

摘要

伴随着信息技术的迅速发展,加快了教育管理的信息化进程,越来越多的教育管理信息系统被用在了教育管理中,教育管理信息系统为教育决策提供了大量的数据支持。如何从海量的数据中快速、准确地提取所需要的信息,成为了管理者面临的难题。把数据转换成数据表或数据图的方式简化了理解复杂数据的难度,也就成了目前教育管理研究的重点。

本文针对中英西南教育信息管理系统中存在的问题,从教育管理指标的角度进行了分析,提出了课题研究的可能性与必要性。通过对教育管理指标分析技术模型的研究,确定了课题研究的过程要素,并在综合分析商业智能技术的基础上,提出了利用微软的商业智能技术实现教育管理指标分析的研究。然后根据设计的技术模型,通过案例对实施过程进行了详细说明,并最终选择适合最终用户的操作平台实现教育数据的分析及可视化。

本文以在校学生人数这一教育管理指标为例,对商业智能技术的实施过程进行了分析。案例以 SQL Server 2008 和 Visual Studio 2008 为开发平台,并结合 Microsoft Office Excel 2007,实现了在校学生人数数据的可视化。本文案例研究的关键环节体现在三个阶段:实现数据填充过程,具体就是数据的提取、转换、加载过程;建立并分析多维数据集,这是实现教育管理指标数据分析的依据;选择适合管理者的操作平台实现数据的可视化。通过商业智能技术实现的教育管理指标分析,大大增强了教育管理者挖掘隐藏在大量数据背后信息的能力,极大提高了教育决策的有效性。

关键词: 教育管理指标、商业智能、决策支持系统

论文类型: 应用研究

ABSTRACT

With the rapid development of information technology, it speeds up the process of education management information. More and more education management information systems are using in the education management, so providing a large number of data for education decision-making. How to acquire the required information quickly and accurately from data-sea, has become a challenge for managers. A method of changing data into data table or data map, which simplifies the difficulty of understanding complex data, and also becomes an important research in education management.

In this paper, for the related problems which existed in the United Kingdom southwest of education information management system, it proposes the feasibility and necessity of the research from the prospection of education management indicators. Through the research of education management indicator analysis technology model, and on the basis of integrated analysis the business intelligence technology, it proposes the method of using Microsoft business intelligence technology to achieve the research of education management indicator analysis. Then, according to the design of the technology model, through a case describe the detailed implementation process. Ultimately, select the appropriate operating platform for end-users to achieve analysis and visualization of educational data.

In this paper, select the total number of students as the education management indicator, to analyze the implementation process of business intelligence. Selecting SQL Server 2008 and Visual Studio 2008 as developing platform, and together with Microsoft Office Excel 2007, the thesis presents the visualization of the total number of students. The key phase of the case study reflects three parts: achieving the process for data filling process, in other words, extracting, transforming and loading data; creating and analyzing the multi-dimensional data sets, this is the basis to realize the data analysis for education management indication; selecting the appropriate operating platform for managers to achieve visualization of educational data. Analysis of the education management indication though business intelligence, which helps to enhance the ability of education

managers to in mining information hidden behind in large amounts of data and tremendously improve the effectiveness of education decision-making.

Keywords: Education Management Indication; Business Intelligence; Decision Supporting System

Type: Application Research

目录

摘要.....I

ABSTRACT..... II

目录..... IV

图目录..... VI

表目录..... VIII

第一章 绪论 1

 1.1 研究背景 1

 1.2 研究的内容 8

 1.3 研究现状 9

 1.4 论文框架 10

第二章 教育管理决策支持系统的开发方法 12

 2.1 教育管理指标分析的技术模型 12

 2.2 商业智能技术综述 19

 2.3 采用的 BI 工具和技术 22

第三章 案例：在校学生人数分析 28

 3.1 开发环境配置 28

 3.2 业务系统数据分析 28

 3.3 创建数据集市 31

 3.4 数据填充 37

 3.5 多维数据集 42

 3.6 数据呈现 44

第四章 案例结果与分析 46

4.1 单一维度学生分布	46
4.2 多维度学生分布	47
4.3 教育管理指标分析结果对教育的意义	49
第五章 讨论	51
5.1 教育管理指标与地方教育主管部门教育规划的制定	51
5.2 关系代数在数据提取中的作用	52
5.3 对数据维度的理解	53
5.4 业务系统中数据质量的监测	55
第六章 总结与展望	57
6.1 论文总结	57
6.2 论文展望	57
致谢	59
参考文献	61
攻读硕士学位期间取得的研究成果	65
学位论文独创性声明	66
学位论文知识产权属声明	66

图目录

图 1-1 论文整体框架结构.....	10
图 2-1 教育管理指标分析的技术模型	12
图 2-2 中英基础教育项目教育管理信息系统登录界面	13
图 2-3 星型结构图	15
图 2-4 雪花型结构图	16
图 2-5 ETL 的体系结构	17
图 2-6 Excel 2007 “数据”功能区的分析服务	19
图 2-7 数据连接向导	19
图 2-8 Microsoft BI 解决方案构架	22
图 2-9 “连接到服务器”对话框	24
图 2-10 Management Studio 主界面	24
图 2-11 新建 Integration Service 项目对话框	25
图 2-12 Integration Service 主界面	26
图 2-13 “解决方案资源管理器”窗口	27
图 3-1 子维表与父维表之间的关系	32
图 3-2 雪花型数据集市	32
图 3-3 EMISDM 数据集市	33
图 3-4 FactStudents 学生事实表	33
图 3-5 DimClass 班级维度表	34
图 3-6 DimSchool 学校维度表	34
图 3-7 DimEthnicGroup 民族维度表	34
图 3-8 DimGender 性别维度表	34
图 3-9 在校学生人数的度量值	35
图 3-10 在校学生人数维度分析	35
图 3-11 定义维度用法	35

图 3-12 EMISSAS 多维数据集结构图	36
图 3-13 成功生效架构图	36
图 3-14 EMISDM 中自动创建的表	37
图 3-15 连接管理器	38
图 3-16 源编辑器	39
图 3-17 列映射图	40
图 3-18 数据源到数据目标的数据流	40
图 3-19 DimClass 维度表中的部分元组	41
图 3-20 DimSchool 维度表中的部分元组	42
图 3-21 DimGender 维度表中的元组	42
图 3-22 DimEthnicGroup 维度表中的部分元组	42
图 3-23 在校学生人数数据源视图	43
图 3-24 在校学生人数多维数据集结构	44
图 3-25 浏览在校学生人数多维数据集界面	44
图 3-26 Excel 2007 呈现数据主界面	45
图 3-27 呈现结果示意图	45
图 4-1 在校学生人数性别维度数据呈现图	46
图 4-2 以学校维度为轴字段的数据呈现图	48
图 4-3 以性别维度为轴字段的数据呈现图	48
图 4-4 两河乡学生分布图	49
图 5-1 数据流	53
图 5-2 维度的层次和类	55

表目录

表 1-1 教育管理信息系统中的各类报表.....	3
表 3-1 学生表.....	30
表 3-2 学校表.....	30
表 3-3 班级表.....	30
表 3-4 民族表.....	31
表 3-5 性别表.....	31

第一章 绪论

1.1 研究背景

随着信息技术、网络技术的日益发展，信息变的无处不在，并且充斥着世界的每一个角落。当今社会，正处于一个知识爆炸，信息海量的时代，人们面对如此大量的信息往往显得不知所措，且很难快速、准确地找出所需的信息。目前，在教育管理中，通过使用各类教育管理软件和信息化平台，提供了支持教育管理直接、客观的数据，提高了教育管理的有效性，也使得其教育信息化水平有了明显提高。

各种教育信息管理系统提供的客观数据在给教育管理规划强有力支持的同时，也积累了大量的数据，通常情况下，我们能够快速认知的只是数据表面呈现出的信息，而隐藏在数据背后的一些重要信息却是很难直接从数据中获得。所以，如何从教育信息管理系统产生的大量数据中提出对教育决策者有用的信息已成为教育管理人员所必须要考虑的问题。

本论文是基于中英西南项目—教育管理信息系统，针对系统中存在的数据可视化进行了相关的研究工作。中英西南基础教育项目为规范和加强学校的日常管理，提高学校和各级教育管理部门利用客观信息进行监测、评估、规划和决策的能力，在项目逻辑架构中设计了教育管理信息系统内容。目的是通过标准系统进行学校基础信息的采集、整理和分析，为教育规划、决策和评价提供客观的依据。但是，在参与教育管理信息系统的测试过程中，发现系统中的数据比较混乱，使用者对系统中给出的大量固定报表数据也不能很好的解读，也就达不到事先预想的目的。针对该问题分析与探讨，转换角度，从人们比较容易接受的图形化报表入手，把固定报表中呈现的大量数据转化成图形报表，以图形报表的形式呈现数据传达的信息，从而形成了本论文的研究课题。本文主要是利用商业智能技术以教育管理指标为标准建立教育管理决策支持系统，实现教育决策者所关心的教育管理指标的数据可视化，以更好地帮助教育管理者做出正确且科学的教育决策，从而提高教育管理的有效性。

1.1.1 教育管理指标

中英西南教育管理信息系统是以学生和教师的个人信息为基础建立起来的,其中最根本的要素就是学生和教师的信息,在系统的建立过程中,始终强调的是建设单一学生和教师的信息卡。学校的主体是学生和教师,要了解一个学校的情况,很显然不可能去了解和掌握每个学生或教师的情况,而是需要把学校中的学生和教师的情况汇总起来,以便更直观、有效地获取学校的信息。因而,也就形成了教育管理指标。

对教育指标的研究源于 20 世纪 20 年代中后期的美国,最先是包含在经济指标和社会指标的研究中,其中教育的社会指标可以说是教育指标的雏形。随着经济与社会的发展,许多领域都开始了研究适合自主发展的指标,教育就是其中一个重要的内容。对教育系统作用及效益的评价开创了一个收集数据、分析教育系统的新时代,同时,世界各国对教育质量也越来越重视,如何提升教育的质量成了教育的主题。自此,对教育指标的研究工作也就开始逐渐发展起来。教育指标是经由收集整理各种教育相关信息,为教育系统提供可解释的资料,以作为教育政策说明与决策参与之用¹。

在教育管理的过程中,必须要依靠正确的教育数据才有可能制定科学、有效的教育决策。教育管理者在实际的实施过程中,也要学会收集、分析和使用数据,教育管理信息系统就是一个以收集和分析数据的系统,通过对数据全面而科学地分析和研究,有助于为教育管理提出有价值的建议。那么,对于中英西南教育管理信息系统有哪些值得我们关注的教育管理指标呢?就系统的使用者层面而言,教育信息系统自下而上分为三种:学校、乡镇和县,对于不同的使用者,教育信息系统提供的报表是不一样的。但是,无论是哪一个层面的使用者,都需要了解这些报表中有哪些教育管理指标的内容和含义,及其在教育实践中有什么样的价值。

在已建成的教育管理信息系统的业务系统中,教育统计指标被分成了三大类:“学校统计指标”、“教师统计指标”和“学生统计指标”,具体每类教育统计指标所

¹ 简茂发,李琪明.当代教育指标:国际比较观点[M].台北:学富文化事业有限公司,2001.2.

包含的分类见表 1-1¹。对于系统不同层级的使用者，不仅要了解这些教育统计指标所包含的信息，而且要能实际应用于教育管理的实践中，依据这些报表中的统计指标，增强教育管理的科学性和合理性。按照教育管理指标收集、分析、整理的数
据，是以教育管理者关心的问题为出发点，所得结果能够很好的反应各教育元素间的关系，方便教育决策者方便、全面、直观地掌握隐藏在教育数据背后的信息，从而更好的为教育决策提供依据。

表 1-1 教育管理信息系统中的各类报表

报表类别	报表名称
学校统计报表	学校基本情况报表
	各类学校统计报表
教师统计报表	教师、学生、班级统计报表
	教师队伍基本情况统计报表
	教师奖励科研情况统计报表
	教师工作情况统计报表
	专任教师学历情况统计报表
	专任教师职称统计报表
	专任教师非学历教育培训统计报表
	专任教师异动情况统计报表
	专任教师任课及成绩统计报表
学生统计报表	家长通知书
	分学年学生变动统计报表
	学生毕业去向统计报表
	单科学习成绩统计报表
	学生成绩统计报表（比较）
	学生成绩统计报表（分数段）
	退学学生情况统计报表
	学生巩固情况统计报表
	寄宿生情况统计报表

¹ 中英西南项目—《教育信息管理系统》

	学生出勤情况统计表
	奖助学金统计报表
	学生基本信息统计表报
	在校生分年龄年级情况统计表
	在校生数统计表

1.1.2 重要教育管理指标举例及其意义

1. 升学率

升学率是检查学校教育教学质量的重要指标。按理说，九年义务教育阶段的小学毕业生都要升入初中，升学率应该是 100%。如果没有实现这一指标，则需要寻找学生未升学的原因。同时，分析不同类别学校的升学率，有助于评价这些学校的办学质量，有助于分析不同学校之间存在的差异，可以为相对落后的学校提供更多的支持和帮助；同一学校历年升学率的比较，有助于分析学校工作的努力程度，有助于反思学校的改革与发展；按民族、性别等的分类比较，可以有针对性地分析是否存在因民族或性别产生的差异性，有助于对少数民族和女生的教育给予更多的关注。

2. 在校学生人数

可以从不同的维度对在校学生人数状况进行分析，例如：民族、性别、是否寄宿等。按照不同民族、性别对在校学生人数分布进行分析，可以帮助管理人员了解少数民族与女学生的构成比例，分析与本地人口中的对应比例是否基本一致或接近。从是否寄宿的角度，可以分析寄宿制对学生学习情况的影响，可以总结取得进步的经验和方法，或者考虑如何加强管理的措施与方法。特别是要关注女生、少数民族等的上学问题。

3. 专任教师的工作量

这里我们以专任教师平均周课时为工作量的计算方法，周课时数是衡量教师工作量的指标，要合理安排每位教师的工作量，以保证满工作量的同时也不能使部分教师超负荷工作。将教师的工作量和教师的教学效率结合在一起，可以分析工作量对教学效率的影响。要为每个专任教师科学安排工作量，这是教育管理人员的能力

表现,在为教师安排教学工作量的同时,还要注意给教师留有一定的自由时间,以便使他们能够开展自主学习和教研工作等。

4. 专任教师的年龄结构

各种不同类型的学校通过分析专任教师的年龄结构,可以清楚的看到各个层级教师的比例,从这个比例中可以分析出老中青教师之间的结构是否合理,在教育管理的过程中,要特别关注教师队伍结构的合理性,防止青黄不接的现象出现。通过分析不同年龄构成的教师队伍对教学质量的影响,特别对一些主要学科要给予针对性的关注。站在乡镇或者县级的角度,需要考虑不同学校之间或不同乡镇之间教师在年龄结构上的相对一致性。

5. 教师、学生、班级比例

通过分析学校教师与学生的数量比,可以从所得比例数中清晰的了解是不是存在不合理的超编教师,或者是教师数远不能满足正常教学的需求等情况,以便能够及时增减教师,以满足正常教学的需要,改变教学的质量。通过分析班级和学生的比例,可以知道每个班的班额数是否符合规定,是否存在大班额的情况,而分析某一科教师数和班级数,可以清楚的看到教师的教学工作量是否符合规定的要求等。

6. 教师队伍基本情况统计

通过分析教师队伍中的教师性别、民族的比例,可以清楚的了解教师在性别和民族层次上的合理性;分析教职工总数中正式教师的数量与全部教职工数量的比,可以从中反映出教师队伍的稳定性;通过分析专任教师的学历结构,可以知道其学历的合格率是否达标,以便关注教师学历进修情况,而对于乡镇和县级部门可以通过关注不同学校、不同乡镇教师学历的分布情况,合理配置不同学历的教师资源;同样,可以分析专任教师的职称情况等。根据各数据分析结果,可以全面考虑如何在学校建立高质量的教师队伍,以保证稳定的教师队伍、足够的教师数量、合理的教师结构等,以保证教师队伍的可持续发展。

对于各种不同类型的学校、不同的乡镇和县,他们所关心的实际教育管理方面的问题存在着不同的侧重点,但大体上都是表格 1-1 报表中所涉及到的问题,这里给出的只是一些在教育中比较容易关心的、比较重要的教育管理指标,其他的教育管理指标还有很多,这里就不再一一进行描述了。

1.1.3 为什么要对教育指标分析

教育指标是收集教育数据的基础,并且能监测教育发展和支持教育决策,越来越成为教育研究的重要领域,教育管理者或决策者可以越来越多的借助于教育指标来分析教育问题。教育指标本身具有很强的目的性,它关注的是教育系统的关键特征,特别要求能够进行某种目的的比较,以及与教育政策相关等等¹。目前国际上达成一致共识的是对教育指标的狭义理解,教育指标是为了更好的实现教育指标信息而在教育统计基础之上进行的一种加工,使其更具可比性。狭义的教育指标成为教育研究的热点,并不是偶然的,它根源于教育系统本身的复杂性。教育系统的日益复杂化也对教育决策造成了一定的后果:第一,提高了教育决策的难度。随着各种信息化技术的使用,教育系统存储的数据日益增多,但是决策者不可能完全分析所有数据以使决策达到最优化。那么,决策者如何才能解决无限增长的数据与自身处理数据能力的有限之间的矛盾?第二,增强了教育决策的风险性。在当今社会,各个领域之间已形成了一个不可分割的整体,不同领域间的密切程度也大大提升了,教育决策的影响也越来越广,不仅可以为教育决策者提供决策制定的依据,而且还可能短时期之内迅速扩大到其他领域,甚至还可能给全体国民造成影响。因此,现在复杂的教育系统给教育决策带来了更大的困难和风险性,但它并不允许在制定教育决策的时候出现失误。

如何在教育管理者自身有限处理能力的前提下避免教育决策的失误?长期以来,社会科学家和行为科学家们一直致力于寻找一种既能对社会系统进行结构化认识,简化和梳理繁华的社会信息,又能对系统的演化和运行过程实时监测的工具²。20世纪60年代,“社会指标运动”的出现实现了这一目标,而教育指标最初正是社会指标的一部分。随后,教育指标在国际组织的重视下越来越受到了关注,且得到了大力发展,教育指标的构建为教育政策地制定提供了主要依据。教育指标能够实现教育现象的数量化,是反映教育发展进程、制定教育发展规划的数据指标。总的来说,教育指标有对教育现象进行客观描述的功能,能如实反应教育发展的现状;

¹ 刘建银,安宝生.教育指标理论研究的几个基本问题[J].中国教育学报,2007.9.

² 同上

教育指标有对教育现象及其发展变化情况进行评价的功能,通过分析比较对教育发展的水平与进程做出判断和评价;教育指标通过数据指标有对整个教育系统运行情况监测的功能,有助于管理者及时发现问题、解决问题,从而更好的对教育系统进行调控,修改和制定更符合实际情况的决策;教育指标可以通过对不同阶段数据的分析,寻找教育发展变化的规律,以便对未来可能发展的趋势或可能出现的问题进行预测¹。教育指标可以通过数据的收集、分析和整理将教育现象数量化,以定性和定量结合的方式呈现教育结果,方便教育管理者了解教育系统的特征,评估教育发展的进程,以此对教育发展做出更有价值、科学的决策。

1.1.4 目前面临的困难

随着信息技术的快速发展,对大量数据的收集变得越来越容易。但是,如何合理的存储如此规模庞大的数据,揭示数据之间的关系及变化趋势,就成为了急需解决的问题。实际上我们正处在一个即拥有大量数据,又缺乏有意义、有实际价值的信息和知识的境地。目前我们面临的困难不是如何拥有数据,而是如何才能“看到”这些数据,也就是如何快速、有效的从这些大量的数据中提取有用的信息或知识,简单的说,就是如何实现数据转化成信息或知识的过程。

中英西南教育管理信息系统是以学校为基础设计的,目的是规范和加强学校的日常管理,在学校中采用规范的管理采集各类教育数据,为学校 and 各级教育部门提供真实、可靠的数据,使各级教育者都能通过这个系统获得改进教育质量的信息,增强教育管理的针对性。各级教育行政部门也可以利用基于学校的基础数据库对各学校进行客观的监测和评估,从而更好的制定符合实际情况的教育规划和决策,提高教育管理者的管理水平和效果。

中英项目教育管理信息系统面向义务教育阶段,收集和储存以学校为基础的各项教育指标,同时,该系统的固定报表服务能够提供国家政策和规划所需要的信息,提高各级教育统计报告的透明度和准确性。但是,固定报表所呈现的数字式的汇总数据只有中层的管理人员能够解读,对于那些高层的决策者并不能完全解读,有时甚至是没有时间去查看那么多的报表。特别是当以学校为基础的数据层层上报后,

¹ 徐娅. 首都教育指标体系研究[J]. 教育科学研究, 2001, 11.

越是高层的教育决策者对于报表所呈现的数据信息越是难解读。

高层的教育决策者面对如此大量、复杂的数据，也同样面临着将教育数据转化成信息或知识的问题，如何才能透过数据的表面现象看到本质？答案就是要提供一种直觉的、交互的，人类思维方式比较容易接受和理解的图形形式，也就是我们所说的可视化环境。它将数据转化为一种视觉形式，以图形的形式传达隐藏在数据现象背后的本质。为了帮助高层教育决策者更方便、更容易地分析大量的客观数据就需要决策支持系统的技术支持，对教育决策者关心的教育管理指标运用商业智能技术进行分析并最终以图形的形式呈现，实现教育数据向信息或知识的转变。

1.2 研究的内容

随着信息技术和各种软件平台在教育领域的使用，积累了越来越多的教育数据，它们为教育管理人员做教育决策提供了客观的依据，但同时，随着教育信息化的发展，越来越多的数据只是被存储在数据库中，仅仅是成了“数据孤岛”，而很少人去关注隐藏在这些数据背后的知识。随着数据挖掘技术的迅速发展，各种各样的数据挖掘软件被应用在了教育管理系统中，试图利用这一技术挖掘出满足教育决策者目标需求的知识，但结果很多时候却是不能够满足目标的需求。所以如何更精确的挖掘隐藏在数据背后的知识，被越来越多的决策者所关注。

目前，借助于商业智能技术开发的教育决策支持系统能够帮助教育管理者根据分析所得的结果做出更科学、可行的决策，它借助已有的现代信息技术，充分利用业务系统的数据，在提取、整理和分析数据的基础上，呈现有用的信息，以帮助教育管理者在加强教育管理、促进教育发展等方面做出及时、正确的决策，并预测未来的发展趋势。商业智能是一个在大量数据基础上，对其进行提炼和重新组合的过程，它可以完成数据到信息再到知识的转变，提高了“数据管理”的能力，并最终为用户提供强有力的决策支持。

本文研究的主要内容，是在对教育管理指标分析的基础上，以教育管理信息系统中现有的教育数据为依据，对决策支持系统进行架构设计，对教育管理者所关心的教育指标，使用商业智能技术进行分析、设计，并最终以图形可视化的形式呈现

教育管理指标所包含的信息。最后，再结合实际情况，对结果进行一些必要的讨论和解释。

1.3 研究现状

本论文研究的主要内容是利用商业智能技术实现教育管理指标分析的研究。伴随着信息技术发展应用而生的商业智能，目前已成为国内外广泛关注的研究方向。高速发展的信息技术为企业提高自身的竞争力带来了巨大的空间，不仅使企业获取了需要的信息，而且企业利用信息技术可以对信息进行再加工利用，从而提高了企业的竞争优势。同时，由于各种软件平台的应用，也使得商业智能更容易和企业结合在一起，促进了商业智能在企业中的迅速发展。在现代社会各企业之间的竞争越来越激烈，为了能够使企业在众多的竞争者中脱颖而出，就迫使企业管理者必须借助一些现代化的科学技术手段，了解企业的运营情况。随着商业智能技术的日趋成熟，越来越多的企业意识到商业智能可以帮助企业管理者做出决策。近年来，国内外商业智能研究者对商业智能都开展了广泛的基础研究和应用研究，所以商业智能在近年来应用较为火热。

从商业智能技术的应用研究，可以看出，目前我国商业智能研究尚不够成熟，涉及的研究方向也多集中在数据仓库、数据挖掘、多维数据库等技术方面，而很难使商业智能在决策方面发挥出应有的作用，需要加大研究的力度。通过对商业智能的应用分析，发现商业智能主要应用在制造业、零售业、金融业和服务业，而在教育中的应用较少，通过仔细分析这类文章，发现目前利用商业智能技术实现教育管理指标分析的研究还没有。随着教育信息化的发展，各种电子化、网络化的教育资源快速的发展起来，给教育管理规划提供了大量客观数据的支持，但是隐藏在这些教育信息背后的知识却是很少被人们所发掘。所以，本论文尝试从教育管理指标的角度，利用商业智能技术对教育信息管理进行研究，通过架构教育管理决策支持系统，使教育管理者就关心的教育管理指标，快速、方便地获得隐藏在大量教育数据背后更有价值的信息，从而更好的进行教育决策的制定或修改，促进教育的发展。

1.4 论文框架

本论文基于中英西南教育管理信息系统中存在的问题，在对教育管理的重要指标进行分析、制作的基础上，主要是利用商业智能技术实现教育管理指标相关数据的可视化。教育管理指标分析能够帮助教育管理者方便地掌握教育系统中各要素的运行情况，通过商业智能分析教育管理指标能够直观、快速地发掘隐藏在教育数据背后深刻、生动、形象的可视化信息。

在论文的结构安排上，共分为六章，如图 1-1 所示。

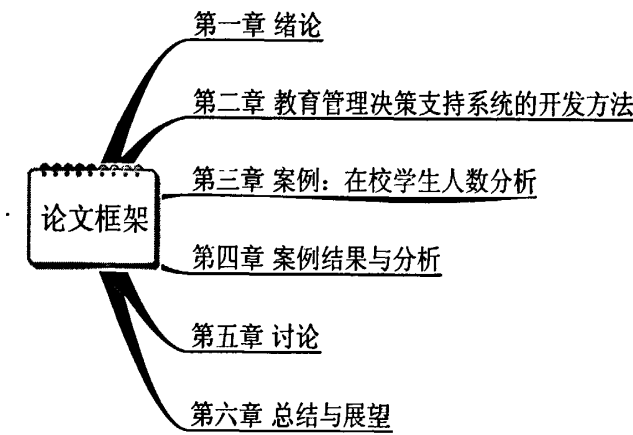


图 1-1 论文整体框架结构

第一章：绪论。主要概述了研究的背景、内容、现状和论文的整体框架结构。其中研究的背景包括了什么是教育管理指标、有哪些重要的教育管理指标，存在着什么样的意义、为什么对教育指标分析以及目前研究面临的主要困难等几个方面。

第二章：教育管理决策支持系统的开发方法。首先，给出了教育管理指标分析的技术模型，并对构成要素过程进行了分析；其次，对商业智能的概念和发展历程进行了综述；再次，对教育管理决策支持系统开发方法中采用的商业智能工具和技术做一说明

第三章：对实施过程的具体分析。以已建立的教育信息管理系统中的数据为基础，以教育管理指标中的在校学生人数这一指标为例对实施过程进行了详细分析。

第四章：案例结果与分析。主要是在第三章具体案例实施的基础上，呈现案例研究的具体结果，并在此基础上对结果进行一定的分析。最后，对教育管理指标分析结果对教育管理的实际意义进行了总结。

第五章：讨论。讨论的主要内容包括：教育管理指标与地方教育主管部门教育规划的制定、关系代数在数据提取中的作用、对数据维度的理解、业务系统中数据质量的监测等。

第六章：总结与展望。

第二章 教育管理决策支持系统的开发方法

决策支持系统是一种基于计算机的信息系统，能够为决策者解决半结构化和非结构化的问题提供支持，以改善决策的水平和质量，并具有交互式、灵活性和适应性强的特点¹。决策支持系统的重点是它能辅助决策者以改进决策的质量，而不是替代决策者实现决策的自动化。教育管理决策支持系统作为教育管理信息系统的一个分支，实际上我们要开发的教育管理决策支持系统，就是一个以计算机为基础，在对学校的教育管理信息进行收集、存储、加工和传输的基础上，将所得到的结果应用于教育管理领域的人机交互系统。

本章主要是根据教育管理者关心的教育现象，从教育管理指标的角度出发，利用商业智能技术实现教育管理指标从数量化向可视化的转变，呈现教育数据可视化实现的过程要素，提供一种利用商业智能技术开发教育决策支持系统的方案，并结合具体的教育管理指标对实施过程进行分析研究。

2.1 教育管理指标分析的技术模型

以已建立的教育信息管理系统中的教育数据为基础，从教育管理指标的角度出发，利用商业智能技术实现教育管理决策支持系统的技术模型，如图 2-1 所示：

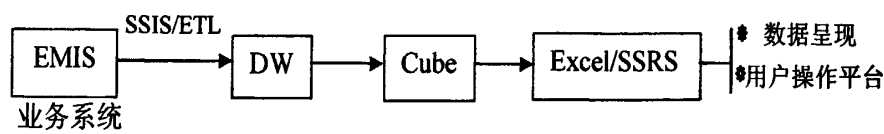


图 2-1 教育管理指标分析的技术模型

教育管理指标分析技术模型的主要操作过程包括：获取数据源—创建数据仓库—填充数据—创建多维数据集—数据呈现/用户操作平台等步骤。

构成过程要素概述：

1. 业务系统²

¹ 李东,梁定澎. 决策支持系统与商务智能[M]. 北京:中国人民大学出版社, 2010. 52.
² 中英甘肃普及九年义务教育项目领导小组办公室, 中英西南基础教育项目办公室.《中英基础教育项目教育管理信息系统使用指南》[M]. 北京:教育科学出版社出版, 2010.

教育管理指标分析技术模型中的业务系统，我们这里使用的是中英基础教育项目中的《教育管理信息系统（Education Management Information System，简称 EMIS）》，如图 2-2 所示，它是一套专门为学校制定的可以用于采集、加工、处理和 分析各教育构成要素等信息的标准化程序。与过去的教育管理信息系统的区别主要 在于它们所服务的对象层次，过去的教育管理信息系统主要是为国家、省、市、县 的教育行政主管部门设计的，重点是为了改善国家或省、市、县的大方向的教育质 量。而中英基础教育项目教育管理信息系统是以学校为基本单位进行设计的，目的 是为了规范和加强学校的日常管理，在学校规范的管理中采集各类教育数据，并可 以通过数据上报的功能，实现各级教育数据的上报、汇总，将其存储在一个综合 的数据库中，为学校 and 各级教育部门提供比较全面、翔实的教育数据。这个数据系 统可以供学校、乡镇和县教育局等多个层面使用。校长和教师可以通过学校层面的 数据，获取改善学校教育质量所需的信息，从而增强教育管理的针对性。对于乡镇 和县级层面的教育行政部门，可以分别利用对应级别的教育数据库对各学校进行直 接、客观的监测与评估，从而为教育决策和规划的制定提供更直接、有效的依据。

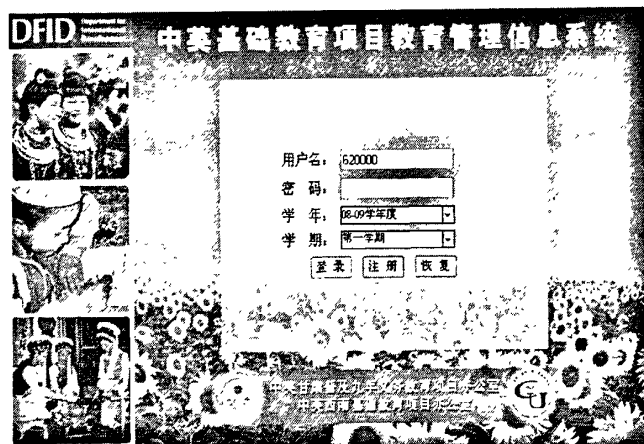


图 2-2 中英基础教育项目教育管理信息系统登录界面

中英基础教育项目教育管理信息系统主要是面向九年义务教育阶段，收集和存 储小学、初中阶段的教育统计指标，其中设置的教育管理内容主要包括学校信息、 教师信息、学生信息、课程信息、学籍信息、考评信息和报表统计等。其中除了报 表统计之外的教育管理内容都可以实现教育构成要素基本信息的录入，而统计报表 则是根据系统中录入的各种教育要素的基本信息进行分类处理产生的，以便能够了

解学校层面的总体数据。

根据教育管理指标需求,开发的教育管理决策支持系统,首要任务就是获取数据源,这里使用的就是中英基础教育项目教育信息管理系统中已存在的数据。我们可以采取两种方式获取需要的数据源:一是,使用 EMIS 单机版获取数据源,通过 EMIS 提供的数据库备份功能,将所有已录入系统的数据进行备份操作,然后下载备份数据到本地磁盘,可得到 Access 格式的数据库,然后将 Access 数据库转换到 SQL Server 数据库;二是,使用 EMIS 网络版获取数据源,EMIS 网络版的配置本身就安装了 SQL Server 及 SQL Server 格式的空数据库,系统经过录入操作之后,它的数据就存储在这个数据库中,可以按照安装空数据库使用的路径去寻找存储了数据的数据库。SQL Server 格式的数据库准备好之后,再将数据库成功添加到 SQL Server Management Studio 中,至此,数据源就可以使用了。

2. 数据仓库

在准备好了数据源之后,还需要建立一个可以用来存储目标数据的数据仓库,以存储实际用到的数据。一般情况下,数据仓库中应包含存储实际数据的事实表和存储维度数据的维度表,并且要设计好每个表所包含的属性名称以及数据类型,表中包含的具体数据允许先为空。事实表中一般包含多个主键,主要是为了建立起与维度表之间关系,而维度表则是从不同的维度对度量组进行分析。

数据仓库(Data Warehouse,简称 DW)是一个面向主题的(Subject Oriented)、集成的(Integrated)、稳定的(Non-Volatile)、随时间变化(Time Variant)的数据集合,以用于支持管理决策过程¹。这是“数据仓库之父”W.H.Inmon 在他的经典之作《建立数据仓库》(Building Data Warehouse)一书中对数据仓库的描述,同时也是众多数据仓库定义中被公认的定义。我们可以从两个方面对数据仓库进行理解:首先,数据仓库是对包含着历史数据的多个异构数据源的有效集成,然后按照特定的主题进行重组并对其进行存放;其次,数据仓库面向分析性数据处理,用于支持管理决策的过程。

总之,数据仓库可以提供用户用于决策支持的数据源,并能把这些数据有效的

¹ William H. Inmon. Building the Datahouse, Fourth Edition[M]. 北京:机械工业出版社, 2007.

集成到统一的环境中以提供决策型数据访问的各种技术和模块，帮助用户更方便、快速的查询所需要的信息，进而为管理者提供决策支持¹。但是，以数据仓库为基础开发的决策支持环境，有时要求数据仓库能够满足所有最终用户的需求，然而各种类型的用户需求是不断变化的，这就需要存储在数据仓库中的数据具有充分的灵活性，能满足各种类型用户的查询和分析，为了适应灵活性的要求，数据仓库需要以规范化的模式存储各种历史数据，而对于特定的用户就需要在很多表的连接中查询所需的数据，就降低了数据仓库的性能，为了解决灵活性和性能之间的矛盾，我们还可以在数据仓库体系中增加数据集市。数据集市（Data Mart）存储了为满足特定用户需求而预先计算好的数据，是一个小型的数据仓库，它能够提高数据仓库中数据的灵活性和性能以满足各种类型用户的需求，所以数据集市是数据仓库特定用户的数据仓库。

在商业智能解决方案中，根据事实表和维度表之间的关系，数据仓库或数据集市的结构常见的模型有星型模型和雪花型模型。在设计逻辑型数据模型的时候，应根据实际数据之间的关系，考虑是按照星型模型还是雪花型模型进行数据的组织。

星型模型是一种多维的数据关系，它由一个事实表和一组维度表共同组成，且所有的维度表都是直接连接到事实表，整个图解像星星一样，故称为星型模型，星型结构图，如图 2-3 所示。

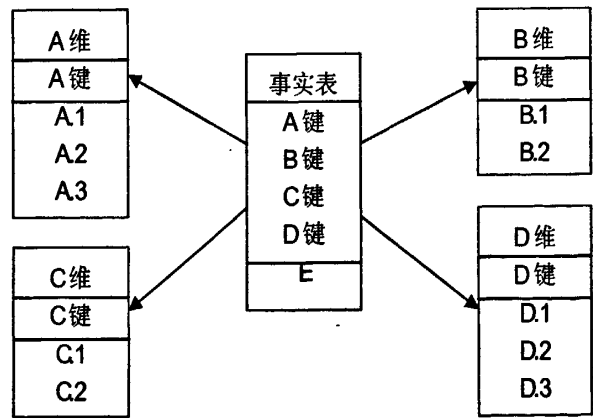


图 2-3 星型结构图

当有一个维度表或多个没有直接连接到事实表上，而是通过其他的维度表连接

¹ 罗建. 可视化数据挖掘方法的研究与实现[D]. 成都: 电子科技大学, 2009.

到事实表上时，即存在父维表和子维表，整个图解就像多个雪花连接在一起，故称为雪花型模型，雪花型结构图，如图 2-4 所示。

雪花型模型是对星型模型的维度进一步层次化的结果，它将原有的各维度表中可能被扩展的表扩展为了小的事实表，形成了一些局部的“层次”区域，这些被分解的表被连接到了父维度表而不是事实表¹。它的优点是最大限度的减少了各表数据的存储量，改善了查询的性能。

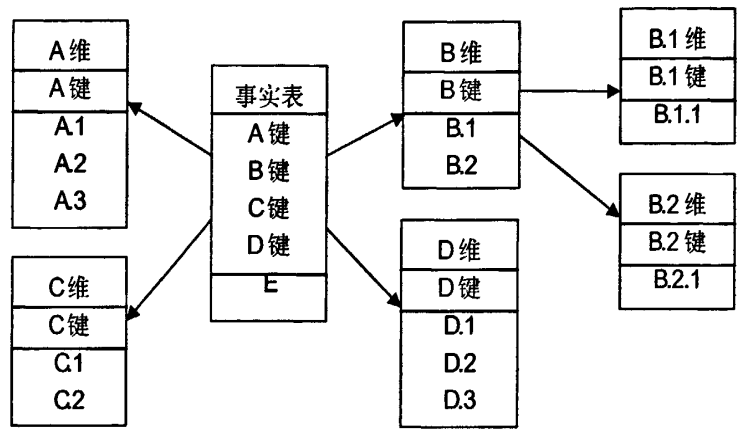


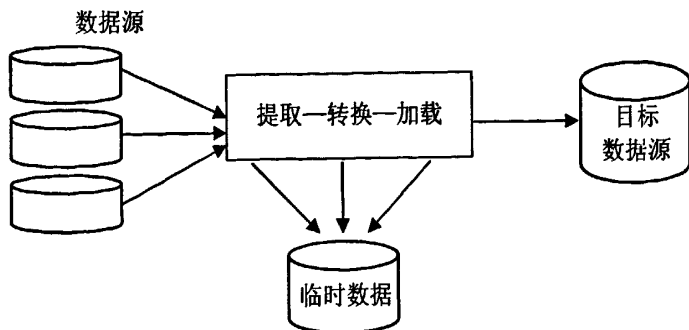
图 2-4 雪花型结构图

3. 数据填充

成功建设好数据仓库或数据集市后，就需要向其中填充数据，即把各种数据源中的数据经过提取、转换、加载（ETL）过程添加到数据仓库中，这个过程中需要注意的是数据仓库中的维度表和事实表中各属性字段类型的匹配问题。

ETL 是 Extract（提取）、Transform（转换）、Load（加载）的缩写，是一整套可以将各类数据源数据提取、转换并加载到多维数据集或数据挖掘结构中的服务，是实现商业智能之前的数据采集的步骤。ETL 的体系结构，如图 2-5 所示。那么，如何在 SQL Server 中来具体实现 ETL 的应用？这里我们使用的 ETL 服务组件是 SQL Server Integration Service（SSIS）。Integration Service（集成服务），顾名思义，就是用来实现数据集成的服务组件，它可以提取和转换多种不同数据源的数据，并将这些数据加载到一个或多个目标中。

¹ 个人图书馆. 星型模型和雪花模型
[EB/OL] .http://www.360doc.com/content/11/0221/20/3356862_94916326.shtml



Integration Service 作为 ETL 在 SQL Server 中的实现渠道,是为数据仓库和数据集市提供数据的主要来源。Integration Service 包含了一个可直接将数据加载到 SQL Server 表和视图中的任务,同时还包含了一个目标组件,该目标组件可以在数据转换过程的最后一步将数据大容量加载到 SQL Server 数据库中¹。在进行数据填充的过程中,要特别注意从外到内的顺序,即先填充各层级的维度表,最后再填充事实表。

4. 多维数据集

在设计好结构良好的数据仓库,并将所需要分析的目标数据填充到数据仓库之后,就为进行商业智能的需求做好了前期准备,后来的操作都是基于填充了数据的数据仓库进行的。然而,对数据的多维分析并不主要是针对数据仓库进行的,而是要从数据仓库中提取子集,我们将其称之为数据立方体。数据立方体是数据仓库实现的一项基本任务,数据立方体的全部或者部分预计算可以大幅度降低响应时间,从而提高联机分析处理的能力²。

建立数据立方体在 SQL Server 中则称为多维数据集,数据立方体中存储的是多维数据集,它允许从多个维度对数据进行建模和分析,一般多维数据集中分为度量值和维度,度量值是用于统计数据,维度是指从哪些方面统计数据。度量值是基于多维数据集的事实表中的一列,通常是数字,并且度量值是分析多维数据集的中心值,是用户最终浏览多维数据集时重点查看的数据;维度是多维数据集的结构特征,

¹ 赵斌. SQL Server 2008 应用开发案例解析[M]. 北京:科学出版社, 2009, 6.

² 一种生成封闭数据立方体的新算法[J]. 弹箭与制导学报, 2010, 6.

它们是事实表中用来描述数据分类的有组织的层次结构¹。

那么,如何在 SQL Server 中实现多维数据集的架构呢?这里我们可以使用的组件是 SQL Server Analysis Service (SSAS)。Analysis Service (分析服务),它提供了分析和查询多维数据集的机制,主要是为了满足分析层的需求。

5. 数据呈现/用户操作平台

选择适合用户的操作平台来呈现数据,是商业智能前端展示的一个重要环节,数据填充部分经过 SSIS 服务组件实现了数据的 ETL 操作,完成了数据源的有效整合和数据的填充,而后通过建立商业智能的 SSAS 项目实现了数据分析所需的数据立方体,在完成了数据提取、转换、加载、整理和分析后,最重要的就是将分析所得通过建立多维数据集,以适合的操作平台呈现给用户,这是商业智能项目的最终目标。

在设置并部署了多维数据集之后,最重要的就是能够从最终用户的角度查看内容。其中最简单的方法就是使用 Microsoft Office Excel 2007 中的透视表视图,打开 Excel 2007,使用功能区中的“数据”选项卡设置到 SSAS 多维数据集的链接。在“获取外部数据”组中,单击“自其他来源”,然后单击“来自分析服务”按钮(如图 2-6),打开一个多步骤的数据连接向导(如图 2-7),输入 SSAS 实例服务器的名称、登录凭据以及从多维数据集生成的透视表和数据的工作簿位置,设置完成后单击“完成”按钮,向导关闭,新设计的透视表将在 Excel 2007 工作簿页面上打开²。

¹ 赵斌. SQL Server 2008 应用开发案例解析[M]. 北京:科学出版社, 2009, 6.

² Lynn Langit, Kevin S. Goff, Davide Mauri 著,张猛,杨越等(译). SQL Server 2008 商业智能完美解决方案[M]. 北京:人民邮电出版社, 2010, 8.

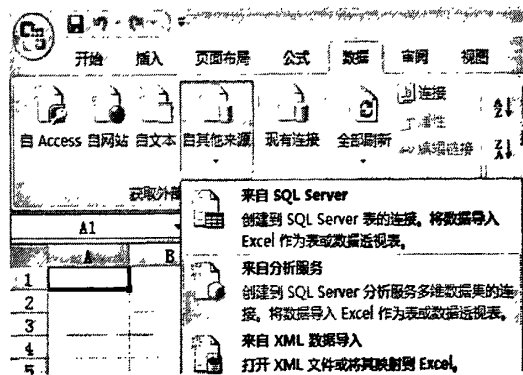


图 2-6 Excel 2007 “数据”功能区的分析服务

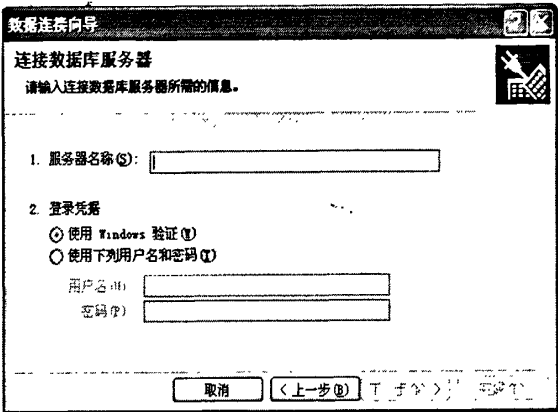


图 2-7 数据连接向导

除此之外，用户还可以使用专门的报表服务实现数据的呈现。SQL Server 2008 四大应用组件中的 Reporting Services（报表服务）就是一种基于服务器的报表平台，它可以用于创建和管理包含多种数据源数据的报表，以呈现多维数据集中的数据。SQL Server Reporting Services（SSRS）提供了各种现成可用的工具和服务，方便人们快捷的制作、部署、发布和管理报表。

2.2 商业智能技术综述

商业智能（Business Intelligence，简称 BI）的概念，最早是 1996 年由加特纳集团（Gartner Group）提出来的。加特纳集团曾这样定义商业智能：商业智能描述了一系列的概念和方法，通过应用于事实的支持系统来辅助商业决策的制定。商业智能提供企业收集、整理和分析数据的技术和方法，并将这些数据转化为有

用的信息,分发到企业的各处¹。根据加特纳集团的定义,商业智能是将与商业运作有关的重要性商业信息搜集、整理和分析,并用来管理商业运作环境的过程,目的是使企业的各级决策者通过这一过程,获得洞察力,以帮助他们做出对企业更有利的策略。当时的商业智能一般定义为由数据仓库、数据分析、数据挖掘和查询报表等部分组成,以帮助企业进行决策而发展的技术应用,其实质就是数据转化为有用知识的过程。

总的来说,商业智能是以数据为基础进行研究的系统,数据是它的根本要素,商业智能系统从开始提出至今大致的发展历程主要可以分为五个阶段。起初,商业智能的定位主要是报表系统,即各企业的用户要求可以通过商业智能将企业业务系统运营的数据有效地展现出来以帮助他们更准确的做决策,主要是批处理预定义查询,简单的说,就是通过商业智能展现的报表可以知道“发生了什么”。随着市场环境的变化和竞争的日益加剧,简单的报表系统以不能满足人们的需求,继而对商业智能系统提出了不同的要求,于是商业智能系统这个阶段具备了一定的分析数据的能力,能够帮助用户进行数据的分析,从而从数据中提取需要的信息,可以进行动态查询操作且这种功能呈逐渐增加的趋势,同时也承担一些决策和业务运营的工作,通过这个阶段我们可以知道“为什么会发生”。2003年开始,商业智能迎来了其第三阶段,特别是国外零售业的快速发展对商业智能系统提出了预测的要求,于是商业智能系统实现了分析、预测和挖掘模型等功能,所以其增加的预测功能能够使我们知道“将会发生什么”。源于MBA的成功案例——沃尔玛案例,商业智能系统发展到了第四个阶段,已经发展成为了一个智能的分析系统,它能够实现数据的持续更新及数据地快速查询响应,通过这时的商业智能系统我们可以知道“正在发生什么”。商业智能系统发展至今,不仅仅是为了能够根据数据的分析结果做业务运营,更重要的是使它发展成为一个实时的数据分析系统,通过商业智能系统能够及时反馈市场信息,起到“数字神经”的作用,从而帮助企业做出更好的决策。总的来说,商业智能系统的发展历程大概经历了报表、分析、预测、操作和动态数据仓库五个阶段,并且还一直处于不断

¹ Bender D, Klein R, Dish A. A Functional Framework for Web Based Information Visualization Systems[J]. IEEE Trans on Visualization and Computer Graphics, 2000, 1.

发展之中。

商业智能虽然经过了很长时期的发展历程,但至今仍没有一个统一的定义,不同的发展阶段对商业智能的定义是不尽相同的。

Microsoft 认为:BI 是任何尝试获取、分析企业数据以更清楚地了解市场和客户改进企业流程和更有效地参与竞争的努力¹。

Gartner 认为 BI 的任务就是获取各种信息源数据,并在其定量分析的基础上,为决策者和组织者提供洞察力²。所以,利用 BI 工具就能够建立帮助高层决策者分析大量客观数据的决策支持系统。近年来,BI 作为当今全球信息技术应用发展的先进技术,被越来越多的用在了决策支持系统的开发中,但事实上,BI 并不是一项新的技术,而是几种不同技术基于计算机的综合应用,经过技术整合后得到的最优化结果能够提供帮助决策的解决方案。

总的来说,商业智能系统就是运用数据仓库、数据挖掘等技术来处理和分析数据,并针对不同领域提供不同的应用解决方案,以协助用户解决日常业务活动中的问题,从而使决策者能够面对不同的环境变化快速、敏捷的做出反应,并制定合理的决策管理系统³。

利用 BI 解决教育决策支持系统的过程就是将来自于教育信息管理业务系统不同的数据源,经过抽取(Extraction)、清理以保证数据的正确性,再将数据经过转换(Transformation)和装载(Load),继而合并在一起存放在数据仓库里,从而得到教育管理信息系统数据的一个全局的视图,在此基础上再利用适当的查询和分析工具、数据挖掘工具等对其进行可视化处理,即将数据转换为可以辅助决策的知识,并最后将所呈现的知识提供给决策者,为决策过程提供支持。本文中利用商业智能的目的是将教育管理指标中涉及的教育数据信息化,并且最大程度地利用商业智能系统将信息化的教育数据转换为信息,最后转换为知识,从而为教育管理者提供最大程度的帮助和支持。但 BI 并不能取代原有的业务系统,而是将其中的数据有效的集成在一起,并通过建模分析,发现隐藏在这些数据背后的知识,了解教育管理的

¹ 黄俊霞.应用商业智能技术构建邮政经营分析系统[J].邮政研究,2006,01.

² 郭艳蕾.商业智能综述[J].商业文化,2008,2.

³ 张云涛,龚翎.商业智能设计、部署与实现[M].北京:电子工业出版社,2004.

现状，把握教育发展的趋势等。

通过阅读大量关于商业智能的文献，发现实现商业智能的技术有很多种，综合考虑它们的成熟度、应用状况、发展前景等方面，这里选择的是我们熟悉的微软的商业智能解决方案。重要的是，微软本身就具有一套完整的商业智能开发工具，可以通过 Microsoft SQL Server Management Studio、Microsoft SQL Server Business Intelligence Development Studio、Office 等其他一些工具紧密集成共同实现商业智能。微软的商业智能之所以具有明显的优势，主要是它拥有优异的技术集成性

（Integration Service、Analysis Service、Reporting Service 等）、OLAP 技术的市场优势、与 Office 系统的紧密集成（Excel 数据透视表等）、众多的合作伙伴、持续的研发投入等。完整的微软 BI 涉及了后端业务数据建模与存储相关的数据抽取、转换与加载、数据仓库、数据集市、数据挖掘以及前端业务桌面可视化呈现、分析与报表等多个模块¹。Microsoft BI 解决方案构架，如图 2-8 所示^{2, 3}：

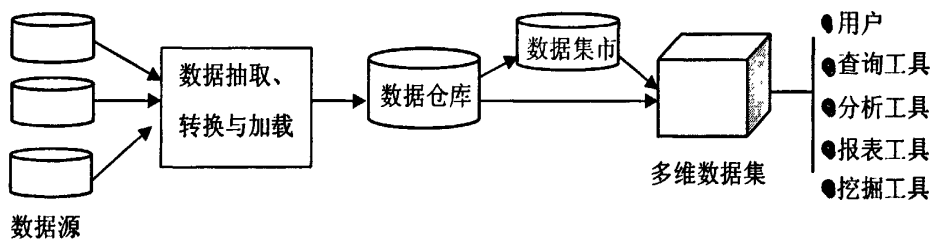


图 2-8 Microsoft BI 解决方案构架

2.3 采用的 BI 工具和技术

Microsoft BI 解决方案的核心组件主要包括 SQL Server 2008、SQL Server Integration Service（SSIS，数据集成服务）、SQL Server Analysis Service（SSAS，数据分析服务）和 SQL Server Reporting Service（SSRS，数据报表服务），其中 Integration Service、Analysis Service、Reporting Service 和数据库引擎并称为 SQL Server 的四大组件。SQL Server 通常是被用作数据源，或者是用在加载数据到多维数据集时，作

¹ 豆丁网. 微软商务智能解决方案 V1.0[EB/OL]. <http://www.docin.com/p-37643564.html>.
² 同上
³ 郭君芳. 教育信息数据的可视化研究[D]. 上海:上海师范大学, 2010.

为数据的临时存储位置对数据进行验证。Integration Service 主要是用于数据的集成和整合，是一个 ETL 工具集，Analysis Service 主要是用于数据分析和数据挖掘，这些服务都具有相同的开发界面 Business Intelligence Development Studio (BIDS)，使用相同的管理界面 Microsoft SQL Server Management Studio (SSMS)，而且每个组件都提供了用于创建商业智能解决方案所需要的对象模板，且提供了用于处理这些对象的各种设计器、工具和向导。通常在使用 SQL Server 2008 建立 Microsoft BI 解决方案的时候，这些核心的组件至少会使用一个，下面就对本文应用案例中可能会用到的 Microsoft BI 工具和技术做一介绍。

2.3.1 数据库技术平台

Microsoft BI 解决方案中使用的数据库技术平台是 SQL Server Management Studio，它是 SQL Server 2008 中最常用最核心的功能组件，这里我们介绍如何利用 SSMS 建立数据库以存储数据。首先，连接到 SQL Server 服务器。在 SQL Server 2008 程序中启动“SQL Server Management Studio”，将会出现“Connect to Server”（连接到服务器）的对话框，如图 2-9 所示，选择“Database Engine”（数据库引擎）类型的服务器，而服务器的名称则为 SQL Server 所在的机器的名字，一般默认为本机，但也可以选择局域网范围中的其他装有 SQL Server 的机器，前提是要具有 Windows 的登录权限。Authentication（身份验证）保持为默认状态：Windows 身份验证（在安装 SQL Server 2008 的过程中已经进行了配置）。点击“Connect”（连接），将打开 SQL Server Management Studio 的主界面，如图 2-10 所示。在界面的左边是“Object Explorer”（对象资源管理器），在对象资源管理器中打开每一个子节点呈现出的是相对应的一些主要操作，在界面的右边是“Object Explorer Details”（对象资源管理器详细信息），如果看不到右边的面板，我们可以在菜单栏中的“视图”中将其调出。

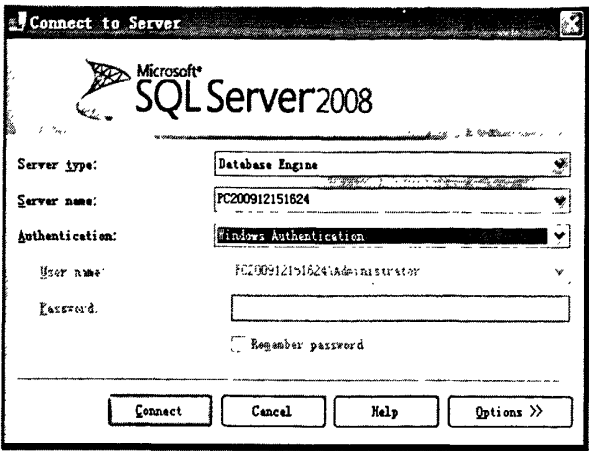


图 2-9 “连接到服务器”对话框

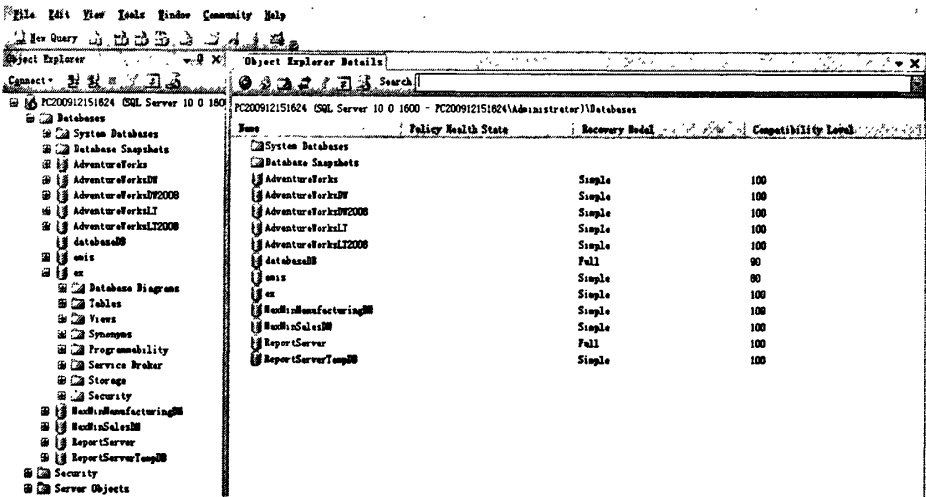


图 2-10 Management Studio 主界面

使用 Management Studio 创建数据库的方法就很简单了，选择“对象资源管理器”中的“Databases”（数据库）节点，右键选择“新建数据库”选项，即可开始创建新的数据库。成功创建数据库后，可以展开已创建的数据库节点，找到“Tables”（表），右键选择“新建表”选项，创建数据库中所要包含的表，等等。

2.3.2 数据集成技术平台

SSIS（SQL Server Integration Service）是 Microsoft BI 解决方案中的关键组件，是实现集成服务的主要组件，它可以用于向 SSAS 的多维数据集或数据挖掘结构提供数据之前，对数据进行导入、清理和验证，并将数据加载到数据仓库、数据集市或者直接载入到分析服务。我们也可以利用 Integration Service 集成服务创建、修改

和删除分析数据库对象。当数据仓库中的数据来源于多个异构数据源时，可以通过 SSIS 的工具集协助 Microsoft BI 解决方案以实现数据的 ETL 工作。

SSIS 的执行单元称为包 (package)，是 SSIS 体系中一个非常重要的对象。它是经检索、执行和保存的工作单元，通常情况下，一个包中包含有控制各个任务先后执行顺序的控制流和控制实现数据从数据源到目标数据地址流动和转换的数据流。SSIS 包通过设置为可重新启动，即从某个预先规定的检查点重新运行包，可以节省执行的时间，特别是在处理来自多种不同数据源的数据时。Integration Service 的包通常是由一个或多个任务组成的，这些任务就是一些控制流元素，当包中包含多这个任务时，它们将按照事先设定的优先约束在控制流中进行连接和排序。其中的数据流任务是一种特殊的任务，它可以实现用户对数据的转化、清除和修改，数据流任务可以实现数据从数据源提取，然后经过转换，再到目标数据的加载过程，即实现数据填充。

在 SQL Server 2008 程序中选择“SQL Server Business Intelligence Development Studio”选项，将会出现 BIDS 的界面，选择菜单栏中的“File”（文件）—“New”（新建）—“Project”（项目）选项，将会出现“New Project”（新建项目）对话框，如图 2-11 所示。

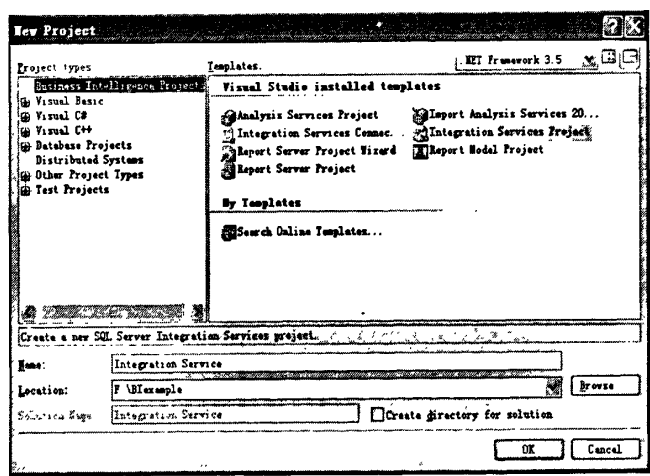


图 2-11 新建 Integration Service 项目对话框

在“Business Intelligence Projects”（商业智能项目）选项框中选择“Integration Service Project”，对此项目进行命名并且设置其合适的存储位置，选择确定按钮，即可进入创建 Integration Service 项目的主界面，如图 2-12 所示。

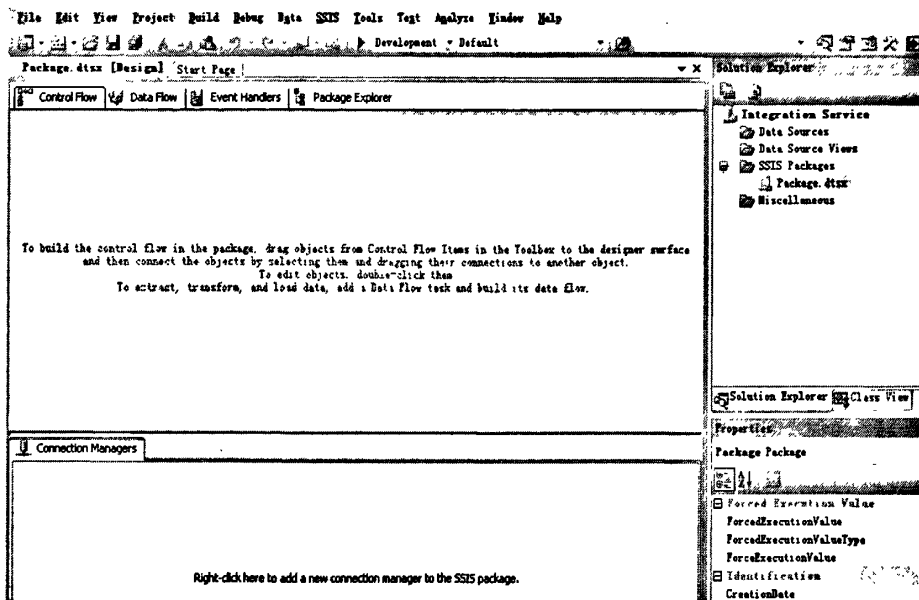


图 2-12 Integration Service 主界面

系统为 Integration Service 项目自动创建一个名为 Package.dtsx 的包,也就是 SSIS 执行的单元,左边的设计器中包含了“Control Flow”(控制流)、“Data Flow”(数据流)、“Event Handlers”(事务处理程序)、“Package Explorer”(包资源管理器)四个标签页,实现数据填充的过程主要用到的是控制流和数据流两个标签页,以实现数据的链接。界面的下方是“Connection Managers”(连接管理器),用户可以在这里添加连接数据源和数据目标的组件,以实现数据的填充。

2.3.3 数据立方的技术平台

数据立方体在 SQL Server 中被称为多维数据集,建立多维数据集我们使用的是 Microsoft BI 解决方案的核心服务—SSAS (SQL Server Analysis Service),它提供了可用于数据仓库存储和查询多维数据集的机制,也可以用来创建包含数据挖掘模型的数据挖掘结构,主要是为了满足分析层的需求。同样, Business Intelligence Development Studio 也是用来开发 SSAS 多维数据集的主要工具。

在 SQL Server 2008 程序中选择“SQL Server Business Intelligence Development Studio”选项,将会出现 BIDS 的界面,按照新建 Integration Service 项目相同的步骤新建 Analysis Service 项目,但在“Business Intelligence Projects”(商业智能项目)选项框中要选择“Analysis Service Project”,对此项目进行命名并且设置其合适的存

储位置，选择确定按钮，成功创建 Analysis Service 项目后，在“Solution Explorer”（解决方案资源管理器）可以看到 Analysis Service 项目包含的各个组成部分，如图 2-13 所示。

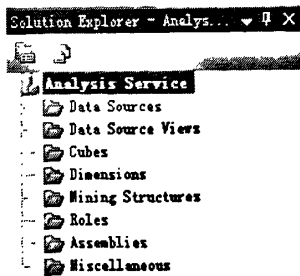


图 2-13 “解决方案资源管理器”窗口

在创建具体的数据挖掘模型和数据立方体时，我们还需要创建数据源和数据源视图，即可在“解决方案资源管理器”窗口中分别右键选择“Data Sources”（数据源）和“Data Source Views”（数据源视图），新建数据源和数据源视图。数据源是用来连接目标数据库的，数据源视图用来连接数据库中对应的数据表和视图。同样的方法，在“解决方案资源管理器”窗口中右键“Cubes”（多维数据集），选择“New Cubes”（新建多维数据集），通过选择不同的维度，来创建不同的多维数据集。

第三章 案例：在校学生人数分析

本文的主要目标是利用 BI 技术实现教育管理指标分析的研究,这里我们以在校学生人数这一教育管理指标为例,呈现利用 BI 技术实施教育管理指标分析的详细过程。对在校学生人数指标的分析一方面主要集中在在校学生人数的各种分布情况,如按民族、性别、地域等,为教育管理决策人员提供真实的学生数据,以帮助其提高教育管理的效率。一方面可以对各种不同分布情况的结果进行比较和分析,同时为学校、教师和班级等多种相关的教育元素的调整,提供基础数据的支持。

3.1 开发环境配置

利用 BI 技术实现在校学生人数分析的实施过程,这里我们选择使用是微软的 BI 解决方案,它是一个基于 Microsoft Visual Studio 2008 开发环境的 Business Intelligence Development Studio 工具,其中包含特定的用于 SQL Server 2008 商业智能的多种技术。具体的开发环境需要安装以下组件和工具:

1. SQL Server 2008 企业版

在安装过程中要注意选择数据库引擎、集成服务、分析服务等项目并进行安装。

2. Visual Studio 2008 及升级包

开发环境成功配置后,程序中将会出现 Microsoft Visual Studio 2008 和 SQL Server Business Intelligence Development Studio 的开发环境,在这里我们可以利用丰富的商业智能的套件新建所需的项目,如 2.3.2 和 2.3.3 所示。

3.2 业务系统数据分析

教育管理决策支持系统分析的数据源主要是中英西南教育管理信息系统中的数据,为了方便建立存储目标数据的数据库,需要围绕在校学生人数的指标,以数据字典为依据,对 EMIS 数据库进行详细分析,通过分析发现学生的信息是由所在班级、学号、姓名、性别、民族和入学年月共同决定的,它们共同组成了学生的事实

表 (Fact Students)。依据建立 EMIS 业务系统时的数据字典分析 EMIS 数据库，可以方便地列出与学生信息相关的各表的主要数据存储情况：

1. XS_XSZB (学生总表)

学生总表主要是记录每个学生基本信息的表，存储着学生的具体信息。该表中的主要字段包括：OC_BJZB (当前班级主键 ID)、C_XH (学号，学生在校就读期间的唯一标识)、C_XM (姓名)、OC_XB (性别)、OC_MZ (民族码)、DT_RXRQ (入学年月)。

2. XX_XXZB (学校总表)

学校总表主要是记录每个学校基本信息的表，存储着学校的具体信息。该表中的主要字段包括：C_XXDM (学校代码)、C_XXMC (学校名称)、OC_SJZGDW (学校所属主管单位代码)、C_DWMC (主管单位名称)。

3. BJ_BJZB (班级总表)

班级总表主要是记录学校每个班级基本信息的表，存储着学校各班级的具体信息。该表中的主要字段主要包括：C_ID (班级代码)、C_BJMC (班级名称)。

4. dm_tydm (通用代码)

通用代码表可以用来和其他一些含有代码的表进行连接，将其中的代码转换为名称，如学生总表中的性别码、民族码，进行和通用代码表的链接可以将其转换为对应的名称。通用代码表中主要存储的是一些代码和相对应的名称，包含的主要字段有：C_DM (代码)、C_MC (名称)。

通过对业务系统学生信息数据的分析，对涉及到的主要表、表中的主要字段以及表与表之间的关系有了一个大概的了解。但业务系统中数据是系统开发人员按照某种特定的法则进行定义的，对于普通的用户来说，它所表示出来的意义不是容易读懂的，为了创建一个用户易懂的数据集市，我们需要对业务系统中各表主要字段进行重新定义，并对各表中的字段进行适当添减。

通过对业务系统数据的分析研究，相关的表主要有学生表 (表 3-1)、学校表 (表 3-2)、班级表 (表 3-3)，然后根据通用代码中关于性别和民族的元组，将学生表中的代码进行适当的转换，这里添加需要添加两个表，我们将其定义为民族表 (表 3-4) 和性别表 (表 3-5)。

表 3-1 学生表

字段名称	数据类型	允许空值
ClassID	nvarchar(50)	☐
StudentID	nvarchar(50)	☐
Name	nvarchar(50)	☐
Gender	nvarchar(50)	☐
EthnicGroup	nvarchar(50)	☐
EnrollmentYear	nvarchar(50)	☐

学生表中存储的主要字段为：ClassID（当前所在班级代码）、StudentID（学号）、Name（姓名）、Gender（性别）、EthnicGroup（民族）、EnrollmentYear（注册学年）。

表 3-2 学校表

字段名称	数据类型	允许空值
SchoolID	nvarchar(50)	☐
Name	nvarchar(50)	Checked
DistrictLocation	nvarchar(50)	Checked
DistrictName	nvarchar(50)	Checked

学校表中存储的主要字段为：SchoolID（学校代码）、Name（学校名称）、DistrictLocation（学校所在乡镇代码）、DistrictName（学校所在乡镇名称）。

表 3-3 班级表

字段名称	数据类型	允许空值
ClassID	nvarchar(50)	☐
Name	nvarchar(50)	☐
SchoolID	nvarchar(50)	☐

班级表中存储的主要字段为：ClassID（班级代码）、Name（班级名称）、SchoolID（学校代码），此处的 SchoolID 是作为外键，用来连接学校表的，这样可以将有关学校的信息从班级表中单独分离出来存储在学校表中，以减少班级表存储数据量。

表 3-4 民族表

字段名称	数据类型	允许空值
EthnicGroupID	nvarchar(50)	☐
Name	nvarchar(50)	☐

民族表中存储的主要字段为：EthnicGroupID（民族码）、Name（民族名称）。

表 3-5 性别表

字段名称	数据类型	允许空值
GenderID	nvarchar(50)	☐
Gender	nvarchar(50)	☐

性别表中存储的主要字段为：GenderID（性别码）、Gender（性别）。

3.3 创建数据集市

1. 数据集市概念模型设计

本案例研究的数据库主要是和学生信息有关的数据，重点研究的是在校学生人数使用的业务系统数据，属于某个具体问题的分析与研究，根据数据仓库和数据集市的描述，案例的实际开发过程中可以采用建立数据集市即面向主题的小型仓库的方法解决问题。数据集市概念模型设计的关键问题就是需要明确事实表和维度表，事实表主要是用来存储实际数据，维度表主要是用来存储事实表中相关维度的信息。

通过对在校学生人数涉及到的数据表之间的关系按照事实表和维度表进行分析，我们发现学生表记录的主要是学生的实际数据信息，而其他表所记录的主要是学生信息的不同纬度的数据，所以将学生表确定为事实表（FactStudents），学校表、班级表、民族表和性别表定为记录学生信息的维度表（DimSchool、DimClass、DimEthnicGroup、DimGender）。FactStudents 表中有多个主键（ClassID、StudentID），可以建立与维度表之间的关系，维度表主要是从不同的维度对在校学生人数进行分析。DimClass 中主要涉及到 ClassID、Name、SchoolID 三个属性，而 SchoolID 既是 DimClass 的外键，又是 DimSchool 的主键，由此可知 DimClass 是 DimSchool 的父维度表，如图 3-1 所示。

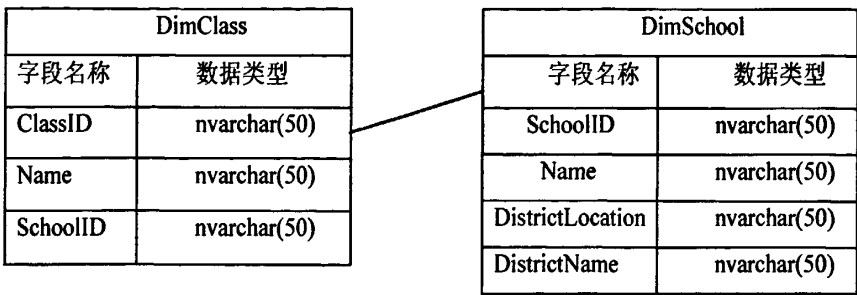


图 3-1 子维表与父维表之间的关系

前面我们已经介绍了数据集市的主要结构：星型结构图和雪花型结构图。通过分析已经确定案例中涉及到了子维表与父维表，并综合考虑各表数据的存储量，为了能最大程度的改善查询的效率，这里我们将采用雪花型结构图，建立数据集市，如图 3-2 所示。

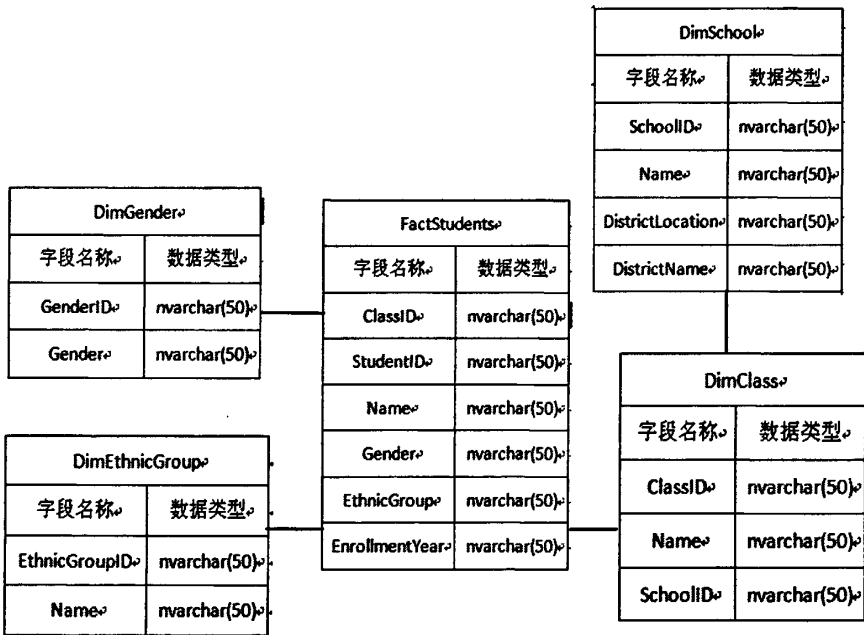


图 3-2 雪花型数据集市

2. 创建数据集市

设计好数据集市的概念模型后，就可以进行数据集市的实际创建工作。根据微软 BI 所提供的技术和工具，可以采用两种方法进行数据集市的创建。一是，利用 SQL Server 2008 提供的 Management Studio 工具，手动创建数据集市；一是，利用 SQL Server Business Intelligence Development Studio 工具，实现数据集市的自动创

建，它采用的是自上而下的方法，并能自动创建数据集市中的事实表和维度表。

(1) 利用 SQL Server Management Studio 工具手动创建数据集市 EMISDM

2.3.1 中我们已经介绍了利用 SQL Server Management Studio 工具创建数据库的基本方法,这里所要创建的数据集市和数据库本身没有很大的区别,即可以按照 SQL Server Management Studio 工具创建数据库的方法来创建数据集市。具体过程请参考 2.3.1。

在利用 SQL Server Management Studio 工具创建数据集市的过程中，要注意选择数据集市的存放位置和恢复模式，并将其设置为“简单”模式。在创建数据集市中的事实表和维度表的时候，要注意的是各表中主键和外键的创设问题，并根据实际情况创设事实表的复合主键,正确设置各表的属性及其数据类型也是非常重要的。成功创建 EMISDM 数据集市后，数据集市如图 3-3 所示，事实表如图 3-4 所示，维度表如图 3-5、图 3-6、图 3-7、图 3-8 所示。

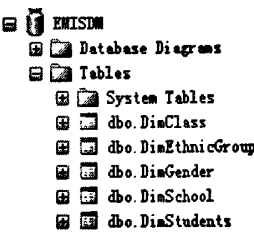


图 3-3 EMISDM 数据集市

PC200912151...etStudents*			
Column Name	Data Type	Allow Nulls	
ClassID	nvarchar(50)	☐	
StudentID	nvarchar(50)	☐	
Name	nvarchar(50)	☐	
Gender	nvarchar(50)	☐	
EthnicGroup	nvarchar(50)	☐	
EnrollmentYear	nvarchar(50)	☐	

图 3-4 FactStudents 学生事实表

PC200912151...o.DimClass*			
Column Name	Data Type	Allow Nulls	
ClassID	nvarchar(50)	☐	
Name	nvarchar(50)	☐	
SchoolID	nvarchar(50)	☐	

图 3-5 DimClass 班级维度表

PC200912151... o. DimSchools		
Column Name	Data Type	Allow Nulls
SchoolID	nvarchar(50)	☐
Name	nvarchar(50)	☒
DistrictLocation	nvarchar(50)	☒
DistrictName	nvarchar(50)	☒

图 3-6 DimSchool 学校维度表

PC200912151... thnicGroup*		
Column Name	Data Type	Allow Nulls
EthnicGroupID	nvarchar(50)	☐
Name	nchar(10)	☒

图 3-7 DimEthnicGroup 民族维度表

PC200912151... o. DimGenders*		
Column Name	Data Type	Allow Nulls
GenderID	nvarchar(50)	☐
Gender	nvarchar(50)	☐

图 3-8 DimGender 性别维度表

利用 SQL Server Management Studio 工具手动创建数据集市，主要是按照设计好的雪花型结构图一步一步来创建的，设计的字段的数据类型和业务系统数据分析中的各表字段是一致的，减少了数据的不匹配问题，但所需要的时间比较长，效率比较低。而利用 SQL Server Business Intelligence Development Studio 工具，可以采用自上而下的方式自动创建数据集市包含的事实表和维度表，所用时间会大大减少，提高创建数据集市的效率。

(2) 利用 SQL Server Business Intelligence Development Studio 工具创建数据集市 EMISDM

利用 Business Intelligence Development Studio 工具创建数据集市，主要是通过创建 SSAS 项目的多维数据集实现，它能够快速的创建 OLAP 多维数据集及包含事实表和维度表的数据集市。首先，需要利用 SQL Server Management Studio 工具创建一个不包含任何事实表和维度表的空的数据集市 EMISDM。然后，按照 2.3.3 中介绍的利用 BIDS 工具新建 SSAS 项目的步骤创建一个多维数据集 EMISSAS，采用的方法是“Generate tables in the data source”（在数据源中生成表）。

在利用 SSAS 分析服务创建多维数据集的过程中，需要注意的关键问题主要有：
Measures（度量值）：在多维数据集中，度量值是基于多维数据集的事实数据表中的一列，通常为数字，它是所分析的多维数据集的中心值，是最终用户浏览多维

数据集时重点查看的数字数据1。简单的说，度量值就是所要研究的问题，案例中也就是在校学生人数，设置为“Fact Students Count”，如图 3-9 所示。

Add new measures:

Measure Name	Measure Group	Data Type	Aggregation
Fact Students Count	Fact Students	Integer	Count ☒

图 3-9 在校学生人数的度量值

Dimensions（维度）：维度是事实表中用来描述数据分类的有组织层次的结构，是分析所研究问题的角度。对于在校学生人数这一问题，我们选择从三个维度对其进行分析，分别是 Dim Class、Dim EthnicGroup、Dim Gender，案例中不涉及时间的问题，这里不需要选择时间维度，如图 3-10 所示。

Select dimensions from template:

Type	Name	SCD	Attributes
☐ Time	Time	☐ Days	

Add new dimensions:

Name	SCD	Attributes
Dim Class	☐	☒
Dim Ethnic Group	☐	☒
Dim Gender	☐	☒

图 3-10 在校学生人数维度分析

通过前面雪花型结构数据集市的模型设计，以及度量值与维度的关系，可知在校学生人数的度量值涉及到三个维度，如图 3-11 所示。

Dimension usage:

Dimension	Fact Students
Dim Class	☒
Dim Ethnic Group	☒
Dim Gender	☒

图 3-11 定义维度用法

利用 SSAS 成功创建多维数据集，如图 3-12 所示。

¹ 赵斌. SQL Server 2008 应用开发案例解析[M]. 北京:科学出版社, 2009, 507.

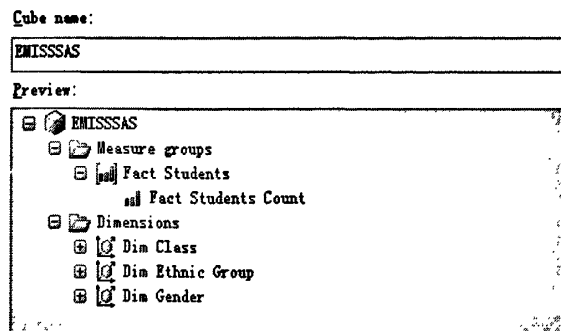


图 3-12 EMISSAS 多维数据集结构图

成功创建多维数据集之后，就需要将数据集中设置的度量值及维度存储到利用 SQL Server Management Studio 工具已创建的数据集市 EMISDM 中。

首先，需要将多维数据集的数据源和已创建的数据集市 EMISDM 进行连接，即创建数据源。在数据源向导中选择已创建的数据集市 EMISDM 以建立数据连接，并选择“Use a specific Windows user name and password”（使用特定 Windows 用户名和密码）的连接方式，以确保后续工作的正常运行。成功创建数据源后，将会显示连接字符串：

Provider=SQLNCLI10.1;Data

Source=PC200912151624;Integrated

Security=SSPI;Initial Catalog=EMISDM

然后，将多维数据集中设置的度量值及维度存储到已创建的数据集市 EMISDM 中，即创建数据视图。本案例按照架构生成向导创建数据视图时，基本上是将向导保持为默认状态，当架构成功生效时，即成功创建数据视图，如图 3-13 所示。

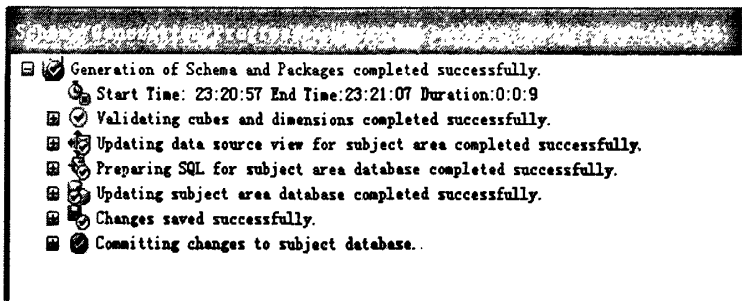


图 3-13 成功生效架构图

最后，利用 SQL Server Management Studio 工具打开之前创建的数据集市 EMISDM，将会发现已成功将事实表、维度表、各属性字段及数据类型填充到了 EMISDM 中，如图 3-14 所示。

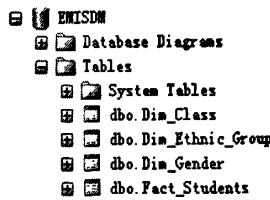


图 3-14 EMISDM 中自动创建的表

利用 SQL Server Business Intelligence Development Studio 工具自动创建的数据集市 EMISDM 中的各表,与前面设计的雪花型结构模式的在校学生人数的事实表和维度表的表示方法有所不同,但实质是一样的,只是表的名称和字段有所差异。这个 EMISDM 数据集市中的 Fact_Students 数据表就是事实表,而 Dim_Class 数据表是 DimClass 维度表,Dim_Ethnic_Gourp 数据表是 DimEthnicGroup 维度表,Dim_Gender 数据表是 DimGender 维度表。打开 EMISDM 数据集市中的每个数据表将会发现这些数据表中的字段非常少,我们需要根据前面设计的事实表和维度表,进一步完善这些数据表。具体需要添加的表及属性参照已设计的在校学生人数雪花型数据集市(如图 3-2)。

至此,利用两种不同的方法都已成功创建了数据集市,但是通过查询 EMISDM 数据集市中的事实表和维度表,发现所有的表都是空的,是没有任何数据存在的。没有数据的数据集市是没有意义的,接下来就要向空的数据集市中填充数据。

3.4 数据填充

在 2.1 节中我们已经介绍了,可以利用 BIDS 工具的 SSIS(集成服务)项目完成对数据集市 EMISDM 的数据填充。首先,按照 2.3.2 中介绍的利用 BIDS 工具新建 SSIS 项目的步骤创建一个名为 EMISSSIS 的项目。然后,就可以利用工具箱中提供的工具组件和连接管理器向各数据表中填充数据。下面以填充 DimClass 维度表数据为例,进行详细说明:

1. 利用 BIDS 工具新建一个名为 EMISSSIS 的 SSIS 项目。
2. 在控制流的工具箱中拖拽一个名为“数据流任务”的工具组件到控制流标签页中,并将其命名为 Class,双击设计好的“数据流任务”进入到数据流标签页。
3. 在 SSIS 主界面下方的“Connection Managers”(连接管理器)中,右击新建

ADO.NET 连接，在“配置 ADO.NET 连接管理器”中选择连接到分析业务系统时准备好的数据源，也就是用作数据填充的源数据，如图 3-15 所示，可以通过“Test Connection”（测试连接），验证数据库连接是否成功。

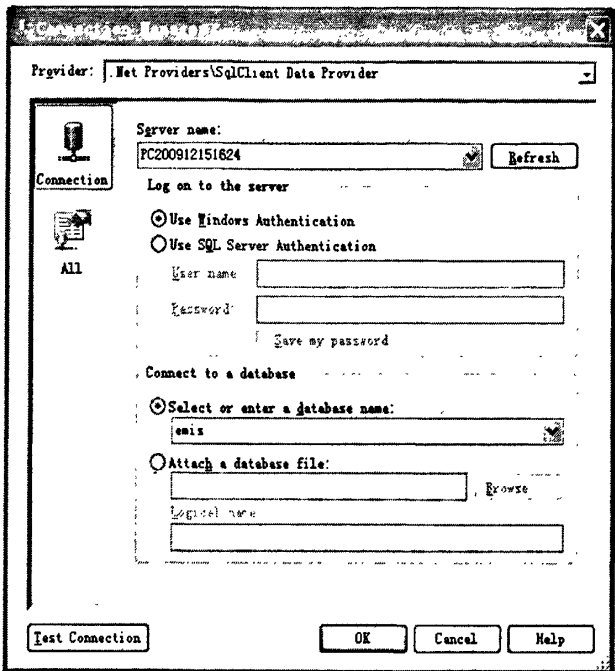


图 3-15 连接管理器

4. 在数据流工具箱中拖拽一个“ADO.NET Source”（ADO.NET 源）到数据流标签页中，双击“打开 ADO.NET 源编辑器”，ADO.NET 连接管理器中默认的是已连接的 ADO.NET 数据源，数据访问模式选择使用“SQL Command”（SQL 命令），输入相应的 SQL 查询语句，即可成功创建需要的 DimClass 数据源流，如图 3-16 所示，点击“Preview”（预览）即可查询。

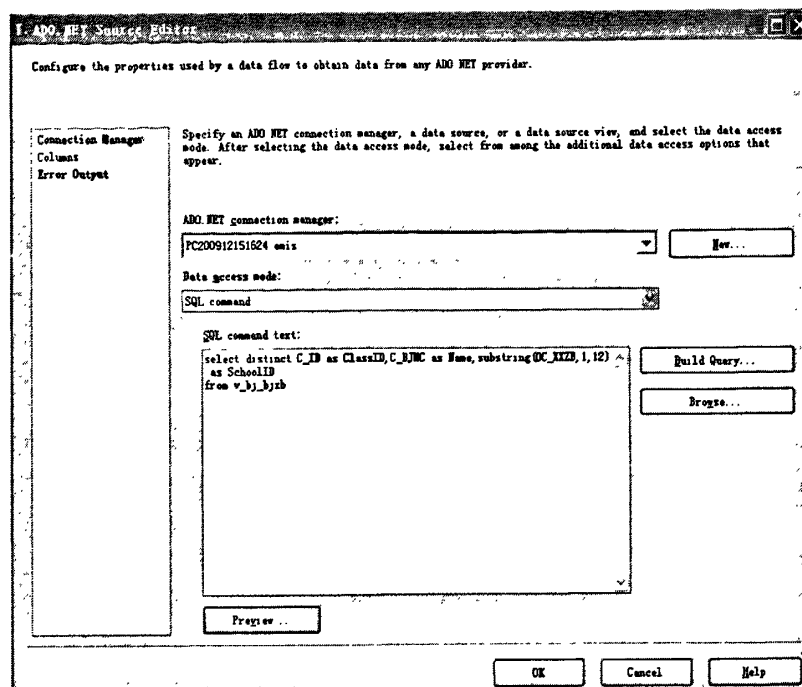


图 3-16 源编辑器

5. 在“连接管理器”中，右击新建 OLE DB 连接，在“配置 OLE DB 连接管理器”中选择连接到需要进行数据填充的数据集市 EMISDM。
6. 在数据流工具箱中拖拽一个“SQL Server Destination”（SQL Server 目标）到数据流标签页中，选择“ADO.NET 源”中的绿色箭头连接到“SQL Server 目标”，双击“SQL 目标编辑器”，连接管理器中默认的是需要填充的数据集市 EMISDM，数据访问模式选择默认状态：表或者视图，这里选择对应的表“dbo.DimClass”。选择“映射”选项，可以查看“输入列”与“目标列”的对应关系，当两者的数据类型完全匹配时，会自动完成“输入列”到“目标列”的映射，如图 3-17 所示。点击“预览”，可看到表“dbo.DimClass”中的查询结果。

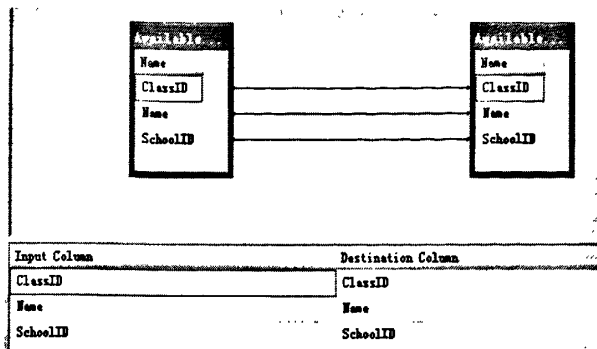


图 3-17 列映射图

至此，就完成了从数据源中提取数据，并将数据传输到 DimClass 维度表中的操作，如图 3-18 所示。然后，点击菜单栏中的“运行”按钮，即可向数据集市 EMISDM 的 DimClass 维度表中填充数据，当“ADO.NET 源”和“SQL Server 目标”都呈现绿色状态时，表示数据填充成功。

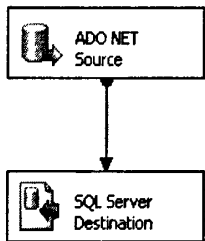


图 3-18 数据源到数据目标的数据流

按照同样的步骤，依次完成表 DimGender、DimEthnicGroup、DimSchool 和 FactStudents 中数据的填充，由于案例中使用的数据源和需要数据填充的数据集市 EMISDM 是相同的，所以这里可以省去步骤 3 和步骤 5。

数据流在实现数据从数据源到数据目标的传输中，可能会涉及到数据的转换和清理，即经过 ETL 操作实现数据填充。在所有表中数据成功填充后，我们需要根据前面设计的在校学生人数的雪花型模式关系图，建立各表之间从低到高的优先级别权限。

接下来，可以按照 2.3.1 中的步骤打开 SQL Server Management Studio 的主界面，选择数据集市 EMISDM，通过 SQL 语句对 EMISDM 中各表进行数据的查询，以检验数据目标 EMISDM 是否成功填充数据。

查询 DimClass 维度表中数据填充是否成功，SQL 查询代码如下：

```
SELECT [ClassID]
      ,[Name]
      ,[SchoolID]
FROM [EMIS_SSAS].[dbo].[DimClass]
```

通过对查询结果的检验，发现数据已成功从数据源中加载到数据集市 EMISDM 中的 DimClass 维度表中，执行结果如图 3-19 所示。

Results Messages			
	ClassID	Name	SchoolID
1	5306280010061120040101	04级1班	530628001006
2	5306280010061120040102	04级2班	530628001006
3	5306280010061120040103	04级3班	530628001006
4	5306280010061120050101	05级1班	530628001006
5	5306280010061120050102	05级2班	530628001006
6	5306280010061120050103	05级3班	530628001006
7	5306280010061120050104	05级4班	530628001006
8	5306280010061120060101	06级1班	530628001006
9	5306280010061120060102	06级2班	530628001006
10	5306280010061120060103	06级3班	530628001006
11	5306280010061120070101	07级1班	530628001006
12	5306280010061120070102	07级2班	530628001006
13	5306280010061120080101	08级1班	530628001006
14	5306280010061120080102	08级2班	530628001006
15	5306280010061120090101	09级1班	530628001006
16	5306280010061120090102	09级2班	530628001006

图 3-19 DimClass 维度表中的部分元组

查询 DimSchool 维度表中数据填充是否成功，SQL 查询代码如下：

```
SELECT [SchoolID]
      ,[Name]
      ,[DistrictLocation]
      ,[DistrictName]
FROM [EMIS_SSAS].[dbo].[DimSchool]
```

DimSchool 维度表数据成功填充，执行结果如图 3-20 所示。

Results		Messages
SchoolID	Name	DistrictLocation DistrictName
46	530628001325 坪政村林口小学	530628001 角奎镇
47	530628001326 坪政村妈逗小学	530628001 角奎镇
48	530628001328 彝良县角奎镇杉林小学	530628001 角奎镇
49	530628001329 彝良县角奎镇杉林村中寨小学	530628001 角奎镇
50	530628001330 彝良县角奎镇杉林村海子小学	530628001 角奎镇
51	530628001333 角奎镇花溪小学	530628001 角奎镇
52	530628001334 角奎镇和平小学	530628001 角奎镇
53	530628007205 树林乡碗厂小学	530628007 树林乡
54	530628007206 彝良县树林乡碗厂村河沟小学	530628007 树林乡
55	530628007207 树林乡碗厂村水塘小学	530628007 树林乡
56	530628007208 树林村鹰山小学	530628007 树林乡

图 3-20 DimSchool 维度表中的部分元组

查询 DimGender 维度表中数据填充是否成功，SQL 查询代码如下：

```
SELECT [GenderID]
      ,[Gender]
FROM [EMIS_SSAS].[dbo].[DimGender]
```

DimGender 维度表数据成功填充，执行结果如图 3-21 所示。

Results		Messages
GenderID	Gender	
1	男	
2	女	

图 3-21 DimGender 维度表中的元组

查询 DimEthnicGroup 维度表中数据填充是否成功，SQL 查询代码如下：

```
SELECT [EthnicGroupID]
      ,[Name]
FROM [EMIS_SSAS].[dbo].[DimEthnicGroup]
```

DimEthnicGroup 维度表数据成功填充，执行结果如图 3-22 所示。

Results		Messages
	EthnicGroupID	Name
1	01	汉族
2	03	回族
3	04	藏族
4	06	苗族
5	07	彝族

图 3-22 DimEthnicGroup 维度表中的部分元组

3.5 多维数据集

3.3 中利用 SQL Server Business Intelligence Development Studio 工具创建数据集市 EMISDM 的方法中已经介绍了多维数据集的创建方法，这里对多维数据集的操作

取决于所选用的创建数据集市的方法。

如果是采用手动方式创建数据集市，就需要新建 SSAS 项目，创建新的多维数据集，其操作步骤可参考 3.3 中利用 SSAS 分析服务创建多维数据集的过程。但在创建多维数据集之前，我们需要先创建数据源和数据源视图，数据源是用来连接目标数据库的，数据源视图用来连接数据库中对应的数据表和视图。

如果是利用 SQL Server Business Intelligence Development Studio 工具自动创建数据集市，则不需要重新创建新的多维数据集，因为前面已经设计好了多维分析的度量值和维度，为多角度分析在校学生人数的数据集市打下了基础。但是，这里我们需要以已成功填充数据的数据集市作为数据源，重新生成数据源视图。重新生成的数据源视图如图 3-23 所示，呈雪花型结构模式，原因是 DimSchool 维度表是 DimClass 维度表的子维表，它们通过 DimSchool 维度表中的主键 SchoolID 与 DimClass 维度表中的外键 SchoolID 建立了关系。

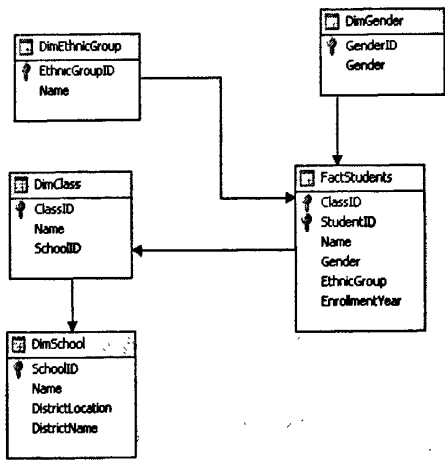


图 3-23 在校学生人数数据源视图

在对多维数据集进行部署后，如图 3-24 显示了在校学生人数多维数据集中的事实表和维度表，可以清晰的辨识带有黄色边框的是事实表，而带有蓝色边框的则是维度表。部署完成后，就可以浏览多维数据集中的实际数据了。方法如下：从元数据窗中拖动有关内容到右边数据显示区即可形成一个初步的报表。如图 3-25 所示，就是从元数据窗口中拖动事实表中的在校学生人数度量值、民族维度表和性别维度表到数据显示区形成的报表。

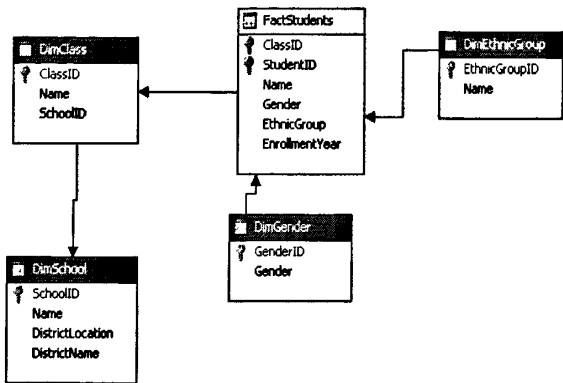


图 3-24 在校学生人数多维数据集结构

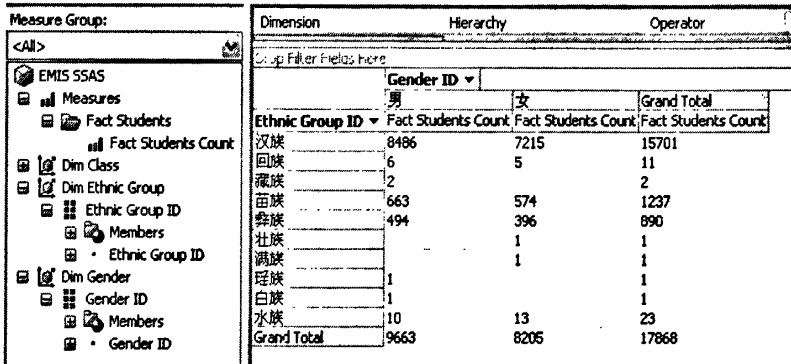


图 3-25 浏览在校学生人数多维数据集界面

3.6 数据呈现

成功创建多维数据集之后，最重要的环节就是选择适合用户的操作平台以呈现多维数据集中分析所得的数据或视图，也是商业智能系统的最终目标。前面已经讲到可以在 SSAS 项目的窗口中直接浏览多维数据集中的数据，但这主要是教育管理决策支持系统开发人员使用的方法，并不适合最终用户。这里我们将选择用户熟悉的 Microsoft Office Excel 2007 实现数据的可视化，同时，也是最简单、最容易操作的方法。

在设置并成功部署了多维数据集之后，打开 Excel 2007，将其连接到已创建的多维数据集，在“导入数据”对话框中选择“数据透视图和数据透视表”的显示方式，可以使呈现的数据更容易、更清楚的理解。如图 3-26 所示，是在校学生人数多维数据集在 Excel 2007 中打开的显示界面，左边呈现的是“数据透视表”，中间呈

现的是“数据透视图”，右边呈现的则是和多维数据集中事实表、维度表及字段一致的内容，我们可以通过拖动事实表到下方对应的“数值”区，因为事实表本身就是用来存储实际数据的，拖动各维度的相关字段到“图例字段”和“轴字段”区，以呈现包含不同内容的数据透视表和数据透视图，如图 3-27 所示，方便最终用户查看。

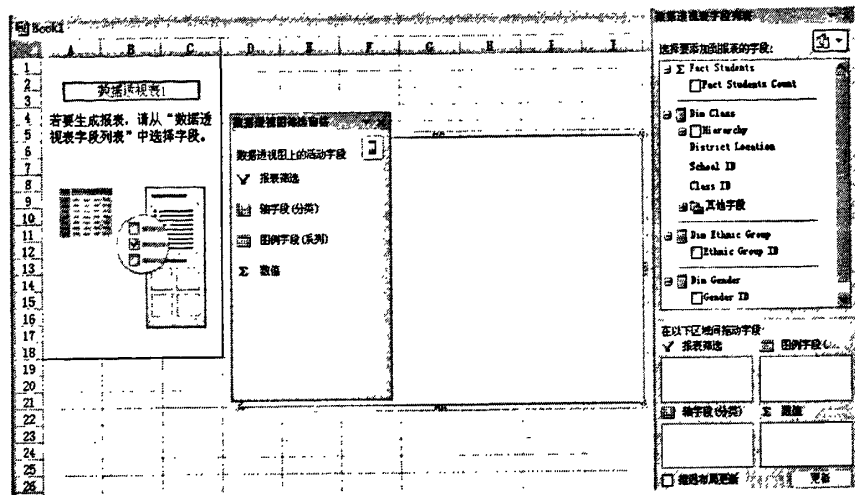


图 3-26 Excel 2007 呈现数据主界面

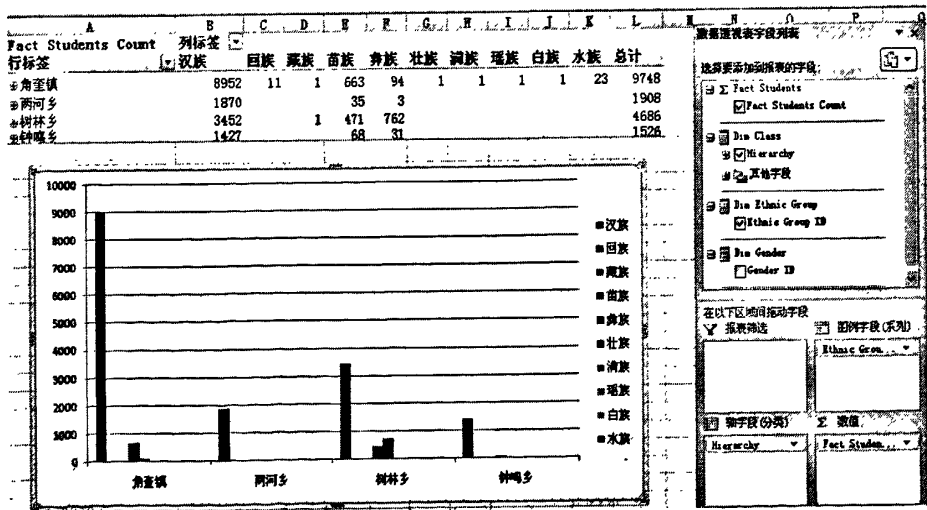


图 3-27 呈现结果示意图

第四章 案例结果与分析

以在校学生人数为案例,对教育管理指标分析的技术模型进行了实施过程分析,并通过用户熟悉的操作平台,最终实现了业务系统中在校学生人数图形化的呈现方式。本案例在对业务系统数据分析的基础上,通过 The Business Intelligence Development Studio 工具实现了学生数据的提取、转换、加载、填充、集成、分析及可视化,这里的“可视化”即数据的前端呈现。通过直观、形象的图形方式呈现的数据,可以方便教育管理者清楚的了解学生的分布状态,有助于教育管理者更好的制定教育决策。本案例中我们可以从民族、性别和学校三个维度出发对在校学生人数进行分析,即可以从单一维度分析,也可以从多个不同纬度分析。

4.1 单一维度学生分布

本案例是从民族、性别和学校三个维度对在校学生人数多维数据集进行的设计,即单一维度可以是民族维度、性别维度或者学校维度中的任何一个,这里以性别维度为例,呈现在在校学生人数的数据透视表和数据透视图,如图 4-1 所示。

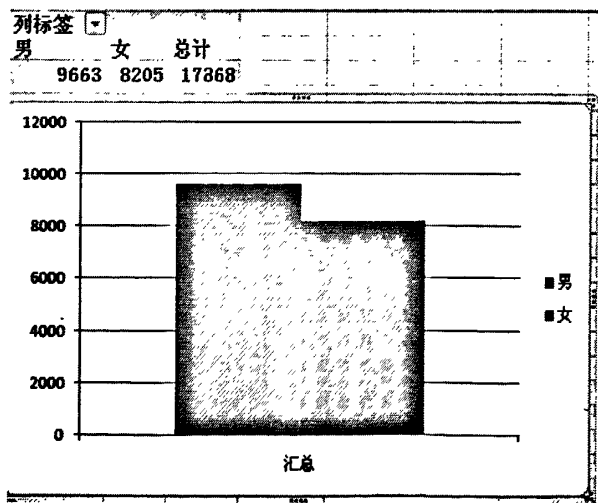


图 4-1 在校学生人数性别维度数据呈现图

如图所示呈现的是不同性别的学生分析柱状图,也可以根据需要更改图形的类别,如饼状图、折线图等,以呈现最能反映实际情况的图形。

从上面数据呈现图中,可以清楚的看到男、女生在在校学生人数中的分布情况,方便了解男、女学生的构成比例,给教育管理者提供了真实、直观的数据。通过和本地人口中实际符合九年义务教育条件的比较,可以从分析结果中得出对应的比例是否基本一致或接近。按理说,两者的比例应该是基本一致的,但如果没有实现这一指标,则需要寻找原因。

同样,可以从民族和学校维度分析学生分布,通过分析各学校在校学生人数和实际学生所在地情况,可以清楚的了解学校的分布是否符合学生实际的分布,应根据学生的实际分布情况,适应城乡发展需要,合理规划学校布局,办好必要的教学点,以方便学生就近入学。通过分析不同民族学生的分布,有助于全面提高少数民族地区的教育发展水平,关注各民族学生的上学情况。

4.2 多维度学生分布

多维度分析学生分布情况,可以是民族、性别和学校三个维度的任意结合。例如,可是民族和学校维度的组合,也可以是性别和学校维度的组合,还可以是民族、性别和学校三个维度的组合等。这里以性别和学校两个维度的组合为例,呈现在校学生人数的数据透视表和数据透视图,通过设置不同的“图例字段”和“轴字段”,呈现的学生分布图是不同的,如图 4-2 所示是以性别维度为图例字段,以学校维度为轴字段设置的结果,如图 4-3 所示是以学校维度为图例字段,以性别维度为轴字段设置的结果,用户可以根据实际需要进行设置,但两者所呈现的结果实际意义是相同的。

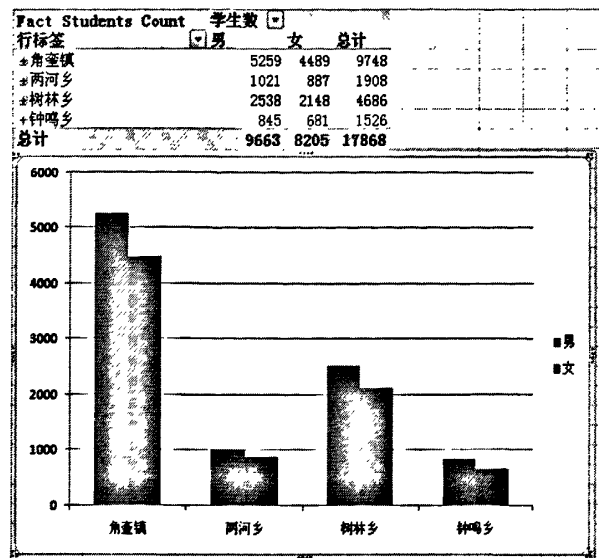


图 4-2 以学校维度为轴字段的数据呈现图

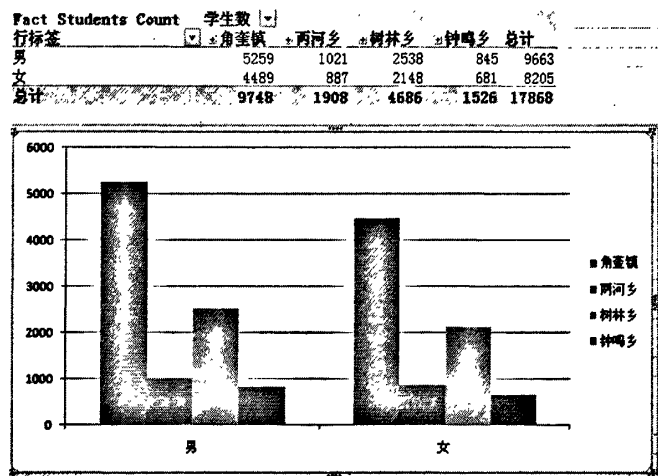


图 4-3 以性别维度为轴字段的数据呈现图

通过上面两个数据呈现图，可以清楚的了解学校的所在地主要包括：角奎镇、两河乡、树林乡和钟鸣乡，其中，对男、女生的分析有着和单一的性别维度分析一样的意义。通过分析得出的结果，可以清楚地知道每个学校所在地包括的学校有哪些，如图 4-4 所示，是两河乡的所有学校及每个学校男、女生学生数。

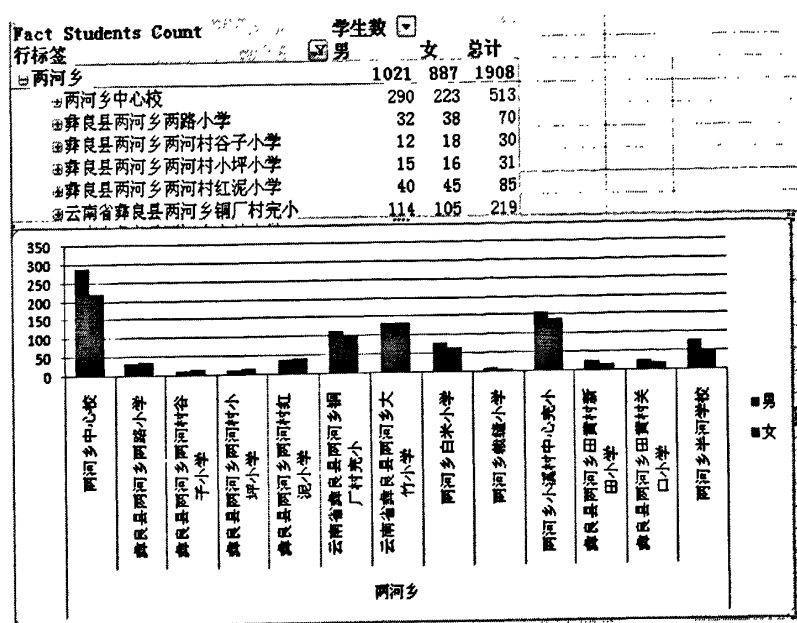


图 4-4 两河乡学生分布图

通过分析每个学校男、女生分布比例，我们发现相对好的学校一般是男生所占比例较高，如中心校，而一般村教学点学校是女生所占比例较高，这就说明了当地对男、女生教育的重视程度存在一定的差别，应该引起教育管理者的重视。

4.3 教育管理指标分析结果对教育的意义

教育管理指标分析结果对教育的意义，主要体现在：提供了支持教育管理真实、科学、直观的数据，使得教育管理者在进行管理的过程中从被动转变为主动，提高了教育管理的效率。

随着信息技术的不断发展，使得越来越多的信息管理系统用在了教育管理中，加快了教育管理信息化的进程，但同时，教育管理信息系统中积累的大量教育数据，也给教育管理者带来了诸多的不便，以至于不能充分发掘教育数据中所隐藏着的巨大的、有价值的信息。而以教育管理者关心的不同层面制定的教育管理指标，正是为了教育管理者能更好、更方便地利用教育数据中的信息进行教育决策地制定，以加强教育管理的质

量。本文通过 BI 技术实现的教育管理指标的分析，就是将教育管理者关心的实际教育问题进行数量化的转换，并在其基础上通过对数据的集成、分析和呈现，提供真

实的、科学的数据支持，以方便管理者能够清楚的了解各种相关教育元素的实际情况，有助于教育管理者更好的制定教育决策。

同时，对教育管理指标的分析，也使得教育管理者在进行管理的时候变的更主动，而不是再被动地去调查、分析教育实施的情况，然后制定教育决策。例如，某学校要求乡镇教育主管部门为其增加教师数量，按照平时的做法，首先，教育部门的负责人要去学校调查学校实际的教师数目，以及与教师有关的其他教育元素的情况，再充分研究的基础上，才能确定是否需要为其增加教师，整个过程显得比较被动。而借助于 BI 技术实现的教育管理指标的分析，教育部门的负责人可以直接通过该校上报的教育数据，方便、快速地查看该校教师与学生的比例或者是教师的工作量，都可以从中得知答案。如果是分析该校在校学生人数，结果发现学生的数量是逐年增加的，师生比超过了相关教育指标规定的标准，则可以考虑为其增加教师，这种方式可以使得教育管理者更方便、快速的制定教育决策。综上所述，通过对教育管理指标的分析可以科学、有效地支持教育管理，为教育决策的制定提供更直观、真实的数据支持，有利于教育的发展。

第五章 讨论

5.1 教育管理指标与地方教育主管部门教育规划的制定

百年大计，教育为本。在我国，党和国家历来高度重视教育的发展，制定了一系列的教育发展规划。2010年，根据党的十七大关于“优先发展教育，建设人力资源强国”的战略部署，又制定了《国家中长期教育改革和发展规划纲要（2010-2020年）》，以下简称《教育规划纲要》。前面，我们介绍过教育管理指标，就是各级教育主管部门通过收集和分析各种相关的教育数据，在将其量化的基础上呈现教育管理结果，方便教育管理者了解教育系统的特征，评估教育发展的进程，制定科学的教育决策。教育管理指标的制定和具体的教育管理数据的量化结果是一对相互作用的两个不同的表示方式，教育管理者依靠教育管理指标收集相关的教育数据将其量化，而得到的具体的教育管理结果又可以反过来指导教育管理指标的制定。那么，《教育规划纲要》中，体现出了哪些方面的教育管理指标呢？

《教育规划纲要》中教育管理指标主要体现在¹：（1）入学率：到2020年，基本实现教育现代化，实现更高水平的普及教育，要基本普及学前教育，巩固提高九年义务教育水平，普及高中阶段教育，毛入学率达到90%，进一步提高高等教育大众化水平，毛入学率达到40%，扫除青壮年文盲；（2）九年义务教育在校学生巩固率：要尽量确保每一位适龄儿童都能够接受九年义务教育，各学校要在加大硬件建设的基础上，高度重视控流保生工作，到2020年，要使九年义务教育在校学生巩固率控制在95%左右；（3）学生分布：根据适龄儿童的实际分布情况，适应城乡发展需要，合理规划学校布局，办好必要的教学点，方便学生就近入学；同时，还要全面提高少数民族地区的教育发展水平，关注各民族学生的上学情况；（4）留守儿童入学率：建立健全的农村留守儿童关爱服务体系和动态监测机制，坚强寄宿学校建设，优先满足留守儿童住宿的需求；（5）专任教师工作量：合理安排每位教师的工作量。

¹ 国家中长期教育改革和发展规划纲要（2010-2020年）

[EB/OL]. http://www.gov.cn/jrzq/2010-07/29/content_1667143.htm

作量,教师要严格执行课程方案,不得增加课时;(6)升学率:加快普及高中阶段教育的需求,到2020年,要满足各阶段毕业生接受更高阶段教育的需求,不能使任何一个学生因高中阶段教育的不完全发展而得不到升学;(7)学生异动:注重因材施教,关注每个学生的不同特点和个性差异,发展学生的优质潜能。对于学习困难的学生要建立帮助机制,同时改进优异学生的培养方式,以使每个学生都能最大程度的发挥潜能;(8)教师队伍基本情况:严格教师资质,提升教师素质,维护教师权益,努力打造一支师德高尚、业务精湛、结构合理、充满活力的高素质专业教师队伍,等等。

《教育规划纲要》中还涉及到其他一些方面的教育指标,这里就不一一做详细说明了,具体情况可参考《国家中长期教育改革和发展规划纲要(2010-2020年)》。

《教育规划纲要》中教育管理指标的制定是大方向的,各地方教育主管部门可以根据当地实际教育情况,制定符合实际教育发展的教育管理指标。

5.2 关系代数在数据提取中的作用

在使用 SSIS 集成服务实现数据填充的时候,我们要用到数据源中的一些数据,即按照实际需求分析选择所需要的数据,也就是要对数据源中的数据进行一定程度的提取。具体实现的操作就是在数据流中,如图 5-1 所示,可以看到数据流任务中包含了两个步骤,首先是从数据源中提取数据,然后将数据载入到数据目标中。加载到数据目标中的数据通常是根据用户对数据集合模式的定义,从数据源中提取并对其进行必要的转换后的数据,以最大程度的满足用户的需求。但随着教育信息化程度的日益提高,教育管理信息系统中的数据迅猛增加,导致数据提取的效率和系统的查询性能下降。在数据提取之前,用户需要把从数据字典中读取的数据集合的描述信息转换成数据的查询语言。对于数据的查询要求可通过关系代数操作表达,又由于关系代数操作与 SQL 语句描述之间存在着对应关系,因此可以将对数据的查询转换为关系代数的运算,从而可以利用关系代数等价转换规则生成优化 SQL 的查询计划,以提高数据提取的效率和系统的查询性能¹。

¹ 萨师煊,王珊.数据库系统概论[M].北京:高等教育出版社,2002.

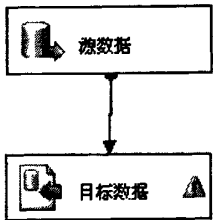


图 5-1 数据流

关系代数最初是由 T.Codd 提出来的，它是一个元组集合，即关系。关系代数作为一种抽象的查询语言，包括了并、差、笛卡尔积、选择和投影等一些基本运算法则，可以用来表达用户提出的查询请求。关系代数等价变换优化算法的基本原理是：把查询问题转变成关系代数表达式，分析得到查询树，从而将全局查询转变为基于片段的查询树，然后利用关系代数等价变换规则的优化算法，对片段上的查询树进行优化，从而最后达到查询优化的目的¹。

在使用关系代数进行数据提取的时候，对于给定的查询问题，我们要注意最大程度的利用关系代数等价变换规则：即尽可能将选择和投影操作下移，提前进行该运算，以减少元组数和关系大小，减少操作次数；把某些选择运算附加在连接运算上，以减少中间结果保存以备后用的时间代价；同一关系中的多个选择和投影同时进行，以避免重复扫描²。总的来说，就是在数据查询和提取的过程中，尽可能减少元组和关系一些不必要的重复操作，减少后操作的操作量和操作的次数，达到优化查询的目的。

5.3 对数据维度的理解

数据组织的存储方式不仅可以是二维的关系表，还可以是以多维的方式对数据进行组织、存储和管理数据库，可以简单地理解为是将数据存放在一个 n ($n=3, 4\cdots$) 维数组中，维表中记录的就是数据中的维度，如利用 Integration Service 创建的多维数据集就是多维度的数据，一个多维数据集最多可以包括 128 个维度。维度是多维数据集的结构特征，是用来描述数据分类的有组织的层级结构的事实数据，用户

¹ 郑勇明, 彭凤梅, 陈越. 基于关系代数等价变换的查询优化处理[J]. 数据库及信息管理.

² 王峥, 王亚平. 关系代数与 SQL 查询优化的研究[J]. 电子设计工程, 2009.

是基于一些描述了相似成员集合的分类和级别进行分析的。维度的含义简单的来说,就是人们观察、分析数据的角度。通常情况下,对数据维度的定义并不是任意的,而是必须要建立在对客观事物充分分析的基础之上,以便于开展深入分析为目标,是一种高层次的类型划分¹。维度是多维数据分析的核心概念之一,在多维数据集的构建中起到了至关重要的作用。

通常我们在分析数据的某个特定的维度时还可以从不同的细节程度作为切入点,而这些维度的不同细节程度就是维度的级别。一个维度往往可以包括多个不同的级别,例如在描述时间维度的时候,可以将时间维度分为年、季、月等不同的级别来描述,那么年、季、月等就是时间维度的级别。另外,我们还可以根据实际需要时间维度的不同级别进行重新定义,又可以分为不同的时间维度级别。维的取值我们称之为维度成员,简称维成员。一般情况下,一个多级别的维的维度成员要在不同维级别上进行取值,然后再将其进行组合。例如,从年、季、月等不同的级别来描述维度,就要分别在年、季、月上各取一个值,然后将其组合在一起就得到了时间的一个维度成员,即“某年某季某月”,有时候不一定要在每个维度级别上都对维度成员取值,可以只在某些特定维度级别上对其进行取值,例如“某年某月”、“某季某月”“某月”等形式的取值,都可以是时间维的维度成员。

多维数据的另一个重要概念是维度内元素的“类”的概念,类是指按照一定的划分标准对维度成员全集的一个划分²。值得注意的是,维度的级别和类是两个不同的概念,维度的级别也就是维度层次,主要是用来对数据进行钻取分析的,以便使用户能够查看不同层次的数据,而维度的分类是对维度取值的划分,目的是为了在不同的类别间进行比较,只有同一层次的维度成员才可以划分为类。以时间维度为例,维度的层次和类的关系图,如图 5-2 所示。

¹ 苏新宁,杨建林,江念南,栗湘.数据仓库和数据挖掘[M].北京:清华大学出版社,2006,67.

² 王丽珍,周丽华,陈红梅,肖清.数据仓库与数据挖掘原理及应用[M].北京:科学出版社,2009,79.

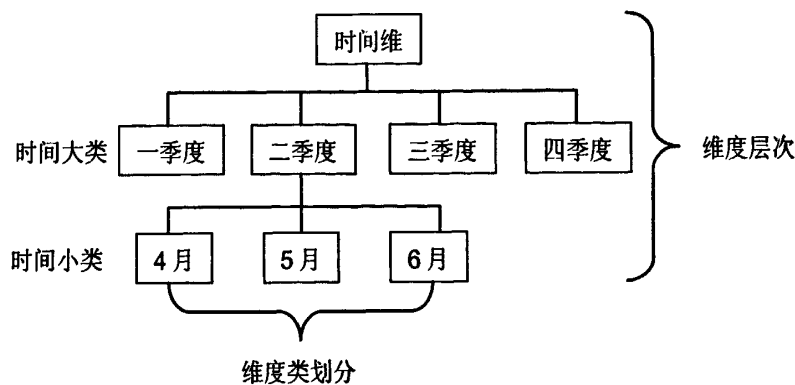


图 5-2 维度的层次和类

5.4 业务系统中数据质量的监测

随着计算机和信息技术的快速发展，当今的数据容量已经达到了万亿字节的水平，过量的数据造成了信息的爆炸，出现的是一种“数据丰富，但知识贫乏”的状态。人们面对如此规模庞大、纷繁复杂的数据却是漫无头绪的，也不便快速的发掘这些数据背后隐藏着的更具有价值意义的决策信息。信息系统的广泛应用，促使人们在获取大量数据的同时，越来越关注数据的质量问题，错误的或者不一致的数据都有可能严重影响数据分析的结果，由此可见，对数据的质量实施管理是至关重要的。

数据质量是一个相对的概念，是质量本身定义的延伸，在不同的时期和不同的应用领域有着不同的定义和评估标准。从数据生产过程中的质量维度看，对数据质量的评估主要涉及到数据的准确性、完整性、一致性以及数据满足用户需求程度等指标。数据挖掘是从大量数据中获取有价值的信息或知识的过程，数据挖掘的结果及质量，在数据挖掘过程中使用的方法与技术同条件下，是与用于挖掘的数据源的质量是密切相关的，高质量的数据挖掘结果依赖于高质量的数据源，所以利用数据挖掘技术可以监测业务系统中数据的质量。我们要特别注意的是，对于数据挖掘，同样的数据在一个挖掘主题下表现良好，而在另一个挖掘主题下可能得不到有意义的结果，也就是说，同样的数据在不同的数据挖掘主题下得到的结果可能是不同的，因此，在实际应用过程中要根据需求分析的结果，选择合适的维度对数据质量进行

监测¹。

数据挖掘是为了满足信息处理新的需求和社会发展的各种需要,而发展起来的一种新的信息分析技术。它能够帮助人们获取大量数据背后隐藏着的特定的数值模式和规律,缩短了数据和信息之间的巨大鸿沟,从而为诸多领域提供了更有价值的、科学的信息,往往被用来预测和支持决策。数据挖掘就是发现数据中的知识,由于发现的方法、应用的领域等的不同,数据挖掘被看作是一门多学科综合的产物,至今,数据挖掘仍没有一个统一的定义。

关于数据挖掘有以下不同的定义:

数据挖掘是利用各种分析工具从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中,提取隐含在其中的、有效的、新颖的、潜在有用的并最终可被理解的信息和知识的过程²。

数据挖掘是从存放在数据库、数据仓库或其他信息库中的大量的数据中获取有效的、新颖的、潜在有用的、最终可被理解的模式的非平凡过程³。

简单的说,数据挖掘就是从大量的数据中提取或挖掘知识⁴。

现代社会是一个被数据充斥着的社会,随着各种信息管理系统中数据的急剧增加、新应用的不断出现以及各种应用之间越来越频繁的相互整合,人们对数据质量的问题越来越重视。数据挖掘技术的应用对各种业务系统中数据质量的监测具有重要的意义。

¹ 王晓华. 电信数据挖掘的数据质量评估技术研究[D]. 浙江: 浙江大学, 2010.

² 谷宏群. 数据挖掘中可视化方法研究 [D]. 武汉: 武汉大学, 2005.

³ 百度百科. 数据挖掘[EB/OL]. <http://baike.baidu.com/view/7893.htm#1>, 2010.

⁴ Jiawei Han, Micheline Kamber. Data Mining: Concepts and Techniques[M]. Beijing: Higher Education Press, 2001.

第六章 总结与展望

6.1 论文总结

本论文主要是在中英西南教育管理系统分析、研究的基础上,针对教育管理者关心的教育问题,制定了教育管理指标,然后再利用商业智能技术实现反映了教育管理指标的教育数据的数据表和数据图的呈现。教育管理者通过直观、形象的图形模式呈现的教育数据,客观、快速地了解当地教育的整体实施情况,有助于教育管理者及时、方便地做出有效的决策,以便提高教育管理的效率。

本文首先在分析业务系统教育数据的基础上,找出能够反映教育管理指标的数据,即在存储了大量教育数据的数据库中提取符合需求分析的数据表及数据,通过对业务系统、教育指标探讨的基础上,提出了如何快速发掘隐藏在数据背后信息的问题。在分析问题的基础上,提出了实现教育管理指标分析的 BI 技术模型:业务系统分析—建立数据集市—填充数据—创建多维数据集—数据呈现方式。在对这一技术模型构成要素过程分析、BI 概述及 BI 技术、工具论述的基础上,总结归纳出了本文所使用的 BI 解决方案,并结合具体的案例对实施过程进行了详细分析。最终借助于 Excel 2007 实现数据表或数据图的方式呈现教育数据,以便教育管理者能够了解教育管理指标的基本情况,同时方便快速、有效地制定教育决策。总之,就是在对教育管理指标分析的基础上,针对教育管理指标所涉及到的数据,通过一系列的操作过程,最终借助于形象、直观的数据呈现方式,方便教育管理者了解各指标的实际情况及发展趋势,为决策地制定提供科学的信息支持。

6.2 论文展望

本文主要是为了实现教育管理决策支持系统的开发,其中选择的研究角度是教育管理指标,实际教育管理中涉及到的指标是很多的且各指标间存在着一定的联系。本文只对在校学生人数这一教育管理指标进行了分析、研究,相对来说是较简单的,

没有涉及到太多的事实表、维度表及属性，实施起来比较容易，而对于整个教育管理系统决策支持系统的开发来说，涉及的事实表、维度表及属性会有很大数量的提升，并且各表之间的关系也会变得相对复杂，接下来会加深对其的研究，以便能呈现完整的数据，帮助教育决策者提高教育管理的效率。

针对本文中的案例研究，最终的用户应该是教育管理者，既然是教育管理，就需要进行实地的实践验证，并对这种教育管理方式进行实验研究，找出存在的问题并加以分析，找出解决问题的方法，以便进一步优化之，从而为教育管理提供最大的帮助。

致谢

在论文完成之际，首先要向尊敬的导师季隽老师表达我最真挚的谢意！本文的研究工作无论从论文的选题、撰写完善，还是研究方案的确定、案例的设计和开发，都倾注了导师非常多的精力，给予悉心的指导。在我三年的研究生学习中，季老师严谨求实的治学精神、扎实的技术功底、敏锐的洞察能力、一丝不苟的工作作风、忘我的工作热情和平易近人的高尚品质，都深深的影响了我，使我受益匪浅。三年来我所取得的每一点成绩都凝聚着季老师的心血，感谢他在学习、生活等各方面给予我的无私关怀和帮助。同时，也要感谢黎加厚老师，三年的研究生学习生涯中，他高深的学术造诣、精益求精的治学精神、丰富渊博的知识和朴实无华的人格魅力都给我以深刻的教育和熏陶，激励着我在前进的道路上奋力向前。

感谢研究生期间，曾给予悉心教育和帮助的叶海松老师、鲍贤清老师、傅英老师、孟琦老师……以及其他教育技术系的老师们，表示最崇高的敬意和最衷心的感谢！

感谢院办的各位老师和数理学院办公室的各位老师，给予我研究生学习和生活各方面的帮助和支持！

感谢所有我的同学们，感谢他们在我学习上给予的支持和生活上给予的莫大的关怀，他们让我的研究生生活变得丰富多彩。同学们之间彼此的交流与互相鼓励，让我感受到了大家庭的温暖，与他们的交流给我很多启示，感谢你们一直陪伴在我的左右，给我温暖和前进的动力！感谢所有朋友对我的帮助和关怀！

感谢我的室友，和我一起见证了欢声笑语、青春年华的岁月，感谢你们的帮助和支持！

最后，特别感谢我的父母和家人对我学业的支持，是他们默默的付出和无私的帮助才使我得以顺利的完成学业。他们为我实现梦想提供了坚强的后盾，使我能够全身心的投入到学习中，当我遇到困难的时候，是他们鼓励我，让我有了战胜困难的勇气和信心。

论文在撰写的过程中参阅了许多专家和学者的研究成果，在此，谨对他们表示

深深的感谢！

再次感谢所有给予我关心、支持和帮助的老师、同学、家人和朋友们！谢谢你们！

参考文献

- [1]Lynn Langit, Kevin S. Goff, Davide Mauri 著, 张猛, 杨越等 (译). SQL Server 2008 商业智能完美解决方案[M]. 北京: 人民邮电出版社, 2010, (8).
- [2]包涵宗. 高校计算机基础课程评价决策支持系统研究与实现[D]. 湖南: 中南大学, 2008, (11).
- [3]曹建军, 刁兴春, 汪挺, 袁霞, 张潇毅. 数据质量控制研究中若干基本问题[J]. 微计算机信息, 2010.
- [4]常达俊. 基于数据仓库和 OLAP 的决策技术研究[D]. 吉林: 长春理工大学, 2009, (3).
- [5]常鑫. 数据仓库中基于模糊数据的数据立方体设计[J]. 计算机时代, 2010, (11).
- [6]陈婧. 基于商业智能技术优化 ERP 系统的应用研究[D]. 上海: 华东师范大学, 2009, (4).
- [7]成媛. 西部地区教育现代化指标体系的构建[J]. 北方民族大学学报, 2010, (6).
- [8]程荣赟. 基于 SSAS 的网站访问日志分析和 BI 前端呈现[D]. 四川: 西南财经大学, 2008, (12).
- [9]迟殿委. 基于数据挖掘的决策支持系统的研究与实现[D]. 江西: 南昌大学, 2008, (12).
- [10]盖欧. 企业商业智能应用模式的研究[D]. 陕西: 西安电子科技大学, 2009, (1).
- [11]郭君芳. 教育信息数据的可视化研究[D]. 上海: 上海师范大学, 2009, (4).
- [12]哈笑. 基于 Web 的商业智能决策支持系统研究[D]. 辽宁: 东北财经大学, 2007, (12).
- [13]胡天状. 数据挖掘技术在教育决策支持系统中的应用[D]. 浙江: 浙江师范大学, 2005, (5).
- [14]吉顺国. 基于数据仓库的学籍管理系统设计与实现[D]. 江苏: 苏州大学, 2009, (10).
- [15]靖伟峰. 基于电子档案袋的决策支持系统设计与实现[D]. 湖北: 华中科技大学, 2008, (6).
- [16]李东, 梁定澎. 决策支持系统与商务智能[M]. 北京: 中国人民大学出版社, 2010.
- [17]李建军. 试论高校内部决策系统及其优化[J]. 高等教育研究, 2009, (8).
- [18]李青. 基于商业智能的广电运营数据资源管理系统[D]. 天津: 天津大学, 2009, (12).

- [19]刘建银,安宝生.教育指标理论研究的几个基本问题[J].中国教育学刊,2007,(9).
- [20]罗建.可视化数据挖掘方法的研究与实现[D].成都:电子科技大学,2009.
- [21]牛潇楠.基于 OLAP 的工商业务分析系统[D].广州:华南理工大学,2009,(12).
- [22]任广新.商业智能及其在商业银行中的应用[D].吉林:吉林大学,2006,(9).
- [23]萨师煊,王珊.数据库系统概论[M].北京:高等教育出版社,2002.
- [24]苏新宁,杨建林,江念南,栗湘.数据仓库和数据挖掘[M].北京:清华大学出版社,2006.
- [25]孙桂丽.区域教育现代化指标体系研究[D].江苏:苏州大学,2009,(4).
- [26]孙继勇.智能化商业信息管理技术的研究和系统开发[D].四川:重庆大学,2002,(5).
- [27]孙杰.基于数据仓库的教学管理系统应用研究[D].安徽:合肥工业大学,2007,(10).
- [28]王军.数据仓库技术在商业智能系统中的应用[D].上海:华东师范大学,2007,(10).
- [29]王丽珍,周丽华,陈红梅,肖清.数据仓库与数据挖掘原理及应用[M].北京:科学出版社,2009.
- [30]王倩倩.电信商业智能系统 OLAP 建模与应用[D].北京:首都经济贸易大学,2009,(3).
- [31]王晓华.电信数据挖掘的数据质量评估技术研究[D].浙江:浙江大学,2010,(1).
- [32]王媛媛.B/S 模式下数据可视化研究及其在商业智能中的应用[D].河北:河北工业大学,2007,(1).
- [33]王晟.基于数据仓库的教务决策支持系统研究与实现[D].福建:厦门大学,2007,(1).
- [34]王哲.高校辅助信息决策支持系统的研究与实践[D].甘肃:兰州大学,2009,(11).
- [35]王峥,王亚平.关系代数与 SQL 查询优化的研究[J].电子设计工程,2009,(8).
- [36]吴长泽.基于 OLAP 的商业智能系统研究及应用[D].四川:重庆大学,2004,(10).
- [37]吴文昭.甘肃省教育管理信息决策支持系统[D].甘肃:兰州理工大学,2006,(5).
- [38]伍永锋.商业智能及其技术[D].贵州:贵州大学,2008,(4).
- [39]徐玲.国际教育指标体系的分析与思考[J].教育科学,2004,(4).
- [40]闫琴.商务智能决策支持系统研究[D].辽宁:东北财经大学,2006,(12).
- [41]杨思清.数据仓库在教育评价系统中的应用研究[D].安徽:安徽理工大学,2007,(6).
- [42]袁桂林,秦玉友.农村教育发展指标研究中的若干基本理论问题[J].教育发展研

- 究,2007,(5).
- [43]袁泉.一种商业智能中的 OLAP 与客户数据挖掘方法研究[D].黑龙江:哈尔滨工业大学,2006,(12).
- [44]詹学朋,王斌会.基于 SQL Server 商业智能的统计决策分析系统[J].统计与决策,2009,(11).
- [45]张云涛,龚翎.商业智能设计、部署与实现[M].北京:电子工业出版社,2004.
- [46]赵斌.SQL Server 2008 应用开发案例解析[M].北京:科学出版社,2009.
- [47]郑炎.基于数据仓库的商业智能系统研究与实践[D].江苏:南京工业大学,2005,(12).
- [48]郑勇明,彭凤梅,陈越.分布式数据库查询优化处理—基于关系代数等价变换的查询优化处理[J].电脑知识与技术,2007,(4).
- [49]钟明.商业智能系统研究与设计[D].四川:重庆大学,2007,(4).
- [50]中英甘肃普及九年义务教育项目领导小组办公室,中英西南基础教育项目办公室.《中英基础教育项目教育管理信息系统使用指南》[M].北京:教育科学出版社出版,2010.
- [51]周艳.浅谈商业智能系统与数据挖掘[J].知识经济,2009,(10).
- [52]Chris Webb,Alberto Ferrari,Marco Russo. Expert Cube Development with Microsoft SQL Server 2008 Analysis Services [M].UK: Packt Publishing Ltd, 2009.
- [53]Brian Larson. DELIVERING BUSINESS INTELLIGENCE with Microsoft SQL Server 2008 [M].USA: The McGraw-Hill Companies,2009.
- [54]Cognos Enterprise Business Intelligence for E-business [N].Technical Discussion Paper, 2000.
- [55]Endsley, Krista. Business Intelligence: Turning Knowledge into Power[J].School Business Affairs,2009,75(26-28).
- [56]Erik Veerman,Teo Lachev.Microsoft SQL Server 2008-Business Intelligence Development and Maintenance [M].USA: Solid Quality Mentors,2009.
- [57]Felix,Kathie.Take ACTION With BUSINESS INTELLIGENCE[J]. Econtent, 2009,32 (22-26).
- [58]Hongjun Lu. Seamless Integration of Data Mining with DBMS and Applications

- [M].Lecture Notes in Computer Science, 2003, (7):78-136.
- [59]IBM. Business Intelligence Certification Guide [M], IBM, 2000.
- [60]Jeffrey Pepper. Building a Data Warehouse: With Examples in SQL Server [M].USA: Vincent Rainardi, 2008.
- [61]Jiawei Han,Micheline Kamber. Data Mining: Concepts and Techniques[M].Beijing: Higher Education Press,2001.
- [62]Microsoft SQL Server 2008-Business Intelligence Development and Maintenance [M].USA: Microsoft Press, 2009.
- [63]Mike Perkowitz, OrenEtzionl. Towards Adaptive WebSites-Conceptual Framework and Case Stud [J]. Artificial Intelligence,2000,118.
- [64]P. Holt Wilson, Eric N. Wiebe.INNOVATIVE DATA COLLECTION TECHNOLOGIES FOR DESIGN EXPERIMENTS[R]. Conference Papers -- Psychology of Mathematics & Education of North America, 2007.
- [65]R.QuinlanSimplifying decision trees [J].International Journal of Man—Machine Studies. 1987.
- [66]Shao Fengjing,An Outlier-analysis Algorithm Based on the Reduction of Boundary Cells Influence[J].Proc.Of the eighth Joint International Computer Conference,2002.
- [67]Sridhar Ramaswamy,Rajeev Rastogi,Kyuseok Shim.Efficient Algorithms for Mining Outliers from Large Data Sets.SIGMOD Conference,2000.
- [68]Tony Bain,Mike Benkovich.Professional SQL Server 2000 Data Warehousing with Analysis Services [EB].<http://www.amazon.com>.
- [69]William H.Inmon.Building the Datahouse,Fourth Edition[M].北京:机械工业出版社,2007.
- [70]ZhaoHuiTang,JamieMadennan.Data Mining With SQLSERVER 2005 [M].USA:Wiley,2005.

攻读硕士学位期间取得的研究成果

发表论文:

[1]王璐璐. 社会计算及其在教育教学中的应用[J]. 软件导刊, 2010, (09):187-188.

参与项目:

- 2009.3—2009.4 参与上海师范大学团委-《爱心学校》网站的设计与制作
- 2009.4—2010.10 参与中英基础教育项目-《教育管理信息系统》，负责系统的测试、培训教材的编写工作等
- 2009.9—2010.9 参与上海市新农村教师专业发展培训项目课程设计与开发

学位论文独创性声明

本论文是我个人在导师指导下进行的研究工作及取得的研究成果。论文中除了特别加以标注和致谢的地方外,不包含其他人或机构已经发表或撰写过的研究成果。其他同志对本研究的启发和所做的贡献均已在论文中做了明确的声明并表示了谢意。

论文作者签名: 王璐璐

日期: 2011年5月22日

学位论文知识产权属声明

本人完全了解上海师范大学有关保留、使用学位论文的规定,即:学校有权保留送交论文的复印件,允许论文被查阅和借阅;学校可以公布论文的全部或部分内容,可以采用影印、缩印或其它手段保存论文。保密的论文在解密后遵守此规定。

论文作者签名: 王璐璐

日期: 2011年5月22日

导师签名: 叶海松

日期: 2011年5月22日

本论文经答辩委员会全体委员审查，确认符合上海
师范大学硕（博）士学位论文质量要求。

答辩委员会签名：

章伟民

主席（工作单位、职称）：

华东师范大学教授

委员：

孟琦

黄美霞

导师：

李隽

叶海松

