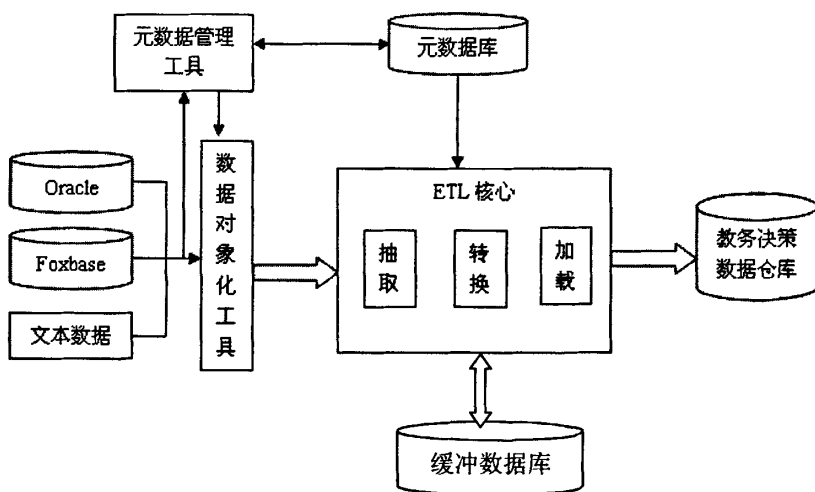


图 4-9 教务决策数据仓库专用 ETL 工具模型图

该ETL工具中数据对象化工具负责将抽取到的数据转换为ETL的统一数据格式。ETL内核负责获取ETL流程规则，把ETL中的规则解析给数据抽取、数据转换、数据加载的各个子过程，然后把规则分配给不同的子过程线程去运行，完成数据的ETL。元数据管理工具用于设置ETL任务的执行方式，协调各个任务流程的

执行，并负责进行运行情况的日志记录。

ETL内核是整个ETL工具的核心部分，其工作过程是先从数据源抽取数据，经过中间处理之后加载到教务决策数据仓库。如果我们把ETL的工作过程看作一个数据流程，就可以把中间处理数据的部分看作处理数据的若干个节点，其中每一个节点完成一项数据处理任务，那么数据抽取由头节点完成，数据加载由尾节点完成，若干个中间节点一起协同工作完成数据转换工作。按照这种假设，ETL中数据转换的工作原理如下图4-10所示。

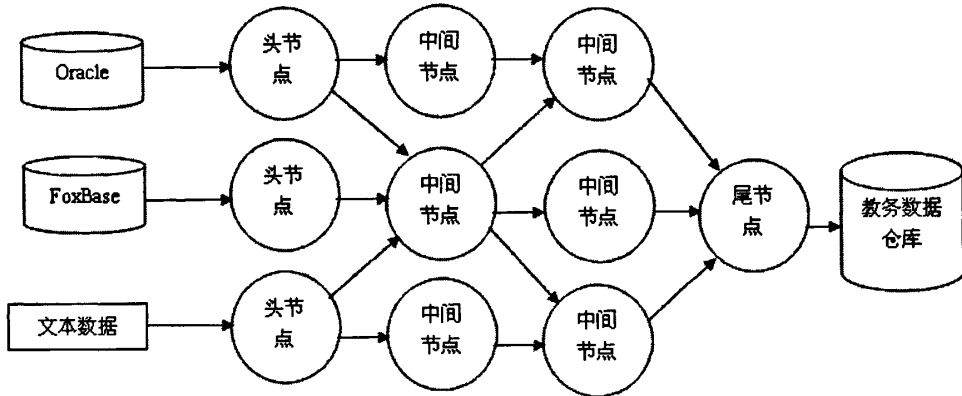


图 4-10 数据转换工作原理图

第5章 教务决策数据仓相关技术实现

5.1 教务数据仓库实现

5.1.1 实施方案选择

在进行数据仓库构建之前，选择一个合适的数据库产品至关重要。自数据仓库技术问世以来，很多厂商分别推出了各自的数据仓库解决方案，如IBM、Oracle、Sybase、CA、NCR、Informix、Microsoft、和SAS等。IBM公司给出的一套基于可视数据仓库的商业智能解决方案，包括：Visual Warehouse (VW)、Essbase/DB2 OLAP Server 5.0、IBM DB2 UDB，以及来自第三方的前端数据展现工具和数据挖掘工具。Oracle数据仓库解决方案主要包括Oracle Express和Oracle Discoverer两个部分。Oracle Express由Oracle Express Server、Oracle Express Objects、Oracle Express Analyzer、Oracle Discoverer四个工具组成。Sybase提供的数据库解决方案包括数据库的建模、数据抽取与转换、数据存储与管理、元数据管理以及可视化数据分析等工具。

教务管理信息系统所用的业务数据库是Oracle数据库，并且Oracle数据仓库解决方案包含了业界领先的数据库平台、开发工具和应用系统，所以在教务决策数据仓库选用Oracle数据库存储数据，并采用Oracle数据仓库解决方案设计数据仓库。

5.1.2 具体实现

数据仓库的实现就是，将所设计的逻辑数据模型转化成物理数据库的过程。在数据仓库的实现过程中，数据仓库的逻辑结构对应物理数据库的表空间，维度，事实实体对应物理数据库的表，实体之间的关系对应表与表之间的外键约束，实体的唯一标识由表的主键来实现，实体属性对应表的列。

在数据仓库的实现过程中要做的工作主要有：表空间、表、分区表、视图、完整性约束、维的创建与实现。教务决策数据仓库，使用两个表空间，一个用作数据表空间 (dataspace)，另一个用作索引表空间 (indexspace)。教务决策数

据仓库所涉及的表较多，并且学生信息表，学生成绩表，教师评价表等这些表都是超大表，所以要为之建立相应的索引，教务决策数据仓库的所有索引均放在索引表空间。此外，教务决策数据仓库中还用到物化视图等。

限于创建过程中所用的SQL语句较多，这里只给出数据表空间和部门维度表的创建语句。数据表空间的创建：CREATE TABLESPACE DATASPACE DTATFILE 'E:\ORACLE\ORADATA\JWJC\JWJC.DBF' SIZE 2048M autoextend on next 32m extent management local；部门维度的创建：CREATE DIMENSION ZFXFZB.BM LEVEL BM IS(BMB.BMBS) LEVEL XY IS (BMB.XY) LEVEL ZY IS(BMB.ZY) LEVEL XZB IS(BMB.XZB) HIERARCHY BM rollup(XZB CHILD OF ZY CHILD OF XY CHILD OF BM)。

5.2 ETL 工具实现

教务决策数据仓库使用Brio作为前端分析工具，这里主要研究ETL工具的实现。

根据图5-1描述的ETL工具架构结构，教务决策数据仓库专用ETL工具具有如图5-1所示不同功能模块。

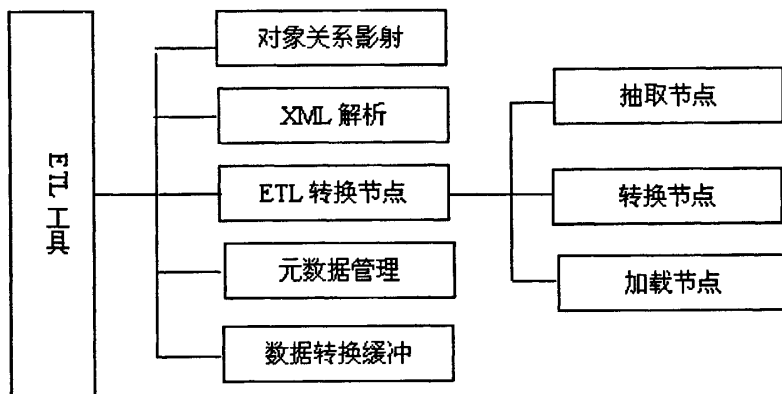


图 5-1 ETL 工具功能模块图

对象关系映射模块实现数据对象化封装；ETL转换节点模块实现具体数据抽取转换加载；数据缓冲模块连接两个转换节点实现缓存转换数据；元数据管理模块负责各节点运行的控制和调度；XML解析模块负责解析数据格式为XML数据。

整个ETL工具实现的工作重心在于ETL转换节点的实现，所以节点是整个ETL工具的核心，它是数据处理单元，组成ETL转换的最小单位。按照功能不同可将节点分为三类：数据抽取节点、数据转换节点、数据加载节点。为提高系统的可扩展

充性，在节点实现中使用了面向对象方法的继承特性。

根据抽取、转化和加载三种节点具有某些相同属性与方法这一事实，在实现中为三种节点抽象出一个基类Node。Node类作为基类给出了涉及到数据转换各子类的方法声明，其具体实现可根据需要通过子类重载来完成。

抽取节点的功能是将基于Oracle数据、基于Foxbase数据，基于文本数据抽取到ETL服务器的内存中。对基于Oracle的数据系统使用DB1_Class类完成数据抽取，基于Foxbase数据系统使用DB2_Class类完成数据的抽取，对于文本数据系统使用Text_Class类完成数据的抽取。

转换节点是类型最多的节点，也是实现功能最多的一类节点。在ETL工具中诸如空值处理、数据字段名字的统一、数据的拆分、数据的合并、不完整字段补充等都是由转换节点完成，每一种功能对应一种类型节点，所以在ETL工具的实现中转换节点的实现难度最大。对于不同的转换功能这里分别为其设计一个单独类作为Node类的子类。目前本系统中为转换节点所用的类有：Null_Class(空值处理)、Unit_Word(字段名字统一)、Table_Cut(表的拆分)、Table_Merge(表的合并)、Field_Cut(字段拆分)、Field_Merge(字段合并)、Field_Complete(字段补充)。

加载节点主要是将经过转换和清洗的数据加载到教务决策数据仓库，可以由一个类Load_Class来完成。

教务决策数据仓库专用ETL工具使用插件技术来解决扩充功能问题，在解决统一管理节点对象的创建问题时，使用工厂模式管理节点创建和添加过程。节点的创建使用CFactory类的方法createC用来生成节点对象完成。实现教务决策数据仓库专用ETL工具数据转换模块各节点功能的类图如图5-2所示。

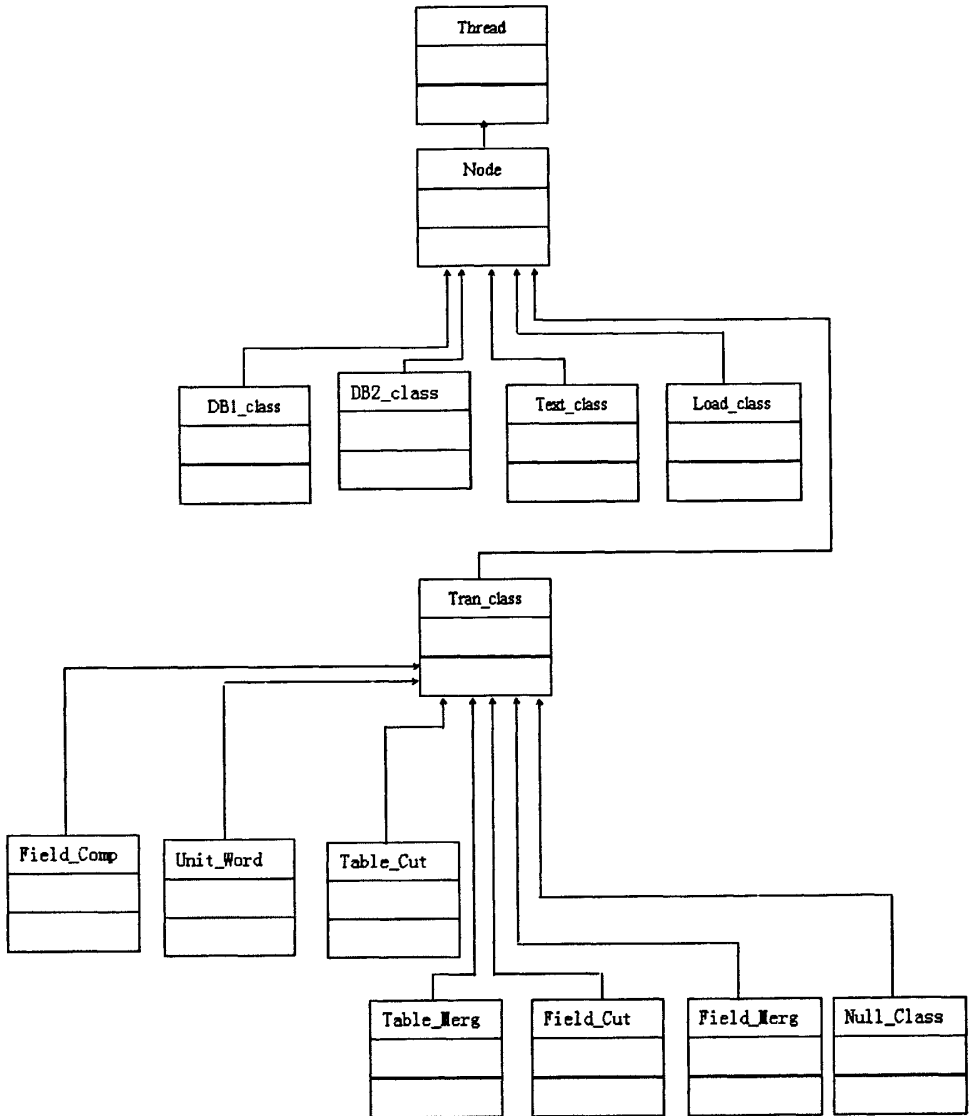


图 5-2 各节点类之间的关系图

第 6 章 教务决策数据仓库典型应用实例分析

作为数字化校园建设的一个组成部分,教务决策数据仓库将异构的教务管理信息集成在一起,起到辅助教务部门决策层领导有效决策的目的。根据前面对教务决策实际需求可知,教务部门决策层领导主要完成围绕学籍、教师、课程、成绩四个主题域的决策与分析。

6.1 异动学籍主题应用实例分析

针对当前教务管理信息系统无法完成对学生学籍进行多维度分析的特点,某高校需要这么一个系统,通过该系统的使用,教务处决策领导能够从部门、地区、民族、性别等多个角度对在校学生的基本信息进行动态分析,判断本校学生结构是否合理、规模是否恰当等。进而达到调整来年招生计划,优化学生结构的目的。

通过从异动原因、异动类别,学院、专业等多个维度对学生学籍异动情况进行动态分析,达到帮助教务部门修订学生培养方案,调整专业结构的目的。

根据上述决策需求,采用 Brio 对教务决策数据仓库中基于学籍主题的数据中 2004 年到 2007 年三年的学籍变动情况进行多维度分析,分析的部分结果如图 6-1, 图 6-2 所示。

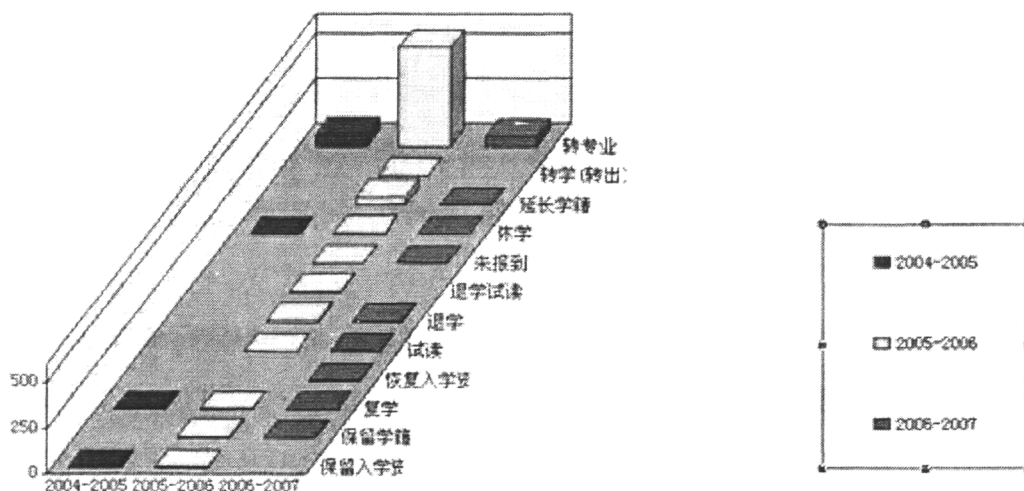


图 6-1 各学年学籍不同变动方式分析图

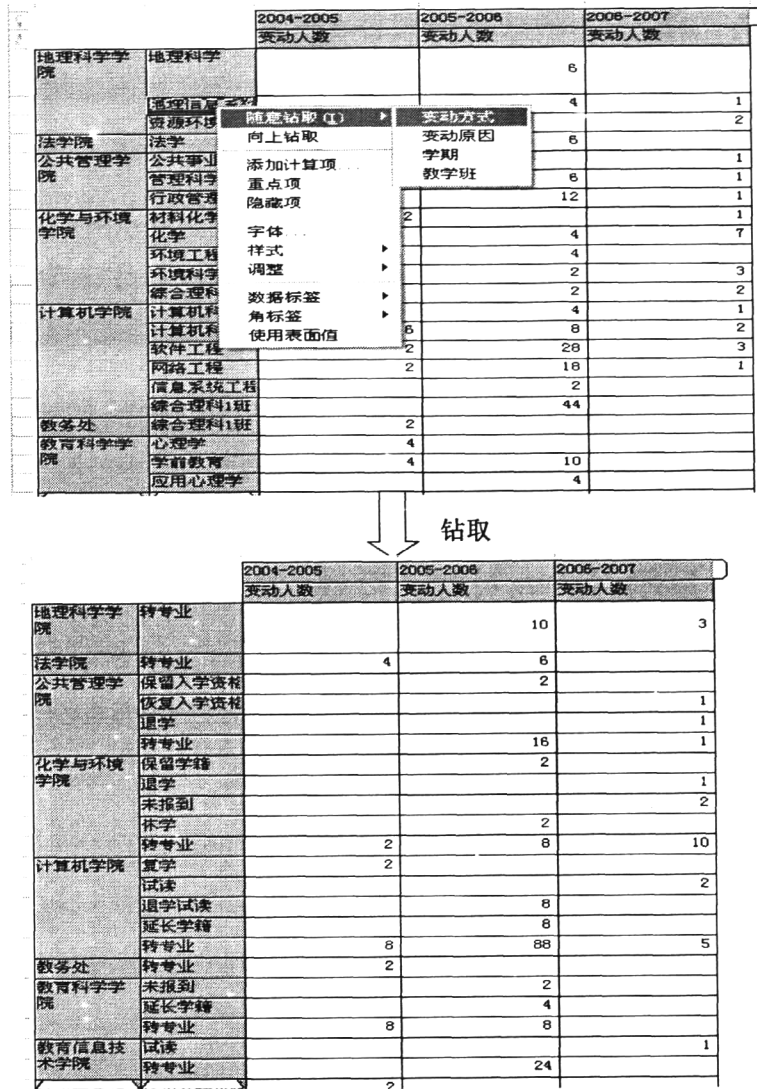


图 6-2 学籍异动钻取分析图

通过分析三年学生学籍变动情况发现，三年来学籍变动中因专业爱好、转入、转出综合班而要求转专业的人数较多；2005-2006 年度学籍变动规模最大，主要体现在变动人数多，变动原因多样化等方面；各学年学籍变动主要集中在计算机学院、经济与管理学院、物理与电信工程学院。

学籍异动主要集中在计算机学院、经济与管理学院、物理与电信工程学院并且 2005-2006 年度达到高峰期，其主要原因在于 2003 年某高校第一次实施综合班制度，当时综合班学生主要集中在上述三个有热门专业的学院，到了 2005-2006 年度到了他们根据自主选择专业的的时间，所以在该年度学籍异动人员数据较大。

针对上述分析结果，教务部门决策领导在接下来的工作中要对综合班学生的

学籍单独处理，此外，还要对异动政策进行认真分析，作出适当调整。

6.2 四六级成绩主题应用实例分析

教务决策数据仓库中成绩主题，主要帮助决策层领导，从多角度对在校生英语四六级成绩、公修课成绩进行分析。为学校制定英语学习指导方针，调整公修课的安排计划提供帮助。

采用 Brio 软件对教对基于四六级成绩主题的数据进行分析，部分分析结果如图 6-3 所示。

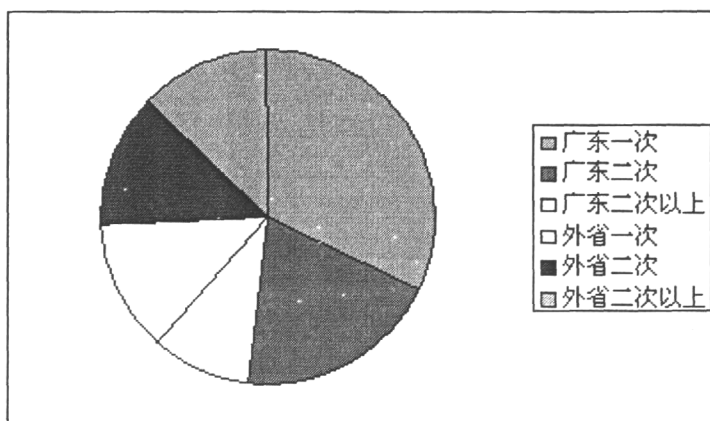


图6-3不同生源地英语四六级通过率分析结果示意图

通过分析发现在四六级通过率方面，虽然最终外地生源和广东生源总的通过率基本上差不多，但是一次通过率广东生源比外省生源要高很多，此外男生的通过率高出女生，计算机学院、物理学院、音乐学院学生通过率不如教科、化学、生物学院，农村来的学生通过率普遍不如来自于城市的学生。

针对这一问题，学校应在新生入校开始，根据学生生源地的不同对其进行引导，适当地提醒外地生源在业余时间多花点时间学习英语，采取相应措施促进计算机、物理、音乐学院学生对英语的学习。

6.3 师资力量主题应用实例分析

在教务决策数据仓库中基于教师主题存放了有关教师基本情况、教学工作量、教学质量相关数据。对教师基本信息分析可以帮助教务处决策领导判断本校教师结构的合理与否，为调整师资结构，制定教师培训计划，确定人才引进类型提供帮助。对教学工作量分析旨在帮助找出影响教师工作量完成的多种因素，为

教务部门制定出合理的教师工作量制度提供帮助。教师质量分析主要帮助教务部门从多维度了解影响教师教学质量的因素,为制定有效提高教学质量政策提供帮助。

根据上述决策需求,对基于师资力量主题的数据从部门、学历、性别、来源地区对本校教师职称进行多维度分析,分析的部分结果如6-4所示。

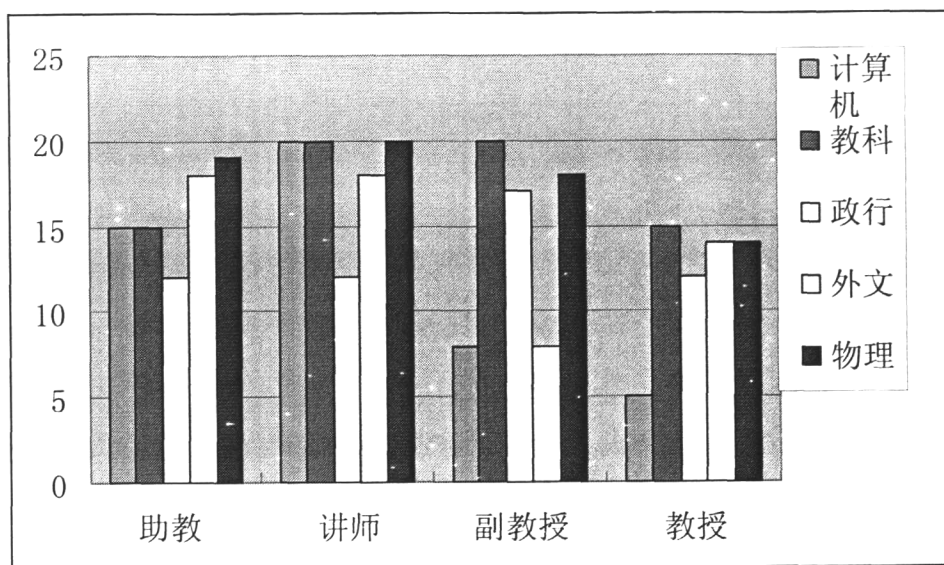


图6-4不同学院不同职称分析图

从分析结果发现教科、生科、化学、物理学院教授人数较多,计算机、外文、音乐三个学院教授数量较少,并且教师的学历普遍偏低,在所有教授中男教授多于女教授、非广东籍教授人数多于广东籍教授人数、职称高的教师学历普遍高于职称低的教师学历。针对此类问题,教务部门决策层领导协助学校其他部门在制定教师引进计划、培训本校教师方面应给出合理建议,要将引进高级职称人才,培养高学历人才的重点放在计算机、外文、音乐学院。

第7章 论文总结

7.1 论文完成工作总结

本论文的重点在于研究如何将数据仓库技术和多维分析技术应用到辅助教务决策中去。通过对日常教务决策调研,以及对历史教务管理系统数据、当前教务管理系统数据以及相关文本教务管理数据分析,提出了通过教务决策数据仓库构建达到辅助教务决策的方案,对教务决策数据仓库的设计进行了研究。本论文主要完成如下工作:

- 1、对教务决策数据仓库模型进行了研究,设计一个具有四层结构的教务决策数据仓库。

- 2、对教务决策的主题域进行分析,使用自顶向下的数据仓库建模工具—信息包图完成教务决策数据仓库概念与逻辑模型的设计。

- 3、采用插件技术设计一个教务决策数据仓库专用ETL工具,将面向对象方法应用于ETL工具的实现,使用Brio软件分析了教务决策数据仓库典型应用实例,并给出了分析结果。

7.2 下一步的研究工作

本文在数据分析方面完成了教务决策数据仓库数据的OLAP分析,并给出了异动学籍主题、四六级成绩主题、师资力量主题的分析结果。教务决策数据仓库的构建是一个长期过程,需要不断地完善,随着数据源数据的不断增加以及教学决策需求的变化,将会有新的分析主题加入到教务决策数据仓库中来,因此,教务决策数据仓库在辅助教务决策的过程中还需不断完善,接下来要做的工作为:

将数据挖掘功能加入到教务决策数据仓库中去,增强教务决策数据仓库对教务数据的分析能力,实现从更深层次对教务数据进行分析,达到有效辅助教务决策的目的。

参考文献

- [1] Nigel Pendse .Drilling Into OLAP Benefits .DM Review .Mar2004, Vol.14: 10-24.
- [2] Conn SS OTLP and OLAP data integration:a review of feasible implementation methods and archtecures for real time data anlysis Digital Object Identifier IO 1109/SECON 2005 1423297 8-10 APril 2005:515-520.
- [3] Inmon W H. Building the data warehouse .[M].3rd ed.NewYork:John Wiley&Sons Inc,2002.31-145.
- [4] Ken OrrData warehouse Technology.http://www.dwinfocenter.org.2005
- [5] 于宁, 王行言, 罗念龙 高校教学决策支持系统数据仓库的研究与实现 计算机工程与设计 .2006 年第 20 期: 3853-3859.
- [6] 龙玉林, 张宪民. 一种可靠的数据仓库中 ETL 策略与架构设计 [J]. 计算机工程与应用, 2005 年 10 期: 172-174.
- [7] 朱德利著 SQL Server 2005 数据挖掘与商业智能完全解决方案[M]. 北京 电子工业出版社, 2007. 71-95
- [8] 郭朝珍, 杨俊杰.基于数据仓库技术和 OLAP 的决策支持系统.计算机系统应用.2000 年第 3 期: 11-13.
- [9] 飞思科技产品研发中心.Oracle9i 数据仓库构建技术[M].北京:电子工业出版社,2003.68-69
- [10] 李雪梅, 何佳洲, 陈世福. 一种基于信息动态打包的数据仓库的设计方法 [J]. 计算机应用研究, 2001 年 1 期: 135-137.
- [11] 朱卫锋, 费奇, 陈学广.维数据仓库建模技术及其在数据建模中的应用. 计算机工程与应用.2002 年 19 期: 201-203.
- [12] 谭振强.基于关系数据库的多维数据存储模型.小型微机计算机系统.2001.22(7): 885-888
- [13] P. Radha Krishna, and Supriya Kumar De.A fuzzy approach to build an intelligent data warehouse.Journal of Intelligent&Fuzzy Systems.2001, Vol. I 1:2332.
- [14] 贺超波.基于学分制体系的高校教学数据仓库与决策支持系统设计关键技术研究[D].硕士,华南师范大学 2007.

- [15] Ken Pohl .Data Warehouse Project Management .DM Review .Mar2006, Vol. 16:2830.
- [16] Pat Minton.The Conformed ETL Architecture .DM Review .Sep2004, Vol.14: 8-12
- [17] Tim Chenoweth, Karen Corral, and Haluk Demirkan.Seven Key Interventions for DATA WAREHOUSE SUCCESS .Communications of the ACM .Jan2006, Vol.49: 11—119
- [19] MIKE HART AND GABI PORTER .THE IMPACT OF COGNITIVE AND OTHER FACTORS ON THE PERCEIVED USEFULNESS OF OLAP .Journal of ComputerInformation Systems .Fall2004, Vol.45: 4756.
- [20] 李建中, 高宏. 一种数据仓库的多维数据模型.软件学报. 2000, 11 7 908-917.
- [21].钟巧华.数据仓库的数据抽取技术.计算机工程: 2004 年第 12 期:62-64
- [22] 龙玉林, 张宪民. 一种可靠的数据仓库中 ETL 策略与架构设计 [J]. 计算机工程与应用, 2005 年 10 期: 172-174.
- [23] WF .Inmon.Building the Data Warehouse .2nd edition .John Wiley and Sons,2000.pp.20-44.
- [24] 陈德军, 盛栩智, 陈绵云, 基于数据仓库的 OLAP 在 DSS 中研究, 计算机工程与应用: 2003 年第 1 期, 30-32.
- [25] 李玉海,张大斌,吕少鹏. 基于数据仓库技术的电信市场决策支持系统探讨.计算机应用研究.2005 年第 6 期: 80-82
- [26] 魏金强.多维分析技术在教学管理系统中的应用与研究[D].硕士,北方工业大学, 2006
- [27] 肖枫, 王国军, 数据仓库技术和 OLAP 的研究, 企业技术开发 2004 年, 第 25 卷, 第 8 期: 6-8
- [28] 李小琳, 何湘东, 李文印, 基于数据仓库及 OLAP 的人口统计预测系统, 吉林大学学报(信息科学版), 2003 年, 第 21 卷, 第 3 期: 313—316
- [29] 曹丹阳.数据挖掘在教学系统中的应用研究[D].硕士,北方工业大学,2006.
- [30] Ralph Kimball Laurs Reeves Margy Ross Warren Thornthwaite 著肖明王永红等编译数据仓库生命周期工具箱: 设计开发和部署数据仓库的专家方法, 电子工业出版社 2004.1 121—143

- [31] 鲍玉斌.数据仓库系统中若干关键技术的研究[D]博士, 东北大学, 2003
- [32] M.A.Hernandez, J.S.Stolfo.Real-world data is dirty:data cleaning and the merge/Purge problem[J] problem [J].Journal of Data Mining and Knowledge Discovery,1998, VOL.2(I): 9-37
- [33] 张宁, 贾自艳, 史忠植.数据仓库 ETL 技术的研究[J].计算机工程与应用, 2002,38(24), 213-216
- [34] 余春红.数据清理方法[J].计算机应用, 2002, 22(12):128-130
- [35] N.Damas, 徐德智, 数据仓库设计方法研究, 企业技术开发, 2004, 23(2), 7-9
- [36] Neely, M. Pamela, A framework and associated software tool for the analysis of source data for a data warehouse, Ph. D, Albany, State University of New York, 2002.47-58.
- [37] V Raman J Hellerstein, Potters wheel.an interactive framework for data cleaning and transformation. Technical Report, University of California at Berkeley, Computer Science Division, 2000.193-211.
- [38] Ralph Kimball,Joe Caserta.The Data Warehouse ETL Toolkit:PracticalTechniques for Extracting,Cleaning,Conforming,and Delivering Data.Manhattan : John Wiley&Sons,2002,10-15
- [39] 陈弦,陈松乔.基于数据仓库的通用 ETL 工具的设计与实现.计算机应用研究, 2004,8: 214-216
- [40] 白洪涛,孙吉贵.一个专用 ETL 程序的实现.计算机应用, 2004,24(2): 102-104
- [41] Michael J.Corey 著数据仓库分析、构建实用指南,机械工业出版社 2000.11 84-95

致 谢

首先我要感谢的是我的指导老师陈启买副教授！在陈老师的悉心指导下，我完成了论文的选题、资料收集以及论文的撰写工作，每一个过程陈老师都抽出大量时间与我讨论并提出了很多宝贵的指导意见。我非常佩服陈老师的渊博学识、严谨治学态度、刻苦钻研精神以及优秀的人格与魅力，并从中受到了很大的启发。三年的研究生生活中，陈老师在学习与生活上给我的关心和帮助让我终生难忘，此外，在论文的撰写与实践过程中，陈老师尽自己最大努力为我提供实验环境，并及时对论文提出宝贵意见。陈老师的教导与帮助，让我终生难忘，陈老师的教诲使我终身受益，在此特向陈老师表示我衷心的感谢！

在论文的完成过程中，刘海老师也给出不少宝贵意见，在此表示深深的感谢。师兄韩冬平时学习上帮助以及在论文完成过程中也给出不少宝贵意见，在此表示感谢。感谢师弟郭莹光、许倡森、黄斯达、陈森平，师妹申利凤、张俊妍等同学在论文的完成过程中提供的帮助。

感谢计算机学院 402 实验室所有师兄师弟师妹们，在这里我度过了最难忘的时光。感谢许爱军、李锋、潘家辉、毛成品、卫建安、吕成伟、梁活民、吴志敏、陈贤明等三年来朝夕相处的同学们，感谢大家营造了一个友好、和谐和融洽的环境，让我度过了美好的研究生生活，感谢计算机学院老师，感谢学校给了我一次学习的机会。

感谢我的父母、妻子，以及其他亲人和朋友，他们长期的支持与关心让我能够最终完成学业。

攻读硕士学位期间公开发表的学术论文

1. 姜永亮, 陈启买, 秦阿利. “基于 Web services 的学籍管理系统的设计与实现”. 现代计算机 2008. 1.
2. 秦阿利, 陈启买, 姜永亮. “基于 XML 异构数据库集成技术的实现”. 华南金融电脑 2008. 2.

附录一

信息包：异动学籍

类别 ↓	→					
	地区	部门	民族	性别	异动	日期
	区域	学院	民族名	性别名	异动原因	学年
	省	专业			异动类别	学期
	市					
指标：变动人数						

信息包：师资力量

类别 ↓	→					
	地区	部门	学历	年龄	职称	性别
	区域	学院	学历名	年龄类别	职称名	性别名
	省	专业				
	市					
指标：人数						

信息包：教学工作量

类别 ↓	维度 →	
	教师	课程
	全部教师	课程性质
		课程类别
指标：课时数		

信息包：教学质量

类别 ↓	维度 →
	教师
	全部教师
指标： 评分分数 评分等级 人数	

信息包：公修课成绩

类别	→			
	教师	学生	时间	地点
	全部教师	全部学生	学年	校区
			学期	课室
			周	教室
			天	
		节次		
指标: 成绩				

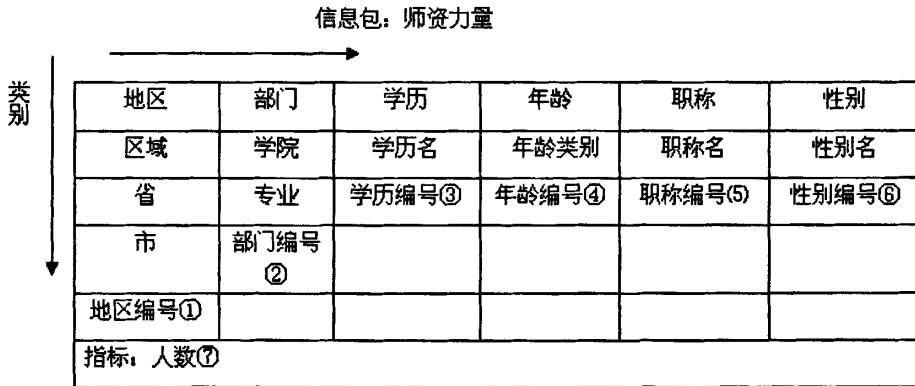
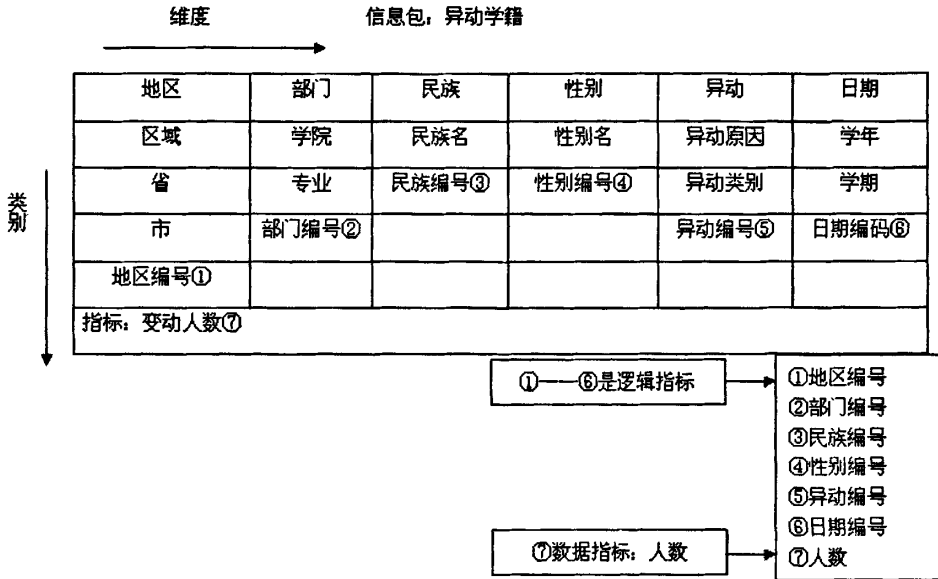
维度 信息包：四六级成绩

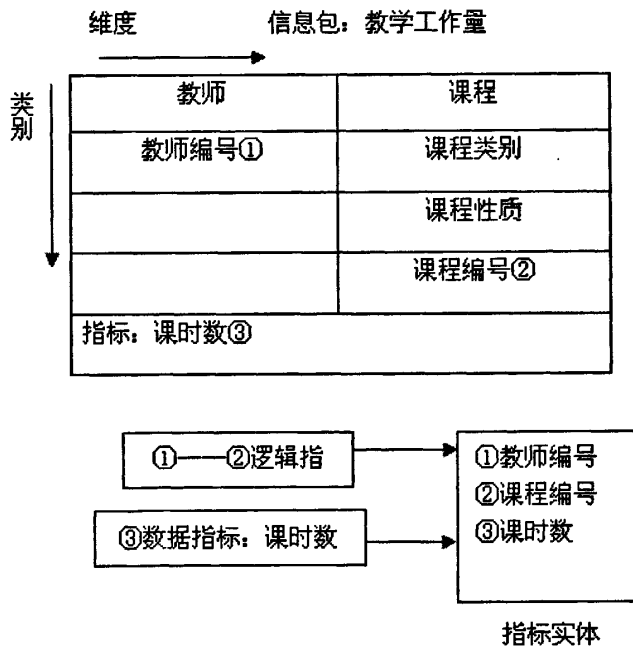
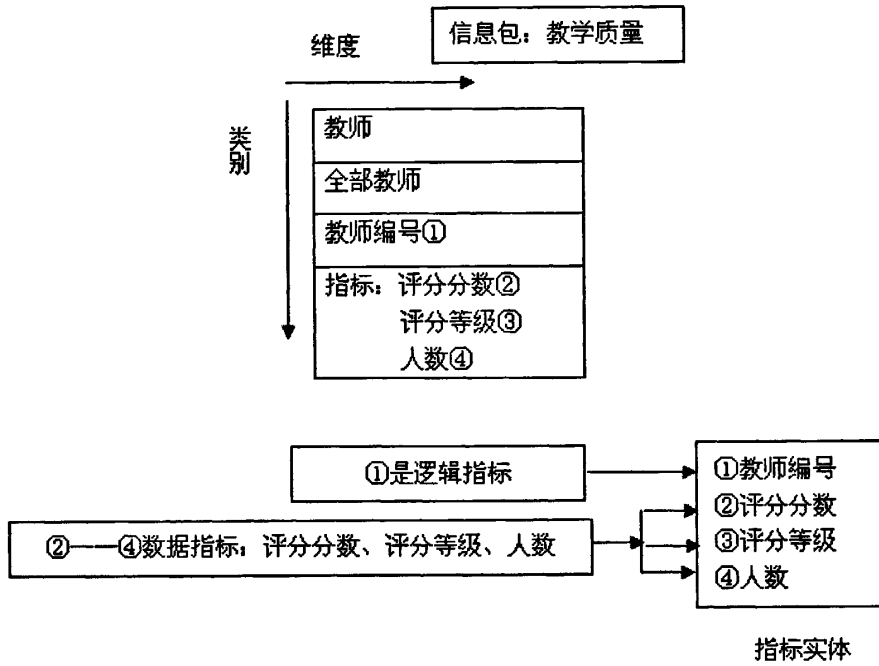
类别	→
	学生
	全部学生
指标: 成绩	

信息包：课程选修

类别	→			
	教师	学生	时间	地点
	全部教师	全部学生	学年	校区
			学期	课室
			周	教室
			天	
		节次		
指标: 人数				

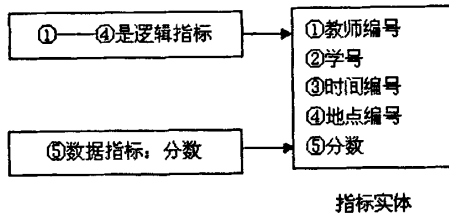
附录二

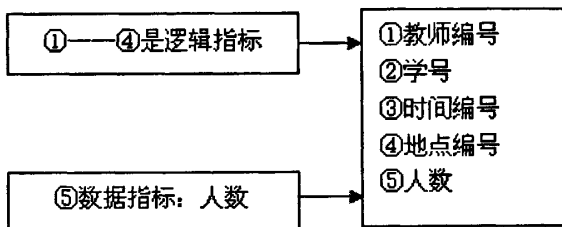




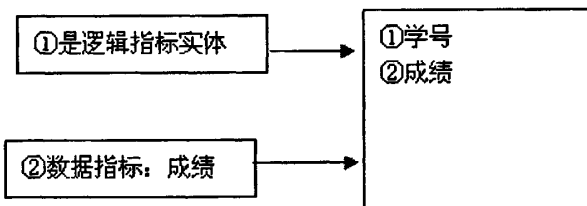
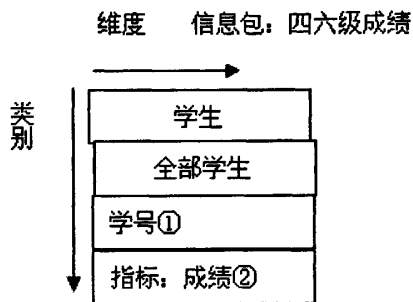
信息包：公修课成绩

类别	教师	学生	时间	地点
	全部教师	全部学生	学年	校区
	教师编号①	学号②	学期	课室
			周	教室
			天	地点编号④
			节次	
			时间编号③	
	指标:分数 ⑤			



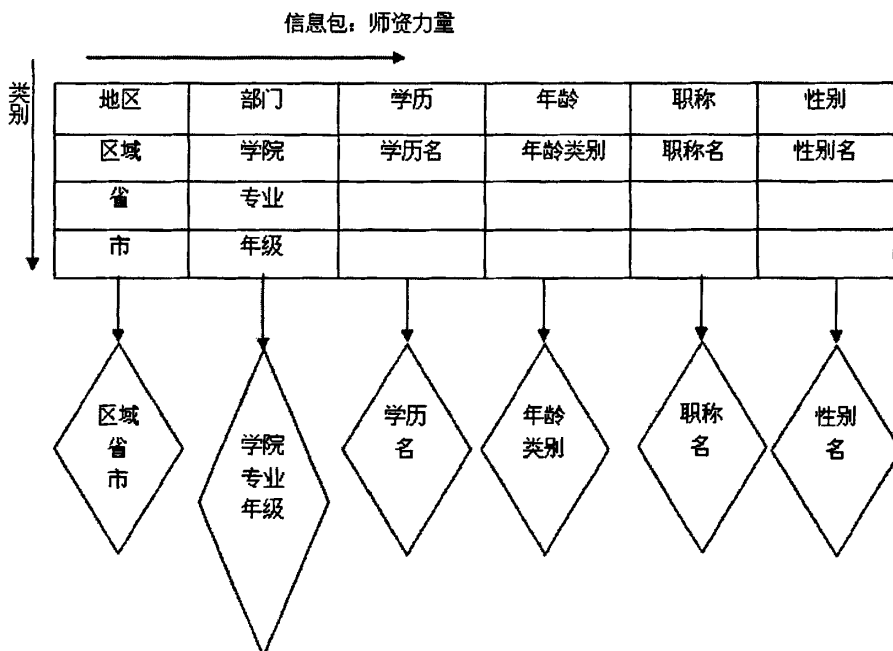
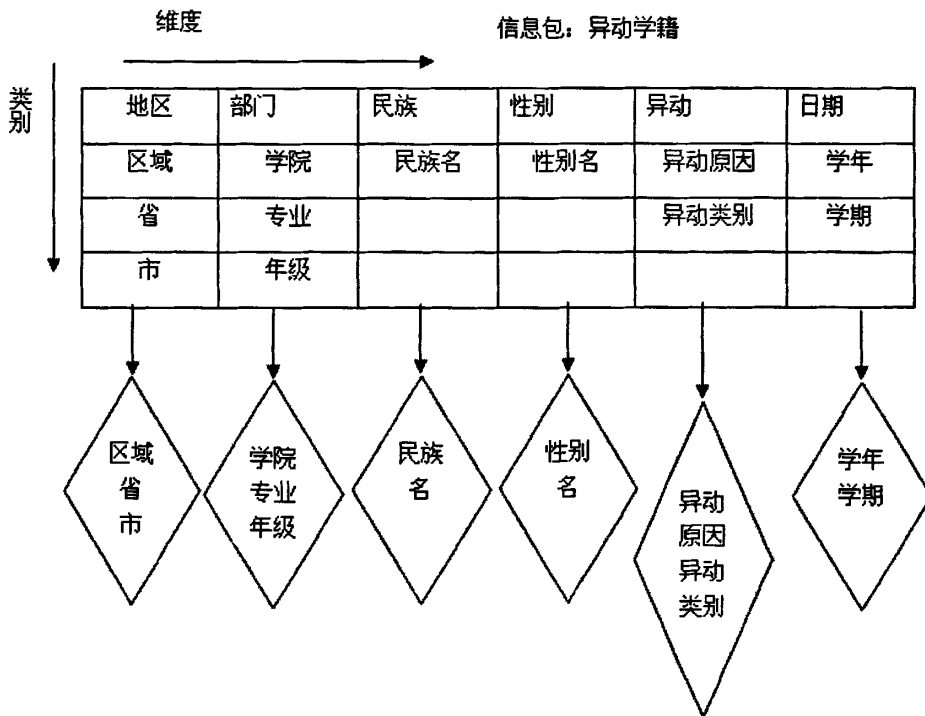


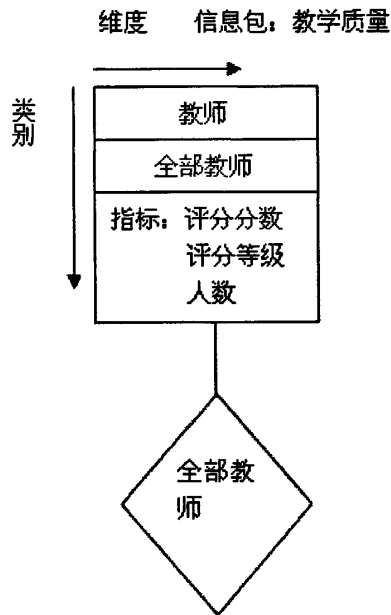
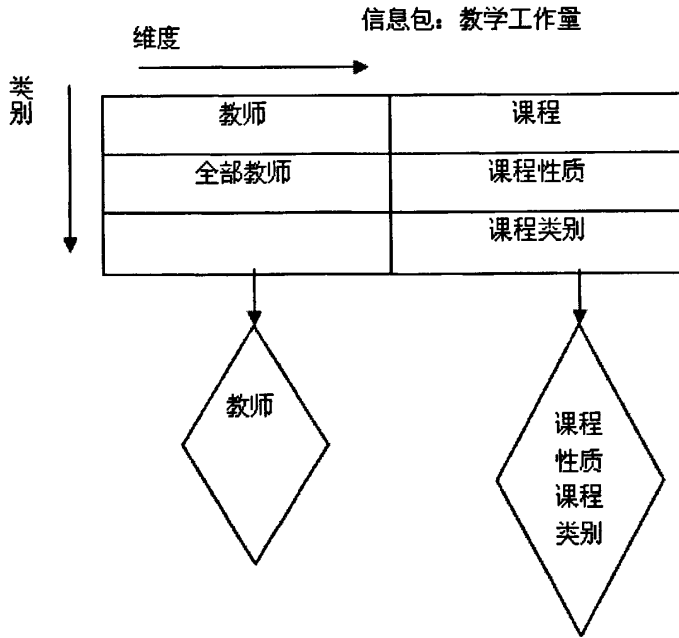
指标实体



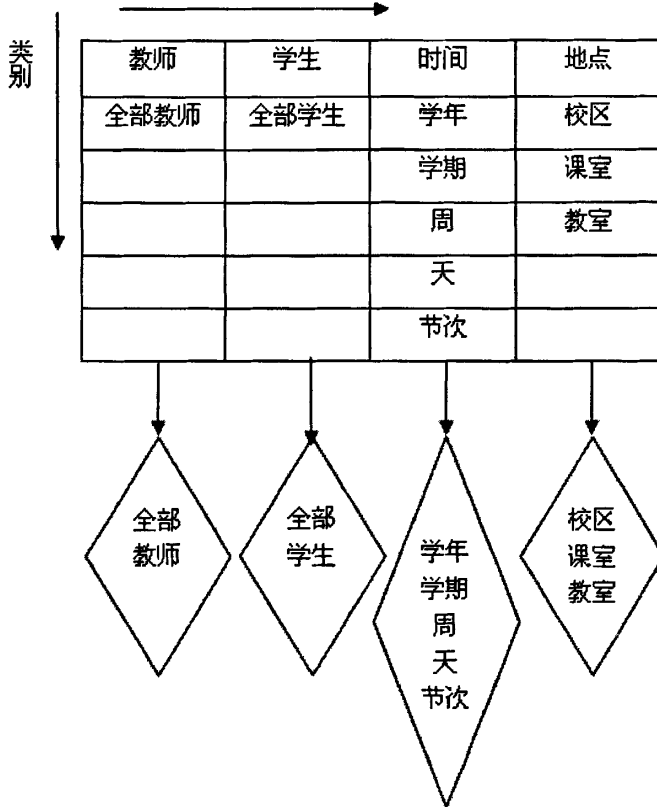
指标实体

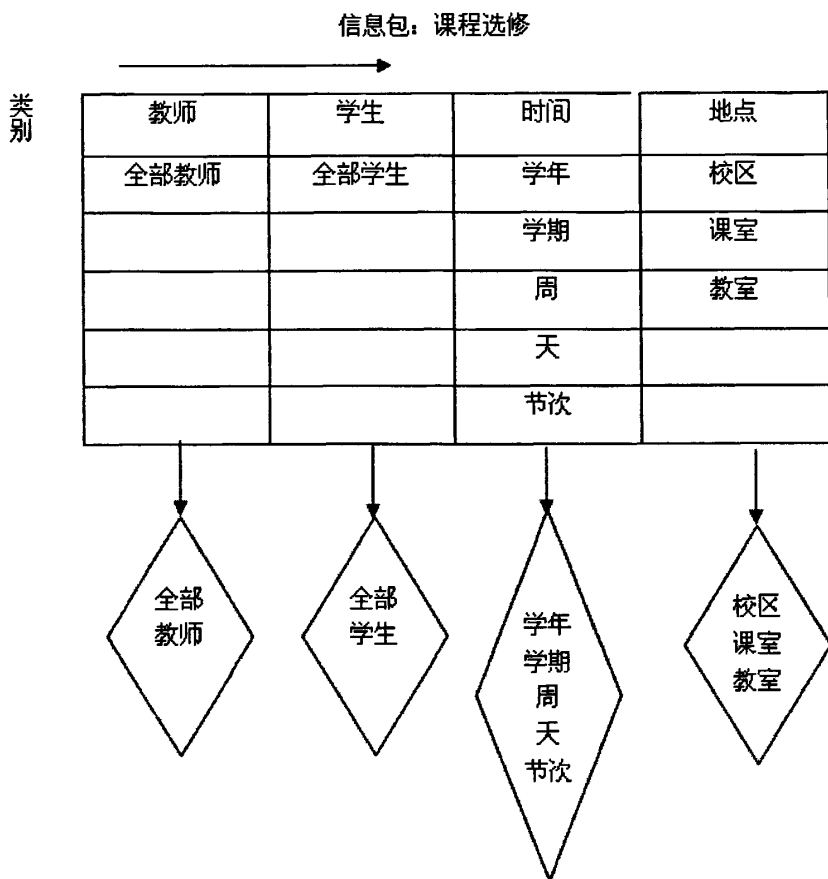
附录三



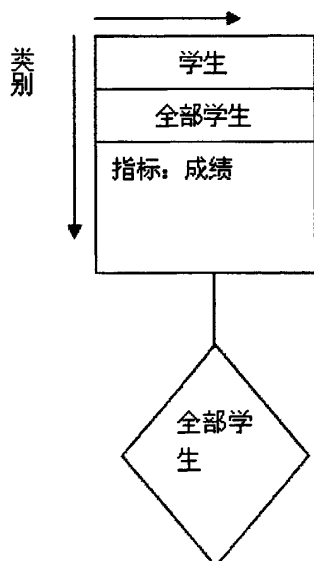


信息包：公修课成绩

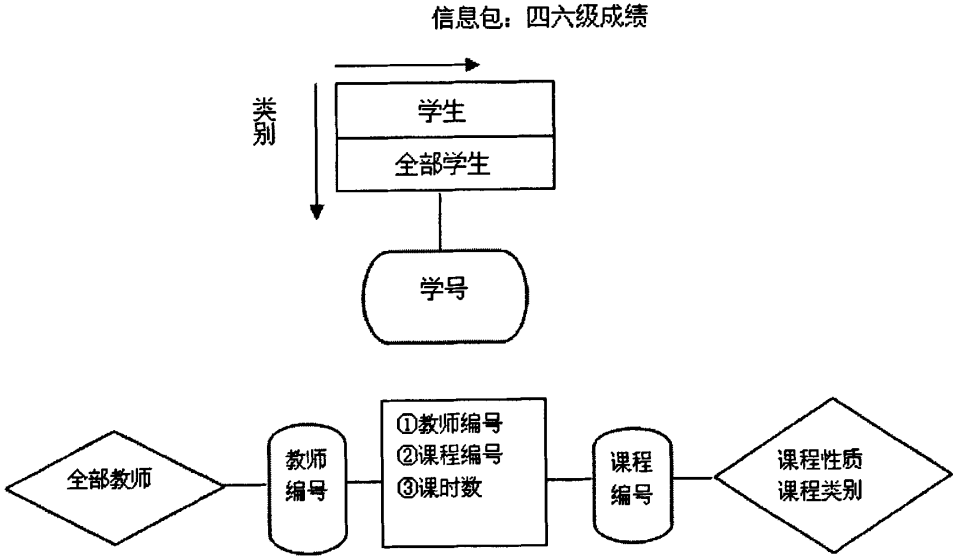
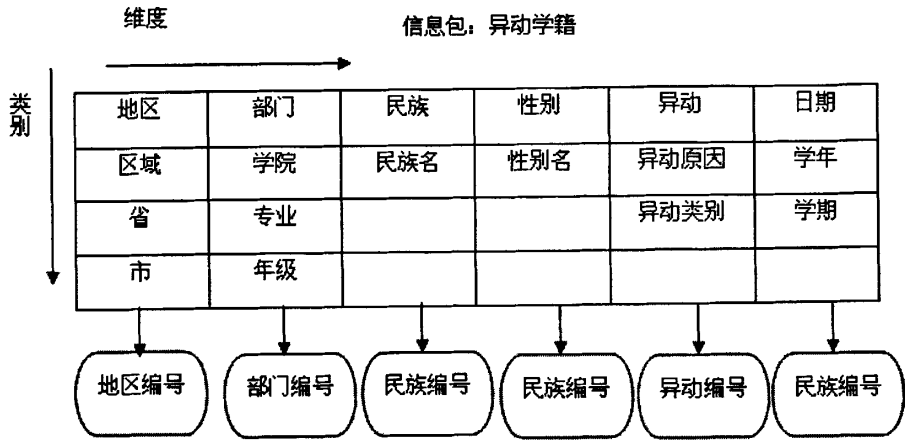




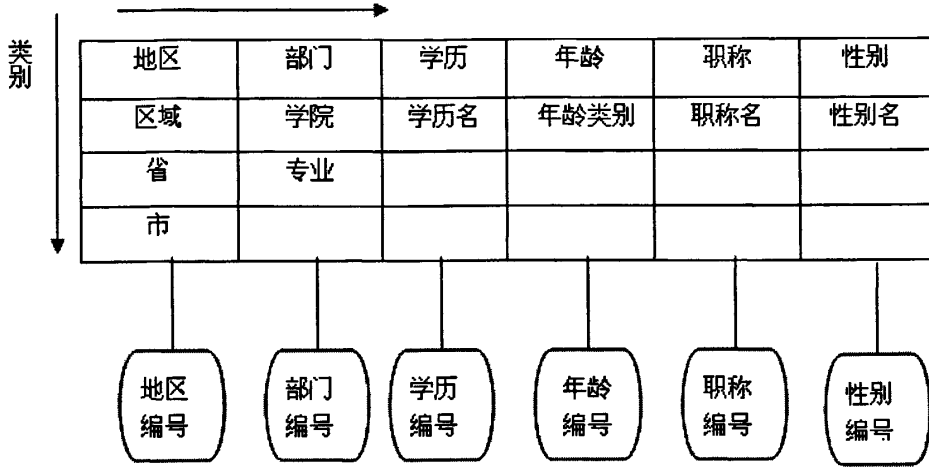
维度 信息包：四六级成绩



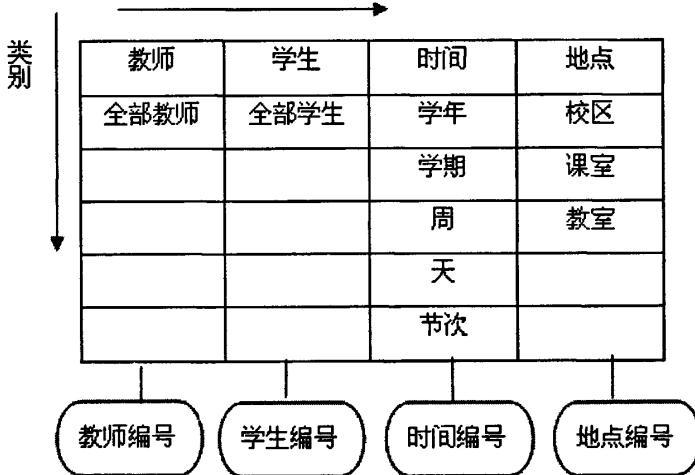
附录四



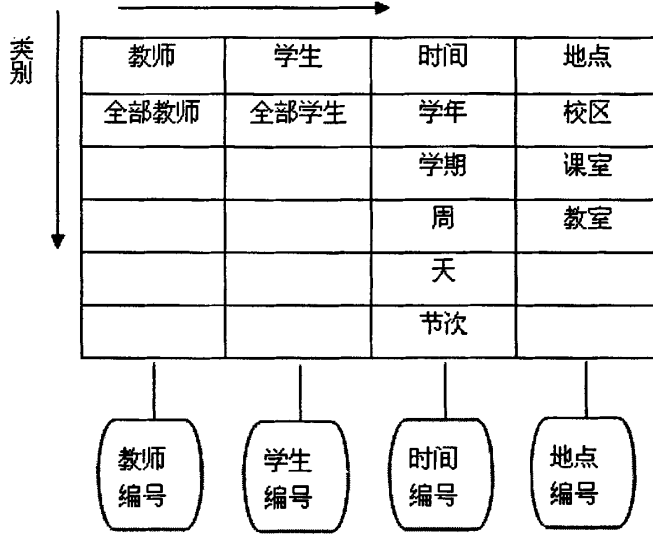
信息包：师资力量



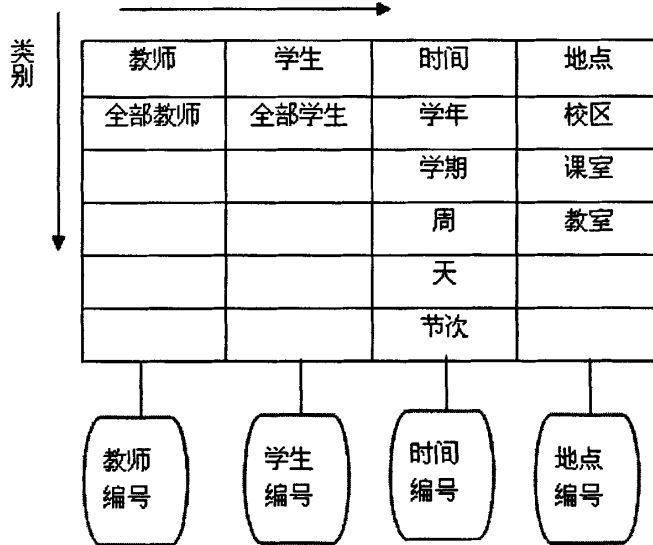
信息包：课程选修



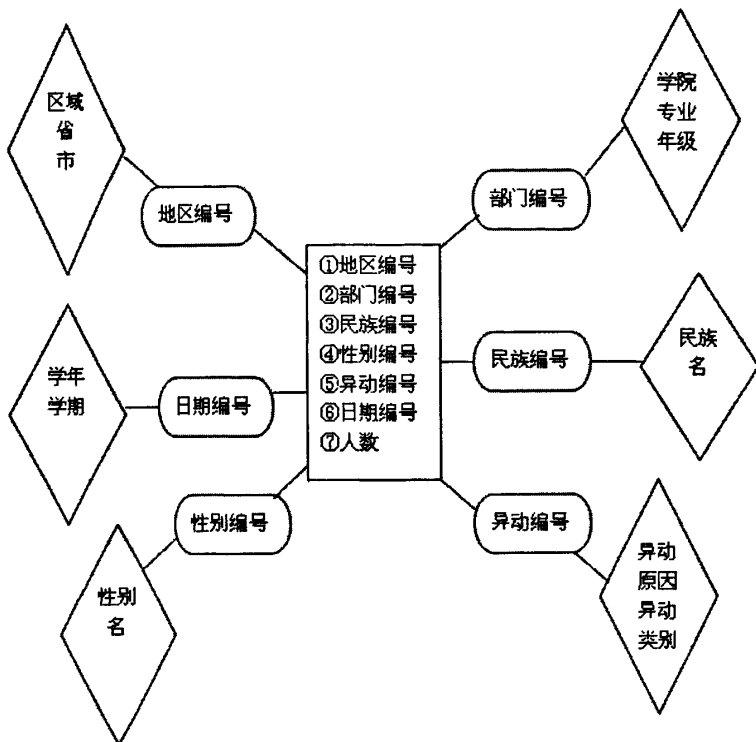
信息包：课程选修



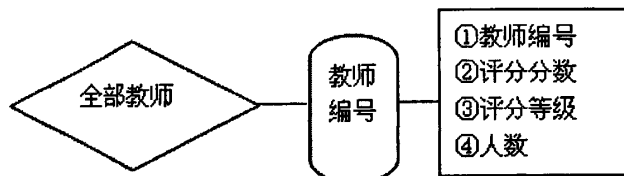
信息包：公修课成绩



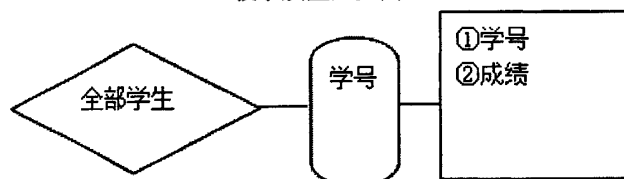
附录五



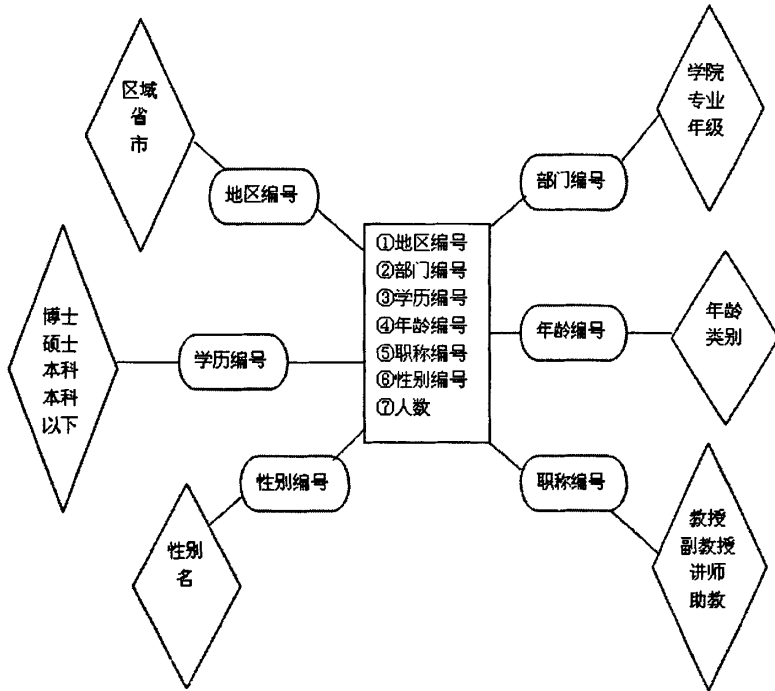
学籍异动星型图



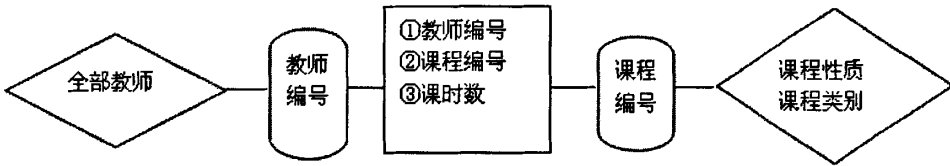
教学质量星型图



四六级成绩星型图



师资力量星型图



教学工作量星型图