

## 摘 要

随着中国电信重组及全面开放 3G 业务，电信行业的竞争日益加剧。中国联通某电信企业利用数据仓库、联机分析等各种技术来设计和实现企业的经营分析系统，从业务支撑系统入手提取出有用信息，统一对信息的加工处理，为管理人员提供准确、科学、及时的决策依据。本论文主要有以下几方面内容：

首先，从理论上介绍与经营分析系统相关的数据仓库和数据分析技术，这些是构建经营分析系统的基础。

其次，介绍了经营分析系统的整体设计情况，包括项目的需求确认和系统建设的目标、总体思路和设计原则，概念模型、分析模型、物理模型的设计以及实现该系统的软件体系结构和实现模块。

在详细设计阶段，论文从数据获取子系统、经营分析子系统、应用层子系统三个层面对功能模块设计和实现进行了介绍。对数据获取子系统的介绍主要包括了 ETL 的设计要求、模块处理流程和模块详细设计；经营分析子系统部分分析主题的功能设计是论文的重点内容，这里详细论述了用户数分析、集团客户分析、个人客户分析、竞争分析、客户流失分析、收入分析等分析模型的设计和实现；本部分还对应用层子系统的功能设计和实现进行了说明。

最后，说明了经营分析系统的部署实现和部分运行界面测试情况。

该系统对现有的运营数据进行了综合利用，加强对运营过程中产生的信息数据的分析和利用，为公司决策提供了依据，加大了对重点业务、重点客户的动态跟踪，加快了企业市场反应速度，提升竞争力。

**关键词：**电信企业；经营分析系统；数据

# Designing and Implement of the Telecom Business Management Analysis System

## Abstract

With the recombination of Chinese telecom companies and the comprehensive opening of 3G business in China, the competition in telecom industry is aggravating day by day. Based on DW, OLAP and other technology, China Unicom Corporation Limited builds the Telecom Industry Management Analysis System, which extracts useful information and unifies the processing procedures so as to offer the accurate, scientific and timely decision support for marketing management. This thesis is focuses on as follows:

Firstly, a theoretical introduction to the DW and data analysis technology which is the basis of the Management Analysis System.

Secondly, an explanation of the overall design of the Management Analysis System, including confirmation of project demands and system-building objectives, the overall concept and underlining principles, conceptual models, analytical models, physical models, and the software architecture and implementation module for system implementation.

In the detailed design stage, it introduces the design and implementation of function modules in three levels: the data acquisition subsystem, the business analysis subsystem, and the application-layer subsystem. The introduction to data acquisition subsystem includes the ETL design requirements, module processing procedures and modules detailed design; the design of analysis theme function in business analysis subsystem which is the key to my these are described in details, including the design and implementation of analytical model for users total analysis, group customer analysis, personal customer analysis, competitive analysis, customer churn analysis, revenue analysis; the design and implementation of application-layer subsystem are described as well.

Finally, a situation description of the implementation and partial interface test of Management Analysis System.

The system utilizes the existing operation data integratedly, enhances the analysis and application of information generated in operation process, provides the basis for corporate decision-making, increases the dynamic tracking of key businesses and key customers, and speed up the enterprise market responses, thus upgrade their competitiveness.

**Key Words:** Telecom Companies; Management Analysis System; Data

---

## 大连理工大学学位论文独创性声明

作者郑重声明：所呈交的学位论文，是本人在导师的指导下进行研究工作所取得的成果。尽我所知，除文中已经注明引用内容和致谢的地方外，本论文不包含其他个人或集体已经发表的研究成果，也不包含其他已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的贡献均已在论文中做了明确的说明并表示了谢意。

若有不实之处，本人愿意承担相关法律责任。

学位论文题目：电信企业运营分析系统的设计与实现

作者签名：蔡均 日期：2009年12月19日

## 大连理工大学学位论文版权使用授权书

本人完全了解学校有关学位论文知识产权的规定，在校攻读学位期间论文工作的知识产权属于大连理工大学，允许论文被查阅和借阅。学校有权保留论文并向国家有关部门或机构送交论文的复印件和电子版，可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印、或扫描等复制手段保存和汇编本学位论文。

学位论文题目：电信企业经营分析系统的设计与实现  
作者签名：蔡均 日期：2009年12月19日  
导师签名：王爱民 日期：2009年12月19日

# 1 绪论

## 1.1 论文研究意义

随着中国电信重组及全面开放 3G 业务<sup>[1]</sup>, 新的竞争格局的逐渐形成, 中国的电信服务迈向了一个新的发展阶段, 电信市场的竞争日益激烈。在外部经营环境上, 客户对电信企业服务质量要求日益提高, 要求各运营商推出各种新的套餐和增值业务; 内部环境上, 电信重组和内部机构调整, 中国目前存在中国移动、中国联通、中国电信三家全业务电信运营商<sup>[2]</sup>, 如何保有并吸引客户, 充分发掘客户的消费潜力成为运营商广泛关注的焦点; 所有这些因素使得电信企业经营所面临的问题更加复杂, 关注自己的核心竞争力, 进一步提升其管理水平及运行效率, 并逐步降低运营成本。电信企业必须考虑如何提升自己的经营分析能力, 从简单的价格竞争提升到较高的理性竞争分析层面。

电信企业运营过程中, 纷纷建立起了自己的数据库系统, 由计算机管理代替手工操作, 以此来收集存储、管理业务操作数据, 改善办公环境, 提高操作人员的办公效率<sup>[3]</sup>。这些数据库系统积累了大量运营方面的数据, 电信企业通过经营分析系统的建设, 来整合企业各方面数据资源<sup>[4]</sup>, 实现分析客户消费的行为特征和业务分析、预测等多方面的信息, 并在此基础上制定针对性的市场销售计划, 从而以更好的服务保留有利润的客户, 用比竞争运营商更低的成本争取到新客户, 扩大市场份额, 降低营运成本和风险。

本电信企业现有的业务系统、计费系统、支撑系统等各生产系统之间存在标准不统一且各种数据源分散的特点, 无法对经营分析提供有效的支撑。企业需要新的技术来弥补原有数据库系统的不足, 需要把已经广泛收集到的数据集成到数据库中, 以从业务数据中提取有用的信息, 帮助他们在业务管理和发展上做出及时正确的判断, 数据仓库技术应运而生。为了把这些孤立的、分散的业务数据有效地关联综合在一起并合理利用, 以提高企业的经营分析决策能力, 中国联通某电信企业经过多年的运营总结, 通过对一些经营分析主题进行明确, 然后对这些分析主题加以建模; 为了实现一系列分析模型, 然后对交换机原始计费文件进行提取、转化, 从十六进制代码中解读出经营分析系统需要的信息字段并从客户系统、帐务系统等系统中提取数据, 并对这些数据清洗、综合, 使这些具有一致性的数据装载进数据仓库, 最后按照经营分析主题的要求通过分析应用程序对这些数据进行分析汇总, 得到分析结果, 通过报表等多维分析工具展现给企业管理人员, 以达到经营分析的目的。

## 1.2 与论文相关的国内外研究综述

### 1.2.1 电信业经营分析系统国内外研究现状

目前电信企业经营分析主要的数据来自于计费系统和网管数据<sup>[5]</sup>。电信系统经过多年运行,存储了大量的历史数据。这些数据中包含着丰富的信息,从这些信息可以了解企业在电信市场经营过程中的现状、竞争对手的经营现状、客户群的分布情况、消费特征、企业经营发展的趋势等,对其进行分析处理可以得到很多有说服力的结论。丰富的企业数据,加上强大的数据处理能力,可以有力地促进企业经营及竞争。

数据仓库技术<sup>[6]</sup>在电信企业经营分析系统构建中得到了许多应用,主要通过对经营中出现的一些情况进行建模、归纳、总结,通过 OLAP、DM 等分析工具对数据仓库中的主题进行联机分析和数据挖掘,对现有经营情况进行分析以及对未来业务进行预测,提供决策依据。美国 AT&T 公司成功利用数据仓库技术在经营分析中就是一个很好的案例。随着通信市场竞争逐渐激烈,AT&T 公司发现自己市场份额有所下降,于是决定利用数据仓库技术加大投入建立更完善的客户档案信息,同时建立预测模型,尽可能准确预测客户流失的概率,及早采取措施,防止客户流失。项目实施后,业务管理部门拿到更及时准确的信息,制定了有效的经营方案,从而提高客户满意度。同时 AT&T 公司根据客户消费规律采取一系列的市场促销活动,保持并发展客户关系,以低成本吸引新客户。利用大量详细的数据和精确的预测模型,业务部门可以准确查出那些可能投向竞争对手的客户,然后采取适当的措施,减少客户流失率,提高市场竞争力<sup>[7]</sup>。

继 AT&T 公司后,世界上一些主要电信运营商陆续采用了基于数据仓库技术的系统:如美国西南贝尔公司、英国 BT、SPRINT 公司、法国电信、澳大利亚电信、加拿大电信等都采用了数据仓库建立智能系统,解决了预防欺诈、客户群体分析等问题,并取得了良好效果。

目前国内电信企业已有大量成熟的数据库业务系统,如计费系统、帐务系统、管理信息系统等。这些系统积累了大量原始业务数据,然而这些系统都是联机事务处理系统,分析处理往往需要通过复杂的 SQL 语句对海量数据进行查询;而且数据库缺乏集中管理,业务数据被存放到分散的异构环境中,同时这些独立的业务系统没有标准的接口,存在数据定义不同、数据冗余,数据值不一致等问题,从而形成一个个信息孤岛。另外各业务系统中数据保存期也不一样,难以加以分析提供经营决策。

2000 年后电信企业经营分析系统基于数据仓库技术在国内电信行业有了一些研究和应用的案例。在中国电信某些地方公司,利用通过数据仓库技术进行了用户话费欺诈

行为的分析。中国移动集团公司从 21 世纪初期开始, 在一些省份进行了经营分析系统的建设工作, 并取得较好成果。2004 年中国移动完成全国范围内的经营分析系统建设工作, 各地市根据当地特点建立了具有不同特色的分析主题。例如: 广东移动经营分析系统利用联机分析技术实现了客户信用度分析、客户流失分析。山东移动经营分析系统建立了营销计划分析、竞争对手分析和与实际业务紧密联系的专题分析, 还利用数据挖掘技术建立了客户重入网、客户流失预测问题的模型。四川移动利用经营分析系统对账务数据、用户特征、呼叫行为数据等进行综合数学分析, 从而对于客户的信用程度、重要程度做出评估, 并分析企业整体客户的分布水平, 发掘潜在客户和重要客户, 找出新的销售机会, 同时制定相应的销售方案, 为企业发展提供决策支持。

中国联通 2004 年也陆续开始了各地市经营分析系统的建设工作, 本电信企业基于数据仓库技术经营分析系统就是中国联通经营分析系统应用的一个项目。

从系统结构看, 电信企业经营分析系统一般包括数据源系统部分、中央数据仓库系统部分、应用系统部分三大部分。电信企业经营分析系统解决方案数据源主要来自电信各业务系统, 包括计费系统、客服系统、营业系统、结算系统、财务系统、网管系统、112 系统、114 系统、网络资源管理系统等。中央数据仓库系统主要解决的问题包括: 建立统一的业务视图, 构建一个高效、线性可扩展的基础平台, 建立一套自动的数据抽取、转换、装载流程, 建立一套符合业务需求、稳健的、有扩展性的数据模型, 建立一套高可靠的安全管理机制。业务应用系统主要功能包括客户管理、客户发展分析、业务量分析、收入分析、营销管理分析、市场竞争分析、服务质量分析、大客户分析、欠费和动态防欺诈行为分析、代理商及渠道分析等。

### 1.2.2 数据仓库技术国内外发展现状

随着计算机应用技术的快速发展, 令各行各业收集数据的能力大力提升, 随之也就带来了“数据爆炸”现象。如何将这些海量数据存储与分析, 令其转换成信息和知识, 辅助决策管理, 成为亟待解决的问题。由此, 数据仓库技术应运而生。

20 世纪 90 年代初期美国信息工程学家 William Inmon 博士在其里程碑式的著作《Building the Datahouse》中提出了“数据仓库”的概念<sup>[8]</sup>, 他认为: “一个数据仓库通常是一个面向主题的、集成的及随时间变化的, 但信息本身又相对稳定的数据集合, 它用于对管理决策过程的支持。”所谓主题, 在这里是指用户使用数据仓库进行决策时所关心的重点方面, 如银行存款情况、贷款情况、客户群情况、利润情况等; 面向主题是指数据仓库内的信息是按主题进行组织的, 为按主题进行决策的过程提供信息; 集成是指数据仓库中的信息不是从各个业务处理系统简单抽取出来的, 而是经过系统加工、汇

总和整理的，保证数据仓库内的信息是关于整个企业的一致全局信息；随时间变化则是指数据仓库内的信息并不只是关于企业当时或某一时点的信息，而是系统记录了企业从过去某一时点到目前的各个阶段的信息，通过这些信息，可以对企业的发展历程和未来发展趋势作出定量分析和预测。所谓信息本身相对稳定是指一旦某个数据进入数据仓库后，一般情况下将被长期保留，也就是数据仓库中一般有大量的插入和查询操作，但修改和删除操作很少。

商业智能（Business Intelligence，简称 BI）的概念最早是 Gartner Group 的 Howard Dresner 于 1996 年提出来的。当时将商业智能定义为一类由数据仓库（或数据集市）、查询报表、数据分析、数据挖掘、数据备份和恢复等部分组成的、以帮助企业决策为目的的技术及其应用<sup>[9]</sup>。

国内外许多大公司在 1998 年以后都推出了各自的数据仓库产品。IBM 于 1998 年 9 月发布 IBM DB2UBD5.2; ORACLE 于 1998 年 11 月发布的 RACLE 8I; SYBASE 于 1998 年 10 月发布 ASE11.9.2; 以及微软 SQL SERVER 都支持数据仓库技术。同时 ORACLE、SYBASE 还提供一套整数据仓库解决方案，包括了数据抽取、转换、清洗、装载(ETL)工具，同时还提供用于对数据仓库中数据挖掘、OLAP 和分析结果展示的工具<sup>[10]</sup>。

发展趋势上，数据仓库技术同 Internet 技术一样，成为信息社会中企业获得竞争优势的关键。“据美国 Meta Group 市场调查机构资料表明，《幸福》杂志所列全球 2000 家大公司中已超过 90%把 Internet 网络及数据仓库这两项技术列入其企业计划，而且有很多企业已经研究、开发与应用数据仓库系统技术。”<sup>[11]</sup>

国外数据仓库技术主要应用在证券行业中帐户分析、客户分析、资金交易分析、证券交易数据分析及防范风险的预警分析、证券业扩大经营中等广泛应用。在银行业中利用数据仓库的强大功能，许多银行建立了银行经营分析管理系统，信贷营销人员查询分析系统，对每一项投资方向做出科学判断，有效控制信贷、投资风险。另外，数据仓库技术在电信、税务等领域内都得到了广泛应用。

和国外应用相比，尽管国内各厂商在数据仓库方面的产品都有较好的性能，但在中国市场上效益不高。国内大部分数据库系统依然建立在传统的 OLTP 上，有一些企业已建或在建数据仓库，但整个数据仓库行业在系统可靠的、完善性、广泛接受性方面还需要有一定提高<sup>[12]</sup>。

国内数据仓库技术的应用也主要集中在保险、金融、证券、电信等行业<sup>[13]</sup>。中国证券登记结算有限责任公司已利用 NCR 公司的 Teradata 数据仓库系统解决方案，深度整合了沪、深两市数据，建成了 8TB 总容量的数据仓库系统，并利用这些海量数据开发出



应用分析系统，为证卷监管部门提供了准确的风险监控和市场交易信息，并为顾客提供了各种证券市场数据增值服务。

### 1.3 论文的目标与结构

#### 1.3.1 基本目标与长远目标

针对本电信企业对市场经营的要求，此系统工程基本目标不仅要建成一个经营分析系统，同时建成一个基于数据仓库的智能化的分析平台。

通过对客户资料及客户行为的分析，达到对不同客户群体彻底的了解，如客户群体构成、消费层次、客户贡献率、客户忠诚度、客户分布情况、客户消费习惯、客户潜在需求等，从而采取针对性营销措施，提高服务水平，扩大市场分额。

通过对客户消费情况的综合分析，掌握通信收入发展趋势，各项产品、各类客户对收入的贡献情况，控制各项产品的生命周期，提出如何增加收入的手段与措施。

通过分析与其他运用商相关联的数据，及时了解市场竞争动态情况，如其他运用商用户数、客户互通行为、竞争手段等，从而使本企业在竞争中立于不败之地。

通过建立分析模型对历史数据的深层次挖掘，为决策提供依据，同时掌握流失客户以及存在流失可能的客户特征，从而提出具体的减少客户流失的营销措施。

经营分析系统的长远目标是可以把企业的商业过程以一种直观而准确的方法描述出来。它可以使经营决策者及市场管理人员学会如何控制事物的发展，如何监视与测量产品性能，使他们能够得出仅靠直觉所无法得到的结论。

#### 1.3.2 论文的组织结构

本论文的结构包括以下几个部分：

论文第一章，介绍论文的技术发展及市场背景，简述该领域国内外研究现状。

论文第二章，介绍经营分析系统相关的数据仓库技术和数据分析技术。主要包括数据仓库相比数据库的一些特征、数据仓库的体系结构，介绍联机分析(OLAP)技术的基本概念、模式结构、常用的分析动作、数据模型等，总结数据仓库与数据分析的关系。

论文第三章，论述经营分析系统的整体设计情况，包括项目的需求确认和系统建设的目标、总体思路和设计原则，概念模型、分析模型、物理模型的设计以及实现该系统的软件体系结构和实现模块。为模型的具体实现打下基础。

论文第四章论述各子系统的详细设计情况。文章从数据获取子系统、经营分析子系统、应用层子系统三个层面对功能模块设计和实现进行了介绍。对数据获取子系统的介绍主要包括了 ETL 的设计要求、模块处理流程和模块详细设计进行了说明；本论文重

点对经营分析子系统部分分析主题的功能设计进行了详细说明，包括用户数分析、集团客户分析、个人客户分析、竞争分析、客户流失分析、收入分析等分析模型的设计和实现；并对应用层子系统的功能设计和实现进行了说明。

论文第五章对经营分析系统的部署实现和部分运行界面测试情况进行了说明。

论文结论部分，总结了论文的主要工作、研究成果及不足之处，展望了今后研究方向。

## 2 数据仓库技术概述

### 2.1 数据仓库的定义和特征

数据仓库没有一个公认且被标准化的定义。20世纪80年代中期，“数据仓库这个名词首次出现是在号称“数据仓库之父”William H.Inmon的《建立数据仓库》一书中<sup>[8]</sup>。不过随着人们对大型数据系统的研究、管理和维护等方面的深刻认识和不断完善，在总结并集中多个企业信息经验之后，都或多或少地指出了数据仓库有如下几个特点且达到了共识：数据仓库中的数据是面向主题的、集成的、稳定的而且随时间不断变化的，建立数据仓库目的是为了更好的支持决策分析。

#### (1) 面向主题

传统操作型系统都是围绕公司的应用进行数据组织的。与此不同，面向主题则是数据仓库中数据组织的基本原则，数据仓库中的所有数据都是围绕着某一既定主题组织、展开的。

由于数据仓库的用户绝大多数是企业的管理决策者，这些人所面对的是一些比较抽象的、层次较高的分析管理对象。从信息管理的角度看，主题就是在一个较高管理层次上对信息系统中的数据按照某一具体的管理对象进行综合、分类所形成的分析对象。从数据组织的角度看，主题就是一些数据的集合，这些数据集合能对分析对象进行比较完整的、一致的数据描述，这种描述不仅涉及数据自身，还涉及数据之间的关联。由于主题是在对较高层次上的数据抽象，这就使面向主题的数据组织可以独立于数据之间处理逻辑，可以很方便地在这种数据环境进行管理决策的分析处理。

#### (2) 数据集成性

数据仓库集成性是指根据决策分析的要求，将分散于不同处的源数据进行抽取、筛选、清理、综合等集成工作，从而消除应用问题的不一致性，以使数据仓库中的数据具有集成性。

#### (3) 数据的非易失性

一般而言，追加到数据仓库里的数据是不允许更改的，它不同于联机事务处理系统允许修改数据库纪录。这样数据仓库里的数据一般不会丢失。数据仓库的数据非易失性指数据仓库中的数据不经常进行更新处理。数据仓库除了进行查询以外，还可定期进行数据加载。数据的非易失性可以支持不同的用户在不同的时间里查询相同的问题时，获得相同的结果。消除了以前决策分析过程中面对同一问题，而结论不同的尴尬。

#### (4) 随时间的变化性

数据仓库的时变性，就是数据应该随着时间推移而发生变化。数据仓库必须能够不断地捕捉业务系统中的变化数据，将那些变化数据追加到数据仓库中去，也就是在数据仓库中不断的生成业务数据库的快照，以满足决策分析的需要。

数据仓库数据的变化不仅反映在数据追加方面，而且还反映在数据删除上和概括数据的变化上。数据仓库中的数据存储期限是有限的，一般保留 6-10 年，超过限期需要删除。数据仓库中的概括数据与时间有关的，概括数据需要按时间进行综合、抽取。因此在数据仓库中的概括必须随着时间变化而重新进行概括处理。

主要表现在以下几个方面：数据保存时间要远长于操作型系统中的数据时间期限；不包含当前值，操作型数据库里含有“当前值”的数据，而数据仓库中的数据仅是一系列某一时刻生成的复杂的快照；包含时间键码，操作型数据的键码结构也可能包含也可能不包含时间元素，而数据仓库的键码结构总是包含某时间元素的。

#### (5) 数据的集合性

数据仓库的集合性意味着数据仓库必须以某种数据集合的方式存储。目前数据仓库所采取的数据集合方式主要是以多维数据库进行存储的多维模式，以关系数据库方式进行存储的关系模式或以两两结合的方式进行存储的混合模式。

#### (6) 支持决策作用

数据仓库组织的根本目的在于决策支持。不同层次的管理人员可利用数据仓库进行决策分析，提高自己工作的管理决策质量、效果。数据仓库为决策者对数据自我分析提供了便利，提供了决策辅助分析的有力工具。

综上所述，经营分析系统是以数据仓库作为一个集成的信息库，把数据从各个信息源中提取出来，按照数据仓库所用的数据模型进行相应转换，并与仓库中现有的数据集集成在一起，以供系统分析使用。

## 2.2 数据仓库体系结构

数据仓库作为一个系统，理论上应该包括 3 个基本部分，其体系结构形式如图 2.1 所示。

(1) 数据获取：此部分负责从外部数据源获取数据，数据被提取出来，进行复制、重新定义格式等处理，准备装入数据仓库。即数据抽取、转换、清洗、加载几部分<sup>[14]</sup>。

在将数据放入数据仓库前需要进行缓存，以等待清洗和转换，这些数据缓存位置即提取仓库。为了保证减少对其他系统的影响，所有对数据的清理、整合都在数据接口区中进行。

(2) 数据存储和管理：此部分负责数据仓库的内部维护和管理，提供的服务包括数据存储组织、数据维护、数据分发及数据仓库的例行维护等，这些工作需要利用数据库管理系统的相关功能。元数据是管理数据的数据，提供数据资源的全面指南。数据集市是数据仓库的一个子集，也叫做部门级数据仓库。

(3) 信息访问：此部分属于数据仓库的前端，面向不同种类的最终客户。主要由查询生成工具、多维分析工具和数据挖掘工具等工具集组合而成<sup>[15]</sup>，以实现经营分析系统的各种要求。

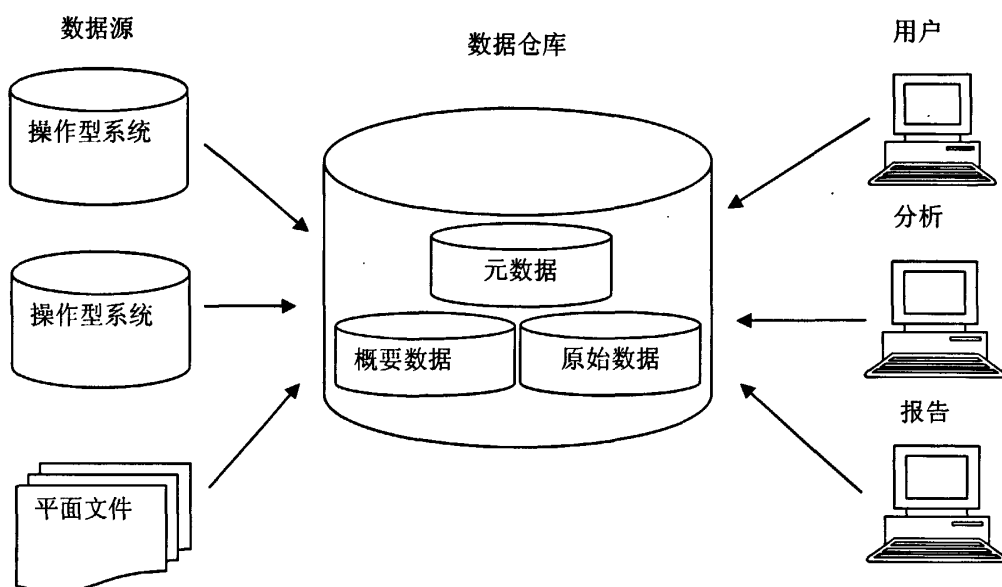


图 2.1 数据仓库的体系结构

Fig. 2.1 Architecture of data warehouse

### 2.3 OLAP 技术

数据仓库是一种管理决策分析的基础。若要有效的利用数据仓库的信息资源，需要有强大的工具对数据仓库中的信息进行分析决策。在线分析处理或联机分析处理就是得到广泛应用的数据仓库使用技术<sup>[16]</sup>。

OLAP 专门用于支持复杂的决策分析，支持信息管理和管理人员决策活动的一种决策分析工具。它可以根据分析人员的要求，快速地对大量数据进行复杂的查询处理，并

以直观的、易理解的形式将查询结果展现给各种决策人员，使他们迅速准确地掌握企业运营情况，了解市场需求。

OLAP 技术主要有两个特性：一是在线性，表现为对用户请求快速响应和交互式操作；二是多维分析(Mult-Analysis)性<sup>[17]</sup>，这也是 OLAP 技术的核心所在。

### 2.3.1 变量、维、事实

(1) 变量：变量是从现实系统中抽象出来，用以描述数据的实际含义。

(2) 维：维是与某一事件中相关的因素在关系模型中的抽象，比如客户打电话这一事件中包含了客户、地点、业务类型、业务提供商等。

(3) 维的层次：维存在着层次，维的层次主要是为了进行向下钻取及向上聚合，最终让用户能查到不同层次的数据。

(4) 维的取值：维的取值也称维的成员，由于维存在层次性，因此当维有多个层次时，维的成员由各个维层次的所有取值的组合构成。

(5) 事实：事实是不同维度在某一取值的交叉点，它是对事件的度量。

### 2.3.2 OLAP 模式结构

OLAP 中的数据是为分析而用的，它们的数据模式结构应该便于分析，在传统的数据库中以 ER 图或二维表为主，而数据仓库中则以多维模式为主，一般有两种结构：星型模式和雪花模式<sup>[18]</sup>。

(1) 星型模式：星型模式是一种多维结构，一般由基本表和维表组成，主体是基本表，而有关维的细节则描述在维表内，以达到简化基本表的目的，基本表与维表间有公共属性相连而使它们构成一个整体。

(2) 雪花模式：在星型模式中维呈单点状，但在很多情况下，维呈层次状即具有一定的深度。这种在维度中有纵向层次所构成的星型模式的扩充称雪花模式。雪花模式比星型模式更为复杂，它有利于数据分析和决策，能更好地反映对象的观察角度与深度，但对维表维护上的困难也比星型模式加大了很多。

在该经营分析系统中，对于各个主题分析模型的组织采用的是星型模式，这样便于维表的更新和维护，且完全能满足对各种主题分析的要求。

### 2.3.3 OLAP 的基本分析动作

多维分析是指对以多维形式组织起来的数据采取切片、切块、旋转等各种分析，以求剖析数据，使最终用户能从多个角度、多侧面地观察数据库中的数据，从而深入了解包含在数据中的信息。

(1) 切片：选定多维数组的一个二维子集的动作叫做切片，即选定多维数组(维 1, 维 2, …, 维 N, 变量)中的两个维：维 I 和维 J，得到的二维子集称多维数组在维 I 和维 J 上的一个切片，表示为：(维 I, 维 J, 变量)。

(2) 切块：选定多维数组的一个三维子集的动作称为切块，即选定多维数组(维 1, 维 2, …, 维 N, 变量)中的三个维：维 I、维 J、维 R，得到的三维子集称多维数组在维 I、维 J、维 R 上的一个切块，表示为：(维 I, 维 J, 维 R, 变量)。

(3) 钻取：维度是具有层次性的，维度的层次实际上反映了数据的综合程度。钻取包含向下钻取(Drill-down)和向上钻取(Drill-up)操作，钻取的深度与维所划分的层次相对应。

(4) 旋转：旋转即改变一个报告或页面显示的维方向。旋转可能包含了交换行和列，或是把某一个行维移到列维中去。

在经营分析系统中，通过 OLAP 服务提供的切片、切块、钻取、旋转等功能，使用户能从多方位、多角度对各种分析模型的数据进行考察和分析，得出经营分析中的一些规律，为决策提供依据。

#### 2.3.4 OLAP 的数据模型

根据 OLAP 服务器端的数据组织方法，OLAP 可分成以下几种结构：MOLAP(Multidimensional OLAP)，ROLAP(Relational OLAP)以及 HOLAP(Hybrid OLAP)。

(1) MOLAP：MOLAP 利用专有的多维数据库来存储 OLAP 分析所需的数据，并以多维视图方式显示。当决策者通过客户端递交分析需求给 OLAP 服务器时，OLAP 服务器直接检索多维数据库把得到结果返回给决策者。

(2) ROLAP：ROLAP 以关系数据库为核心，用二维表组织数据表达多维概念。ROLAP 一般采用星型模型和雪花模型来组织数据，ROLAP 的底层数据库是关系数据库。当用户通过客户端工具提交多维分析请求给 OLAP 服务器，服务器动态将这些请求转换成 SQL 语句执行，分析结果转化为多维视图返回给用户。

(3) HOLAP：是 ROLAP 和 MOLAP 技术优点的有机结合，当客户端用户提交一个分析请求时，系统透明地从多维数据库中获取综合数据或从关系数据库中提取细节数据。

MOLAP 具有查询速度快，所需磁盘空间较大特点；ROLAP 不需要额外的磁盘空间、并行处理能力强、维度变化和数据变化适应性强但查询速度稍差特点；在该经营分析系

统中数据量大，如果采用 MOLAP，则要大量的磁盘空间，所以选择 ROLAP 模式，但为了提高查询速度，系统通过经营分析程序，对一些分析主题的数据预先进行了汇总。

## 2.4 数据仓库与数据分析的关系

数据仓库是数据分析数据来源，因为数据仓库的数据是针对某些分析主题的数据集合，并且仓库中的所有数据等都具有一致性。数据分析是对数据仓库中的数据进行分析、决策的工具，数据分析中的数据都是从数据仓库中获取。数据分析包括联机分析处理 (OLAP)和数据挖掘(DM)两种分析工具<sup>[19]</sup>。OLAP 是一种验证性数据分析，用户提出某种问题或假设，OLAP 负责从上到下、由浅到深的展现相关详细信息，供用户判断提出假设是否合理。DM 是一种挖掘性质数据分析，它能自动地发现事物间隐含关系和特征模式，并且可以利用这些特征模式进行有效预测<sup>[20]</sup>。

## 2.5 本章小结

本章主要对与经营分析系统开发相关的数据仓库技术进行了概述，其中包括了对数据仓库的定义，数据仓库系统的特点进行了说明；对数据仓库体系结构的数据获取、数据存储和管理、信息访问三层功能进行了说明。本章还对数据仓库中的联机分析处理数据分析工具进行说明：首先介绍了 OLAP 基本概念，然后说明了 OLAP 基本动作、数据模型，最后对数据仓库、联机分析和数据挖掘三者的关系进行了说明。



### 3 经营分析系统的整体设计

#### 3.1 经营分析系统需求确认

为实现对信息数据的智能化加工和处理,为市场经营、决策提供及时、明确、合理的决策依据,利用业务支撑系统产生的宝贵数据资源,从大量的业务数据中整理提取出有用的信息建立经营分析系统。由于目前本电信企业本地数据量相当庞大,业务系统较多的现状,不可能从企业庞大的业务处理系统中随时获取所需的信息,因此必须引入数据仓库技术来支持企业的经营分析对信息的要求。数据仓库是可以将大量企业数据转换成可靠的、商用的决策支持信息的最好解决途径。采用数据仓库建立经营分析系统可以帮助电信企业建立完整的客户资料档案,整合计费信息与业务信息,提高客户了解并提升服务水平。根据上述需求分析,该经营分析系统设计应达到如下目的:

(1) 数据仓库的组建,主要包括建立基于一些分析主题的数据模型,设计数据基本表及与基本表相关联维表,定期对主题基本表、相关联维表数据进行汇总<sup>[21]</sup>;为提高系统对海量数据的查询能力,要充分考虑到系统的物理设计,以及逻辑设计与物理设计的关系等。

(2) 根据数据仓库中定义主题的需要从客户系统、原始计费信息系统、帐务系统等各业务系统中对各种需要的数据进行获取、清洗、整合形成统一标准的数据装载进数据仓库中。由于现有的业务系统自成体系、互不相联、信息独立,没有形成整体提供决策信息,管理者只能得到一些粗陋、分散的报表和数据,而将数据进行整合是决策依据的基础,需要把这些整合的数据导入,实现各级领导对各项业务快速、灵活、准确的查询分析,并提供规范的月度经营情况类报表。对于原始计费系统中的文件代码为十六进制,还要对原始计费数据进行翻译、解读。得到系统需要的字段值,综合后装载入库组成每个用户通话详单以供专题分析使用。

(3) 分析模型的展示及应用,主要包括生成经营报表及对一些主题进行 OLAP 分析,其中 OLAP 分析主要包括了业务发展情况分析、客户发展情况分析、竞争情况分析、集团客户分析、个人客户情况分析、亲情电话分析等主题模型展示与实现,并能对潜在流失客户提供预警及明细<sup>[22]</sup>。

#### 3.2 整体设计原则

经营分析系统按照三层结构的方式进行设计,所谓三层结构指数据获取层、数据存储层和数据应用层三层,在应用上使用基于 WEB 三层架构形式。在系统实施过程中遵循的原则包括:

(1) 必须按照数据一致性原则将数据转移到数据仓库中，确保数据仓库中数据格式及形式上完全一致。

(2) 充分利用各类数据源，将位于不同数据库、不同平台中的数据按照相关规则集中在一个库中。

(3) 为提高经营分析系统数据分析的效率，存储的数据应是一些经过预先加工、计算的汇总数据。

(4) 达到随机、准确查询和分析效果。通过一致、快速交互访问各种可能的信息查询视图，协助决策人员、管理人员、分析人员掌握深层次的规律，如实准确的反映了企业经营状况。

### 3.3 构建数据仓库模型

#### 3.3.1 模型构建目标

(1) 统一的数据仓库。即在一个数据仓库中，包含了企业所有部门的分析信息，数据高度集成，访问同一控制，能够满足所有部门的分析需求。

(2) 有效的信息系统。系统分析信息模型应该支持所有报表，同时提供对应的分析，信息模型中的相应的数据信息应与经营指标的定义一致。

(3) 独立的数据模型。系统将经营分析需要的数据从操作性的数据环境中提取出来，实现数据的积累与集中，达到事务处理系统与经营信息的分离。

(4) 灵活的组织形式。经营分析需求的多变性需要模型的灵活性和扩展性，信息模型的体现除应按数据的本质特征整合数据，同时在维的设计方面要体现一定的灵活性，以达到适当地适应需求变更<sup>[23]</sup>。

#### 3.3.2 模型构建总体思路

在构建经营分析信息模型时，先根据企业目前对经营信息的统计分析需求和经营信息数据的本质，确定宏观分析主题和微观的子主题，及企业统计、分析时关心的维对象和维层次。设计时应主要考虑以下方面：

(1) 维和度量的设计。在分析主题的设计中具体应包含那些维和度量，在维的分层和各层的取值时，本着尽量使各维的交叉取值都有意义原则和尽量使维各层都可以向上汇总原则设计。只有不同维的交叉取值有意义，维的度量才有意义；维各层可以向上汇总并且向上汇总度量的值有意义，才能支持 OLAP 分析型报表的“上钻、下钻”。

(2) 基本表的设计。一个分析主题可对应一个或多个基本表，一个子主题在系统中与基本表一一对应。基本表中主要包含维字段和度量字段(指标)两部份。在每个基本表中描述和每个子分析主题关联的维对象和度量，维对象主要是定性地描述子主题的观察角度，度量主要是定量地的表示该子主题在各维约束条件下联通各部门关心的量值。

(3) 存储模型的选择。设计基本表和各层维的关联关系时，可采用星型模型或雪花模型，因为星型模型效率高、可扩展性好，本系统的分析主题采用星型模型设计。设计时对维和维的层次进行合理划分，以保证模型的可扩充性。提供数据的适当聚集以满足用户的一般查询且数据模型应支持企业现有的报表和分析型需求。

### 3.3.3 概念模型的设计

按照主题来组织数据仓库，将企业的经营模型映射到数据仓库的高层模型中，就可得到系统的概念模型，目前本电信企业经营业务系统主要由客户、帐务、营销、服务、服务使用、客服、结算、资源等八个业务组成，业务间存在一定关系，具体关系情况如图 3.1 所示。

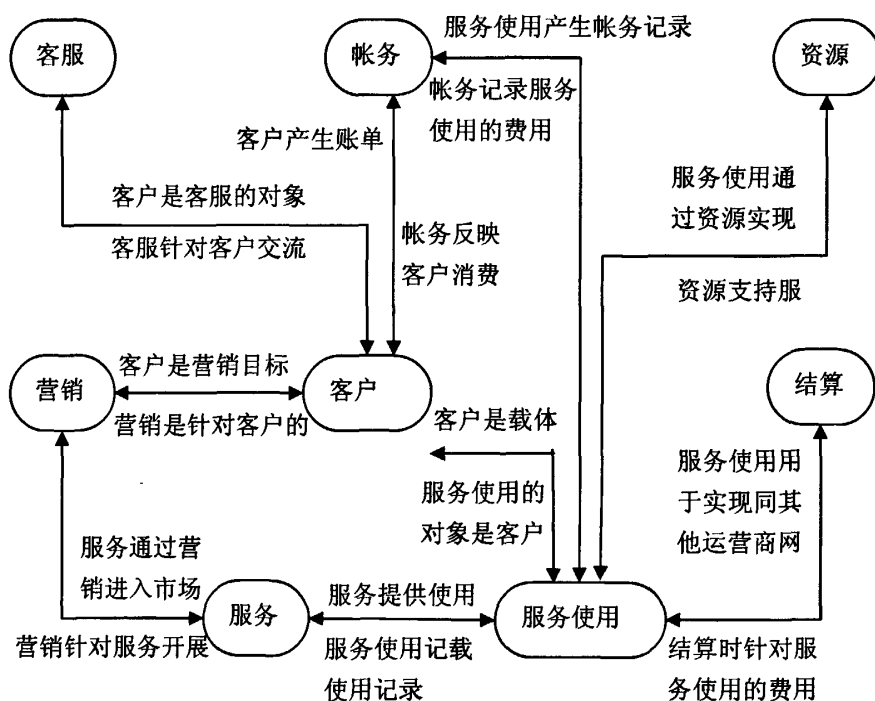


图 3.1 经营分析系统的概念模型

Fig. 3.1 Conceptual framework of BI

该模型主要反映本电信企业的业务关系，包括所有主题、主要实体及其相互之间的关系等，所以本电信企业经营分析系统的分析主题也应该由客户分析主题、帐务分析主题、营销分析主题、服务分析主题、服务使用分析主题、客服分析主题、结算分析主题、资源分析主题等八个分析主题组成。下文对与客户、帐务、服务、服务使用相关的主题进行分析。

### 3.3.4 分析模型的设计

根据概念模型得到分析主题，为了在 OLAP 中能对这些主题进行多层次、多角度、多方位的分析，论文对部分分析主题建立了相应的分析模型，具体分析模型功能描述如下：

(1) 客户发展情况模型。以客户发展情况为分析主题，在不同时间段内，按客户年龄层次、性别、客户类型、消费层次、业务品牌等维度对客户总量情况等进行分析。

(2) 新用户发展情况模型。以新用户发展情况为主题，提供在不同时间段内，按年龄层次、性别、客户类型、消费层次、业务品牌等维度对网上有效新增加客户数以及新增客户数的构成情况进行分析。

(3) 收入情况分析模型。以话费收入情况为主题，不同时间段内，按年龄层次、性别、客户类型、消费层次、业务品牌等维度对收入总量、收入增量、收入结构、客户欠费情况及结构情况进行分析。

(4) 市场竞争分析模型。以客户话单为基础，通过提取网间通话话单，对其他运用商客户发展状况、客户呼叫行为、其他运用商策略追踪及影响进行分析。主要包括了其他运用商用户总量分析、其他运用商新增用户分析、其他运用商流失用户分析。

(5) 流失用户分析模型。对流失客户的属性，主要以套餐、代办点、客户群作为考察点，把流失客户的欠费情况、呼叫转移次数(电信、移动)、话费突变情况作为主要因素对流失的用户进行分析。

(6) 集团客户资料分析模型。主要按照有执照号码的用户，在不同时间段内，按部门、年龄层次、代办点类型、用户状态、入网时长、消费层次、性别、套餐、用户类型等维度对集团客户的构成、新增、欠费等情况进行分析。

### 3.3.5 物理模型的设计

在分析系统中，每个分析主题下包含多个子主题，如果对于每个子主题都直接从各种数据源中直接提取数据，系统运行效率将会大为降低，为了提高系统效率，本系统首先从客户资料系统、营帐等系统中对各种数据进行提取、清洗形成各种与分析主题相关

的基本表和维表，对于每个子主题模型在这些相关基本表的基础上进一步分析汇总得到。根据经营分析主题要求及实际情况，目前系统主要的事实基本表有客户资料基本表、客户每月历史状态表、客户帐单表、通话详单表，其它各种分析模型都是在几个事实基本表的基础上进一步分析汇总得到；目前系统的维表主要包含了用户类型维表、套餐类别维表、年龄层次维表等。下文给出部分基本表及维表的物理结构：

客户资料基本表：该表主要包括了每个客户的用户姓名、性别、年龄层次、生日、职业类别等自然属性；结算帐号、消费层次、用户状态、信用额度等使用情况属性；套餐类型、业务类别等业务属性，具体结构如表 3.1 所示：

表 3.1 客户资料基本表  
Tab. 3.1 Basic table of customer data

字段名	类型描述	说明
YHBH	NUMBER(10)	用户编号
DHHM	VARCHAR2(12)	电话号码
YHXM	VARCHAR2(40)	用户姓名
YHSR	DATE	用户生日
JSH	VARCHAR2(20)	机身号
SIM	VARCHAR2(20)	SIM 卡号
KHSJ	DATE	开户时间
BTSJ	DATE	报停时间
DZSJ	VARCHAR2(8)	呆帐时间
JSZH	VARCHAR2(14)	结算帐号
IDCARD	VARCHAR2(20)	证件号码
XYCE	NUMBER(8)	信用初额
JSBZ	VARCHAR2(1)	寄送标志
KH_MONTH	VARCHAR2(8)	开户月份
AGE_LEVEL	VARCHAR2(1)	年龄层次
IDCARD_TYPE	VARCHAR2(1)	证件类别
SEX	VARCHAR2(1)	性别
KHQ	VARCHAR2(6)	集团编码
ZY	VARCHAR2(2)	职业类别
KIND	VARCHAR2(1)	用户类型
INCOME	VARCHAR2(1)	收入级别
AGENT	VARCHAR2(6)	代理商

(转下页)

表 3.1 客户资料基本表 (接上页)

Tab. 3.1 Basic table of customer data

字段名	类型描述	说明
FEE_LEVEL	VARCHAR2 (1)	消费层次
INNET_TIME	VARCHAR2 (1)	入网时长
STATUS	VARCHAR2 (1)	用户状态
THJB	VARCHAR2 (1)	通话级别
BANK	VARCHAR2 (4)	银行信息
BILL_TYPE	VARCHAR2 (1)	缴费方式
DWXZ	VARCHAR2 (1)	单位性质
XQAH	VARCHAR2 (2)	兴趣爱好
EDUCATION	VARCHAR2 (1)	教育程度
INNET_TYPE	VARCHAR2 (5)	套餐类型
XYED	VARCHAR2 (1)	信用额度
YWLB	VARCHAR2 (1)	业务类别
XQ	VARCHAR2 (2)	行区
TCQYSJ	DATE	套餐启用时间
ZZH	VARCHAR2 (12)	执照号 (VIP 集团客户标志)

客户每月历史状态表：该表主要包含了用户性别、用户年龄层次等一些自然属性；用户每月消费情况、业务使用情况等属性，具体结构如表 3.2 所示。

客户帐单表：该表主要用来记录客户帐单情况，除了包括了客户状态、套餐类型、客户群、客户类别等属性外，还记录了各种费用的构成情况(主要包括了本地移动费、长途费用、短信、各种漫游费用等各种新业务使用费用)及上个月话费情况、费用变化、费用级别情况等，具体结构如表 3.3 所示。

通话详单表：该表详细记录了客户每次通话的具体情况 & 通话客户的各种属性。其中通话具体情况主要包括了用户号码、用户编号、月份、本方归属地、对方号码、对方归属地、小区代码、基站代码、通话时长、呼叫日期、呼叫时间、入中继代码、出中继代码；通话用户各种属性主要包含通话客户类型、套餐类型、年龄层次、入网时长、集团编码、代理商类型等，具体结构见附录 A。

与客户每月历史状态表、客户资料基本表、客户帐单表、通话详单表等基本表相关联的还包括了用户类别、年龄段(年龄层次)、套餐类型、集团编码等维表，本文不再详细列举。

表 3.2 客户每月历史状态表  
Tab. 3.2 Monthly statement of customer status

字段名	类型描述	说明
YHBH	NUMBER(10)	用户编号
XQ	VARCHAR2(2)	行区
YWLB	VARCHAR2(1)	业务类别
MONTH	VARCHAR2(6)	统计月份
SEX	VARCHAR2(1)	性别
INNET_TYPE	FVARCHAR2(5)	套餐
KHQ	VARCHAR2(6)	集团编码
KHSJ	DATE	开户时间
INNET_TIME	VARCHAR2(1)	入网时长
AGENT	VARCHAR2(6)	本月开户点
STATUS	VARCHAR2(1)	月末状态
KIND	VARCHAR2(1)	用户类型(合约、预付费)
AGE_LEVEL	VARCHAR2(1)	年龄层次
FEE_LEVEL	VARCHAR2(2)	费用层次
YJJE	NUMBER(10, 2)	预交款结余
FEE_CHANGE_LEVEL	VARCHAR2(1)	话费改变层次
ARREAGE_SUM	VARCHAR2(1)	欠费总额
ARREAGE_COUNT	VARCHAR2(1)	欠费次数
ARREAGE_TIME	VARCHAR2(1)	欠费时长
HZGSM_COUNT	VARCHAR2(1)	呼转到移动级别
HZPSTN_COUNT	VARCHAR2(1)	呼转到联通级别
DHHM	VARCHAR2(12)	电话号码
HZGSM_NUM	NUMBER(6)	呼转到移动次数
HZPSTN_NUM	NUMBER(6)	呼转到电信次数
CALL_TIMES	NUMBER(8)	本月通话次数
CALL_DURATION	NUMBER(8)	本月通话总时长
SUM_FEE	NUMBER(10, 2)	本月话费总额

表 3.3 客户帐单表

Tab. 3.3 Customer billing table

字段名	类型描述	说明
YHBH	NUMBER(10)	用户编号
DHHM	NUMBER(12, 0)	电话号码
MAKE_MONTH	NUMBER(6)	统计月份
MONTH	NUMBER(6)	帐单月份
YWLB	VARCHAR2(2)	业务类别
JSZH	VARCHAR2(14)	结算账号
KIND	VARCHAR2(1)	用户类别
STATUS	VARCHAR2(1)	状态
YD	VARCHAR2(6)	营业点
INNET_TYPE	VARCHAR2(5)	套餐类型
KHQ	VARCHAR2(6)	客户群
KHSJ	VARCHAR2(8)	开户时间
JFF	NUMBER(1)	缴费否
CZZ	VARCHAR2(6)	操作者
TOTAL_FEE	NUMBER(10, 2)	总话费
HALF_FEE	NUMBER(10, 2)	上月总话费
FEE_LEVEL	NUMBER(2)	费用级别
LAST_TOTAL_FEE	NUMBER(10, 2)	上月总话费
FEE_CHANGE	NUMBER(10, 2)	费用变化
FEE_CHANGE_LEVEL	NUMBER(2)	费用变化级级别
THSJ	NUMBER(10)	通话时间
RQ	VARCHAR2(8)	日期
FEE01	NUMBER(9, 2)	本地移动费
FEE03	NUMBER(9, 2)	一省外漫游移动费
FEE04	NUMBER(9, 2)	本地国内长途费
FEE05	NUMBER(9, 2)	本地长途附加费
FEE09	NUMBER(9, 2)	省内漫游移动费
FEE10	NUMBER(9, 2)	省内漫游长途费
FEE11	NUMBER(9, 2)	省外漫游长途费
FEE15	NUMBER(9, 2)	国际漫游(含漫游长途)
FEE25	NUMBER(9, 2)	数据业务费
FEE35	NUMBER(9, 2)	点对点短信费



### 3.4 系统的体系结构和实现模块

#### 3.4.1 经营分析系统架构

本电信企业经营分析系统以数据仓库、联机分析处理等先进技术为基础，通过数据整理和数据分析对海量数据和历史数据加以利用、分析，提供决策支持。系统总体架构设计如图 3.2 所示：

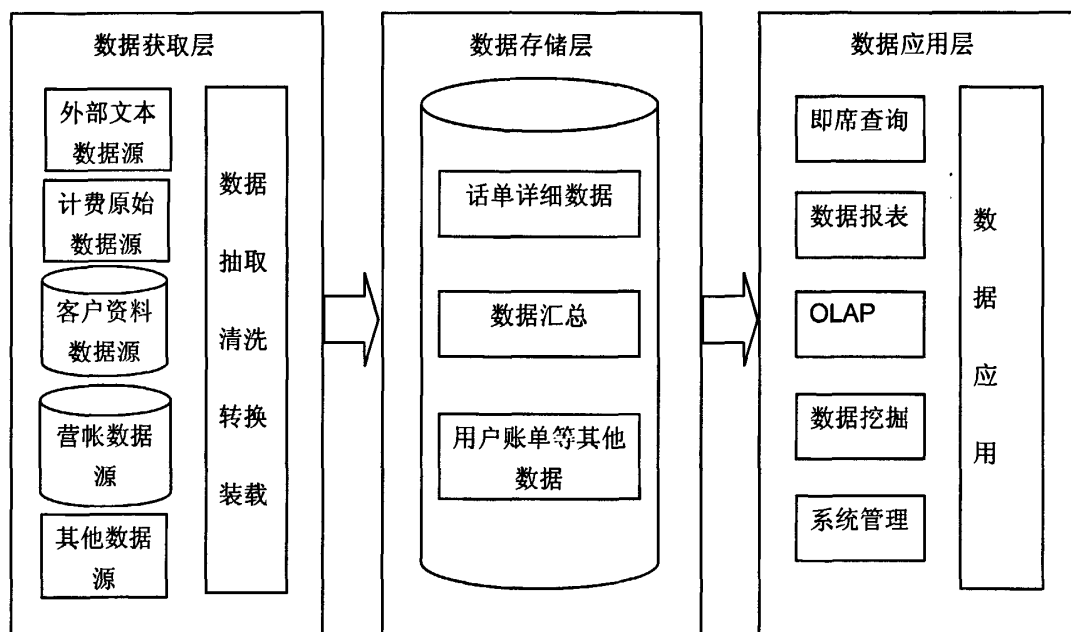


图 3.2 经营分析系统的架构

Fig. 3.2 Framework of BI

(1) 数据获取层主要是从交换机存储单元中提取原始计费文件，并对原始计费文件按照数据仓库构建的要求进行翻译、从客户资料数据库中获得用户数据，从营帐数据库中获得帐务等其它数据并通过对数据的转换、清洗、整合，然后装载到数据仓库。

(2) 数据存储层对数据仓库中的数据进行集中存储和管理，并根据不同主题对数据进行汇总和集中，以提高系统快速查询和分析能力。

(3) 数据应用层主要对数据仓库中的数据进行分析查询并通过前端展示工具对报表、OLAP、数据挖掘的结果进行分析展示。

#### 3.4.2 经营分析系统体系结构

经营分析系统体系结构主要由数据源、平台层、应用层三个层次组成，每层由不同模块来实现不同的功能，具体结构如图 3.3 所示。各功能模块说明如下。

营帐和客户资料所有数据都在业务支撑系统(BOSS)中,通过定期从 BOSS 中抽取经营分析系统所需要的数据字段,并对其中的一些数据进行转换、清理、整合形成经营分析系统的基本表和维表放入数据仓库。其它数据源是指为提高整个系统的扩展性和分析业务的需要将来要采集的数据,如网管数据。

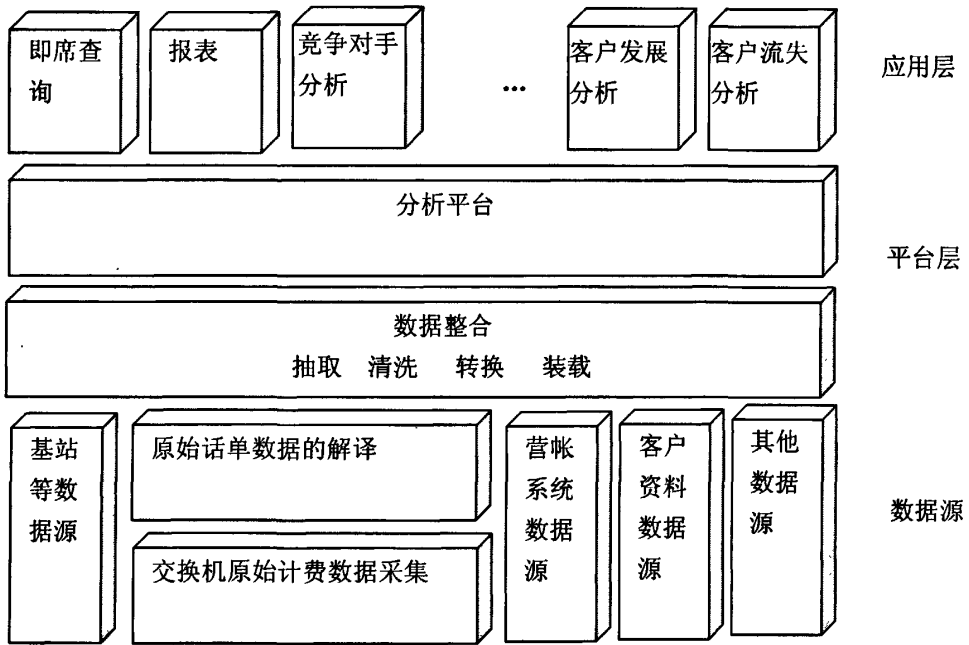


图 3.3 经营分析系统体系结构  
Fig. 3.3 Architecture of BI

**交换机原始计费数据采集:** 由于本电信企业目前有西门子、华为二种交换机型,且这二种交换机型分布在不同的地方,为了提高系统可靠性,在计费结算中心放置一台计费数据采集服务器,定期从交换机中采集原始计费数据,供原始话单数据解译程序解读。

**原始话单数据解译:** 对采集到的原始交换机计费数据实施解译,把计费文件中按一定格式构成的十六进制代码转化成十进制代码,并抽取系统需要的数据字段装载到 ORACLE 数据库中。

**基站等数据源:** 主要包括对基站数据、号码归属地等其它数据源进行采集。

**数据整合：**从初级计费数据库中提取话单数据、各种帐务数据并对用户数据等进行综合形成分析平台所需的标准数据装载进数据仓库。数据整合平台主要是对数据进行收集、转换、综合、整理、储存，作为分析平台基础。

**分析平台：**以应用层的分析主题为依据，对基本表的数据进行多维度、多层次的汇总，以空间换时间原则，预先通过对各个基本表中的数据根据不同条件整理汇总获得结果以供应用层使用和展示。

**应用层：**对分析平台中的经营分析结果以表格或图形的形式显示以供企业领导、市场部门等决策者对经营分析情况进行了解。主要包括客户发展情况分析、市场竞争分析、收入分析、各种分析报表等经营分析主题的查询和展现。

### 3.4.3 系统的实现模块

根据系统需求提出的系统实现目标，整个经营分析系统的构建由数据获取层子系统、经营分析子系统、应用层子系统三个部分组成。其中每个子系统又有若干实现模块组成，完成经营分析系统不同功能。系统总体功能模块图如图 3.4 所示：

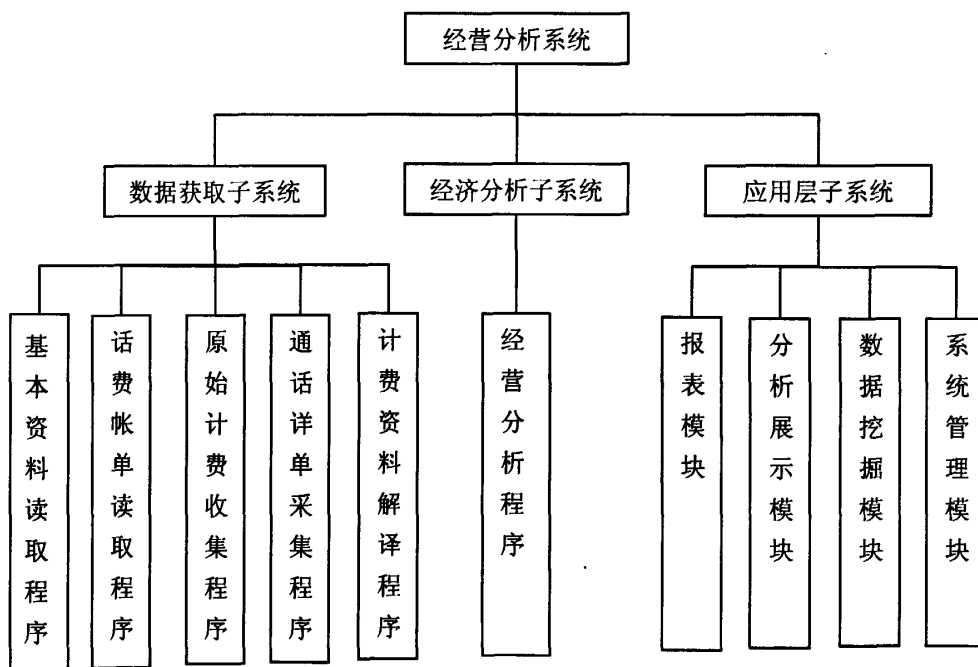


图 3.4 系统功能模块图

Fig. 3.4 System function module diagram

(1) 数据获取子系统：主要完成从不同的数据源中获取各主题分析所要的数据，并对获取的数据进行转换、清洗、组合形成一致的数据装载进数据仓库。该子系统主要包括基本资料读取程序、话费帐单读取程序、原始计费收集程序、通话详单采集程序、计费资料解译程序。

(2) 经营分析子系统：该系统是经营分析系统的核心部分，主要运行经营分析各项程序。该程序主要对从数据获取层所获得的高质量的各种基本表及计费数据进行统计、汇总，并把所有这些不同粒度的汇总数据按照不同主题装载到数据仓库，以供 DM 和 OLAP 使用。主要包括了对客户发展分析、集团用户分析、个人用户分析、流失用户分析、亲情电话分析、收入分析，其他运用商分析等模型的实现。

(3) 应用层子系统：该系统包括了联机分析模块、简单数据挖掘模块、系统管理模块和报表模块。

### 3.5 本章小结

本章根据经营分析系统的需求分析和系统建设的原则和目标，提出了经营分析系统数据仓库模型的设计思路，并构建了该系统数据仓库概念模型以及与模型相关基本表和部分维表的逻辑模型，并对客户发展情况、新用户发展情况、收入情况分析、市场竞争分析、流失用户分析、集团客户资料分析等六大分析主题模型的功能进行了介绍。本章最后阐述了该经营分析系统的体系结构和实现系统的功能实现模块。

## 4 经营分析系统功能模块设计

### 4.1 数据获取子系统功能模块设计

#### 4.1.1 数据获取子系统设计要求

数据获取子系统主要功能包括数据的提取、清洗、整合、装载，需要从不同的数据源中获得数据。由于各系统数据量庞大，为保证系统效率，数据获取子系统各模块设计中应达到以下要求：

- (1) 模块设计及应用期间要保证所有原业务系统的安全可靠性，对原业务系统影响要降到最小程度；
- (2) 数据在清理后应保证数据的质量和正确性，对一些错误的值和缺省的值进行处理，使得进入仓库中的数据具有一致性；
- (3) 要保证数据获取模块足够的健壮性；在进行数据提取、清洗、整合、加载时要保持系统高效性、可靠性，保证数据能在预定的时间内整合、加载完毕。

#### 4.1.2 ETL 模块处理流程

从各种数据源抽取、清洗、整合、加载数据到数据仓库的工作统称为 ETL。本系统主要从业务支撑系统（BOSS）中的客户系统、帐务系统中抽取、清洗数据，并从交换机计费系统中采集、解读话单，然后对这些数据整合，使各类数据达到一致性，继而形成经营分析程序所需要的基本表和维表，然后加载进数据仓库。

集成的 ETL 工具很多，例如 Informatica 的 PowerMart，Oracle 的 Warehouse Builder(OWB)，IBM 的 Warehouse Control Center 等，但在实际应用上都有许多不足，为了提高系统性能，针对该系统对所有程序进行开发。具体 ETL 模块处理流程如图 4.1 所示。

在 ETL 实现的模块设计中，为减少对原有业务支撑实时运行系统影响，与其它数据仓库 ETL 的设计相比本系统增加了一个接口数据区，作为各个数据源到数据仓库数据的过度过程，以提高 ETL 系统的安全可靠性。

数据获取子系统主要功能模块包括基本资料读取程序、话费帐单读取程序、原始计费收集程序、通话详单采集程序、计费资料解译程序几部分组成，各个程序的调用按照一定的规则进行。具体数据流程设计见图 4.2 所示。

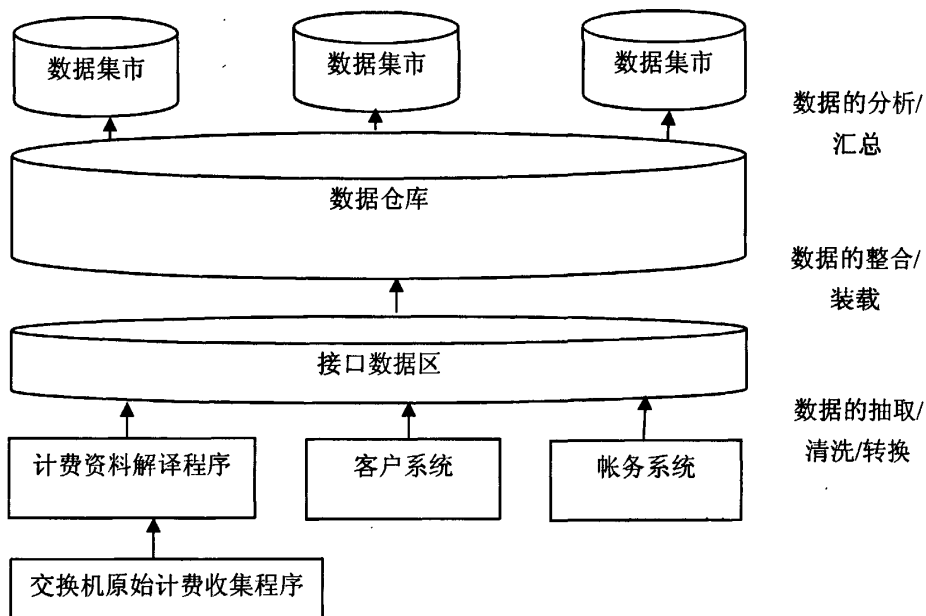


图 4.1 ETL 模块处理流程图

Fig. 4.1 Flow chart of ETL module processing

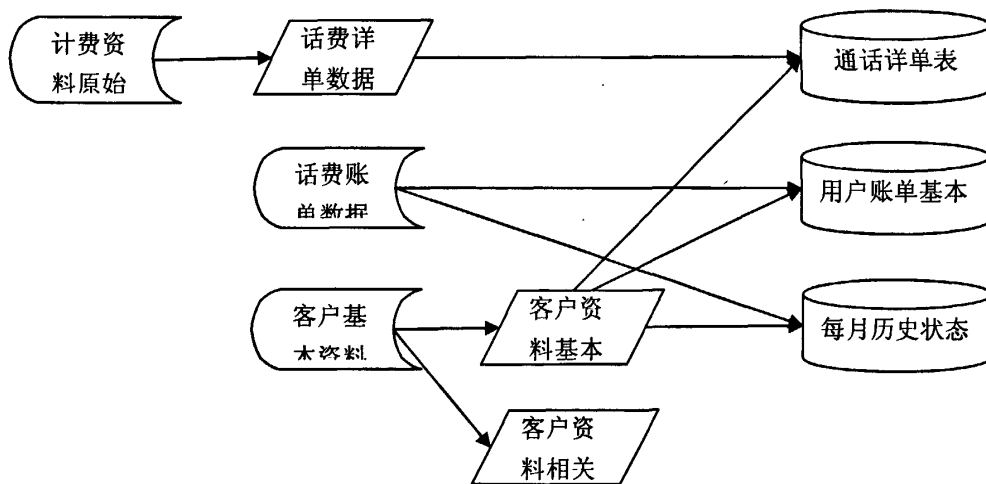


图 4.2 ETL 数据流程设计图

Fig. 4.2 ETL data flow design diagram

(1) 基本资料读取程序从客户系统把客户档案、客户类别、银行帐号结算信息等数据读到数据接口区后,对数据进行清洗和综合、生成一些要分析的数据维表和客户资料基本表,最终得到客户资料基本表和若干与客户资料相关的维表。

(2) 话费帐单读取程序从帐务系统中把需要的各种帐单数据读取到数据接口区后,进行清洗、综合,得到用户帐单基本表、每月历史状态表和用户费用等一些其它数据表。

(3) 通过原始计费收集程序得到计费资料的原始数据,通过计费资料解译程序对原始数据进行解读,通过 SQLLOADER 函数装载进数据接口区。与数据接口区的通话详单数据和客户资料基本表进行整合形成与其它数据源一致的呼叫记录数据,加载到通话详单表中。

从以上 ETL 模块流程设计过程中可以看出,在 ETL 的执行过程中不同模块的调度是有先后顺序的。首先要读取用户基本资料数据然后才能运行话费帐单读取程序和通话详单采集程序。

#### 4.1.3 ETL 模块的详细设计

ETL 模块主要包括基本资料读取程序、话费帐单读取程序、原始计费收集程序、通话详单采集程序、计费资料解译程序几部分组成。其中基本资料读取程序、话费帐单读取程序、通话详单采集程序由 ORACLE PRO \*C 开发。

##### (1) 基本资料读取程序

程序名称:JBZLDQ-INPUT,具体流程图如图 4.3 所示:

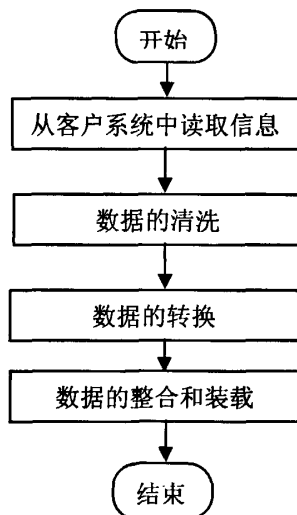


图 4.3 基本资料读取程序流程图

Fig. 4.3 Flow chart of basic data reading program

加载方式：每天一次，数据完全覆盖。

(2) 话费帐单读取程序

程序名称：HFZDDQ-INPUT，具体流程图见图 4.4。

加载方式：用户帐单表为每天一次，每月历史状态表为每月一次。

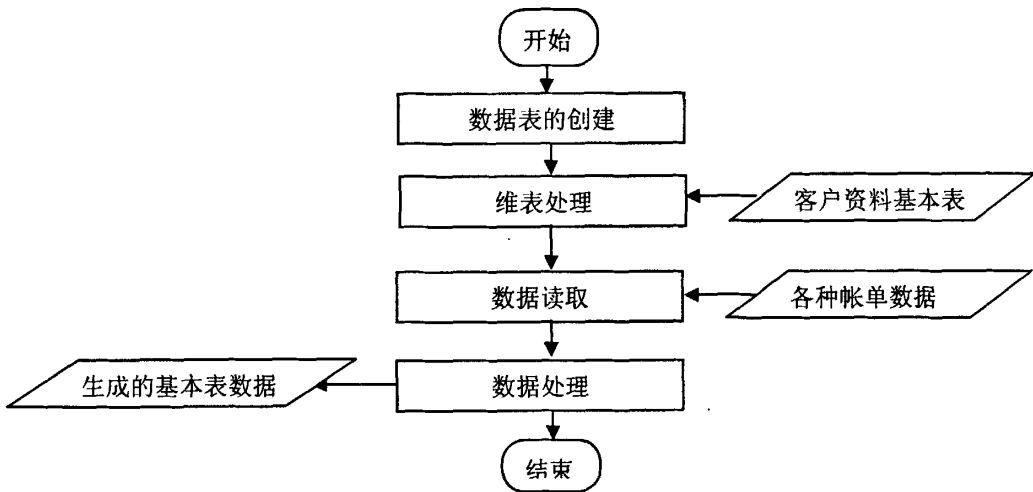


图 4.4 话费帐单读取程序流程图

Fig. 4.4 Flow chart of phone bills reading program

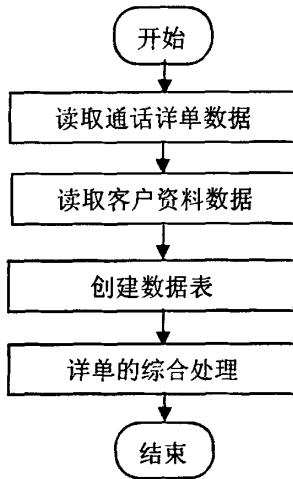


图 4.5 通话详单采集程序流程图

Fig. 4.5 Flow chart of call detailed record collecting



### (3) 通话详单采集程序

程序名称：THXDCJ-INPUT，具体流程图如图 4.5 所示。

加载方式：每天追加一次且次序要后于基本资料读取程序。

### (4) 原始计费收集程序

程序名称：SJ-AUTOFTP，具体流程图如图 4.6 所示。

加载方式：实时加载

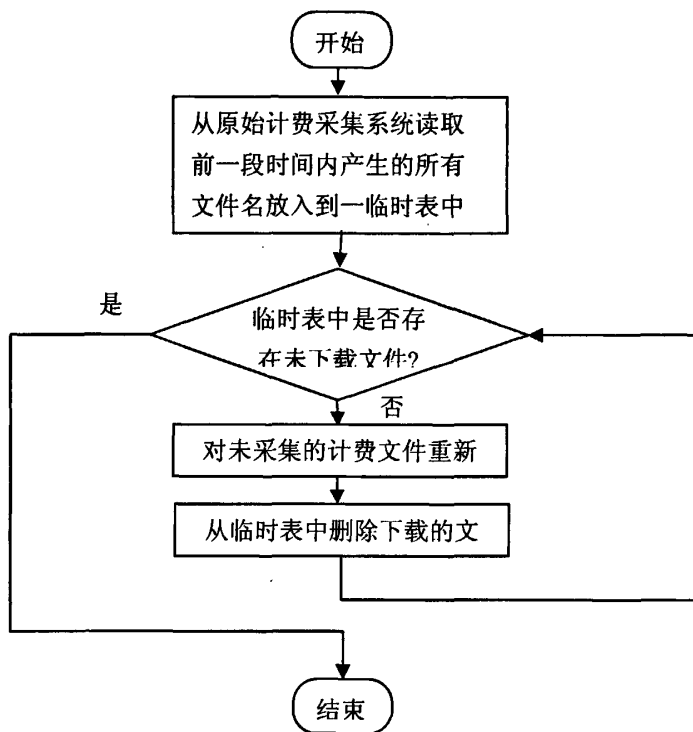


图 4.6 原始资料收集程序流程图

Fig. 4.6 Flow chart of raw data collecting program

### (5) 计费资料解译程序

计费资料解译程序包括话单解译、话单资料管理两部分功能。其中话单解译模块主要包括了对华为、西门子交换机的原始话单进行解译并把结果装载进数据库中，并提供日志查询功能。话单资料管理模块实现对话单数据的管理，包括数据备份、删除等。本文对话单解译程序模块设计进行说明。

话单解译模块主要功能是对华为、西门子交换机原始计费文件进行解读，并把解读出的话单加载进接口数据库。该模块还包括了数据自动装载、选择解读入库、上次操作记录、已读记录等附加功能。由于篇幅有限，本文只对话单解译模块设计进行简要说明，具体见图 4.7 所示。

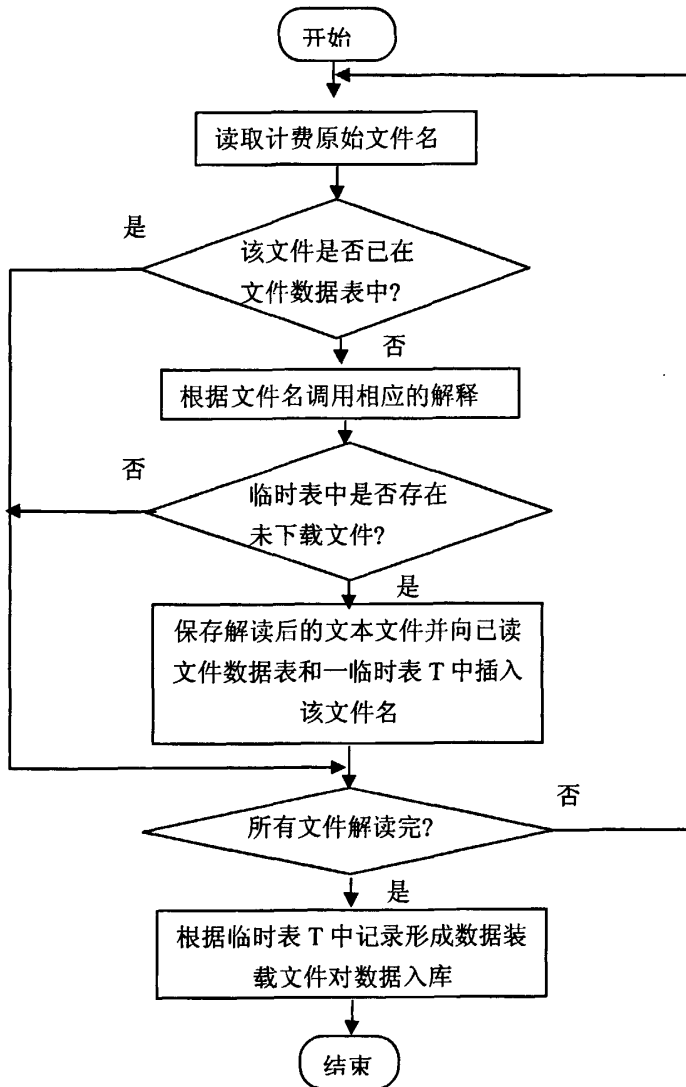


图 4.7 话单解译模块流程图

Fig. 4.7 Flow chart of bill converting module

## 4.2 经营分析子系统功能设计

本电信企业经营分析系统以实际业务需求分析为基础，结合目前各个实时系统的实际条件，经过总结和归纳确定经营分析主题，常用的主题包括包括了用户数分析、集团客户分析、个人客户分析、竞争分析、客户流失分析、收入分析等，下文列举这些分析主题模型的功能设计。这些模型是通过经营分析程序从各个数据源经过 ETL 后得到基本表后汇总分析得出，经营分析程序则是由若干函数组成，每个函数完成一个独立主题分析或产生一个报表结果，每个函数运行的参数和时间由经营分析程序进行调度和设置。

### 4.2.1 用户数分析功能设计

功能描述：以月份为基础，按照用户类型、客户编码、用户状态、消费层次等属性维度对客户数的构成和发展趋势展开分析，并能对不同月份各个分析属性实际情况进行比较。

业务功能结构图（图 4.8 所示）。

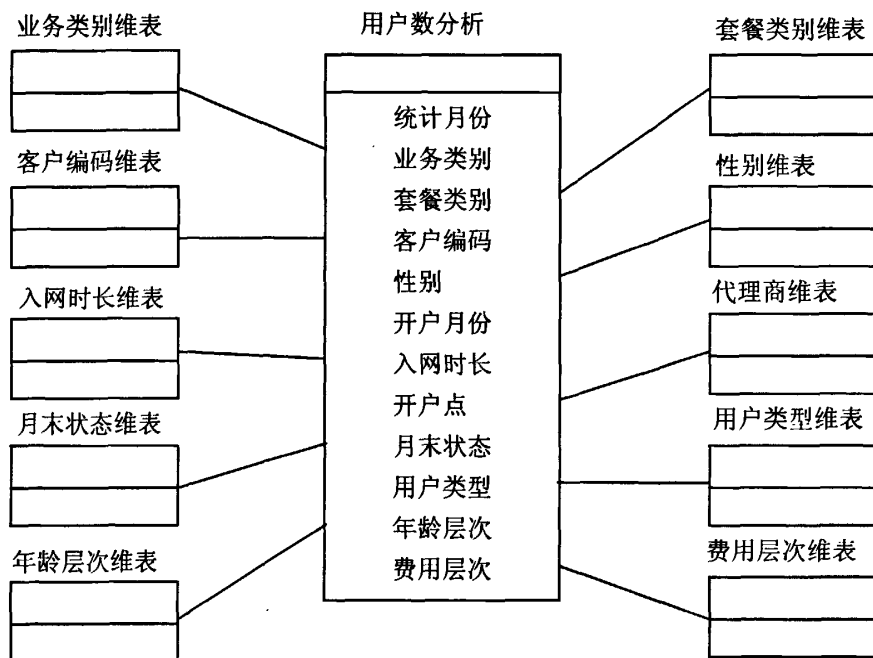


图 4.8 用户数分析功能结构图

Fig. 4.8 Diagram of customer sum analysis function

所需数据源：客户每月历史状态表。

定义函数名：customer\_sum\_analy，具体流程图如图 4.9 所示。

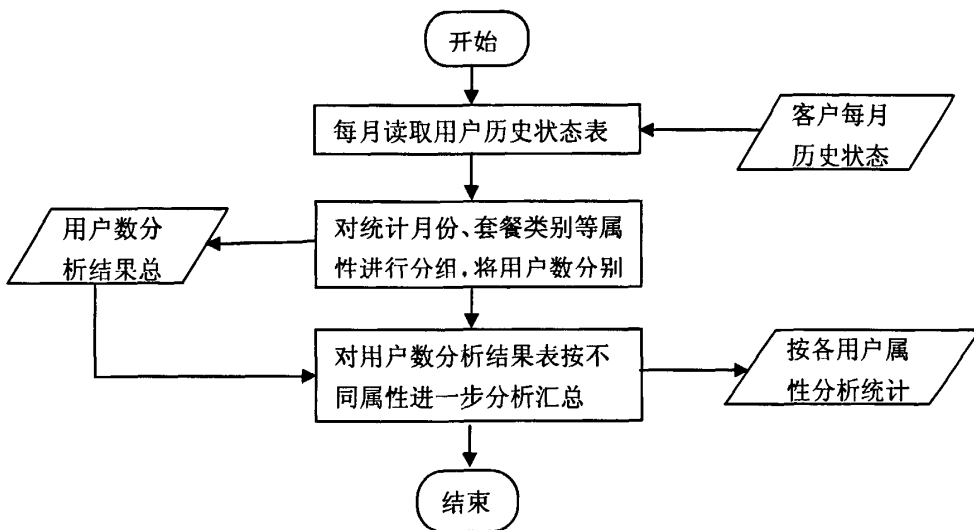


图 4.9 用户数分析模型实现流程图

Fig. 4.9 Implementation flow chart of customer sum analysis model

#### 4.2.2 集团客户分析功能设计

功能描述：主要包括了集团客户构成(总量)分析、集团客户新增分析、集团客户收入分析、集团客户欠费分析四个部分。

其中集团客户构成分析以时间为主要分析条件，对客户总量及构成情况进行分析，集团客户构成分析功能结构见图 4.10。

集团客户新增分析与集团客户构成分析结构基本一致。

集团客户欠费分析主要对集团客户欠费的用户数以及欠费额度情况进行分析，集团客户欠费分析功能结构见图 4.11。

集团客户收入分析主要对国内长途话费合计、国内长途漫游费合计、短信息费用合计等指标进行多角度多层次的分析，集团客户收入分析功能结构见图 4.12。

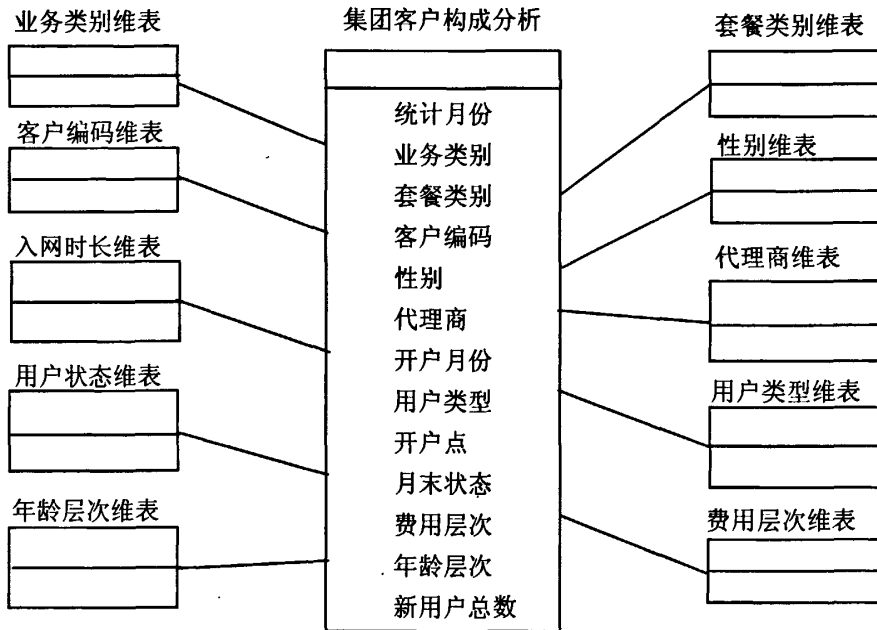


图 4.10 集团客户构成分析功能结构图

Fig. 4.10 Diagram of group customer structure analysis function

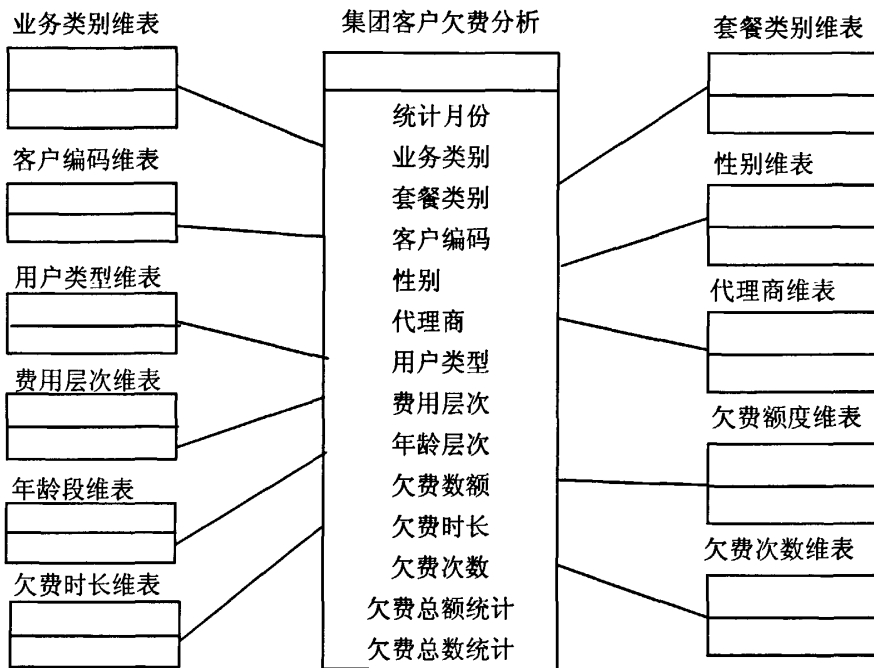


图 4.11 集团客户欠费分析功能结构图

Fig. 4.11 Diagram of group customer overdue charge analysis function

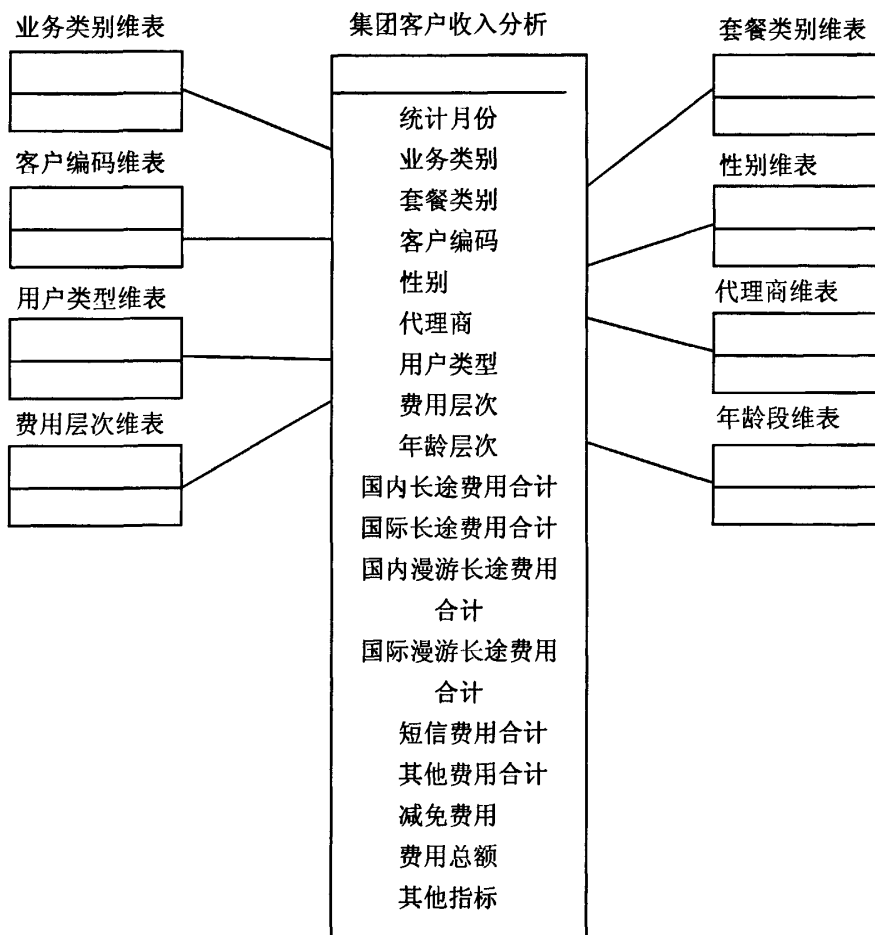


图 4.12 集团客户收入分析功能结构图

Fig. 4.12 Diagram of group customer income analysis function

所需数据源：客户每月历史状态表、客户资料基本表、客户每月帐单表等。

定义函数名：group\_customer\_analy，具体实现流程图如图 4.13 所示。

#### 4.2.3 个人客户分析功能设计

功能描述：个人客户分析是在个人客户资料基本表、客户每月帐单表和客户每月历史状态表基础上对个人客户电话号码情况进行查询，掌握客户状态、缴费、欠费历史情况。

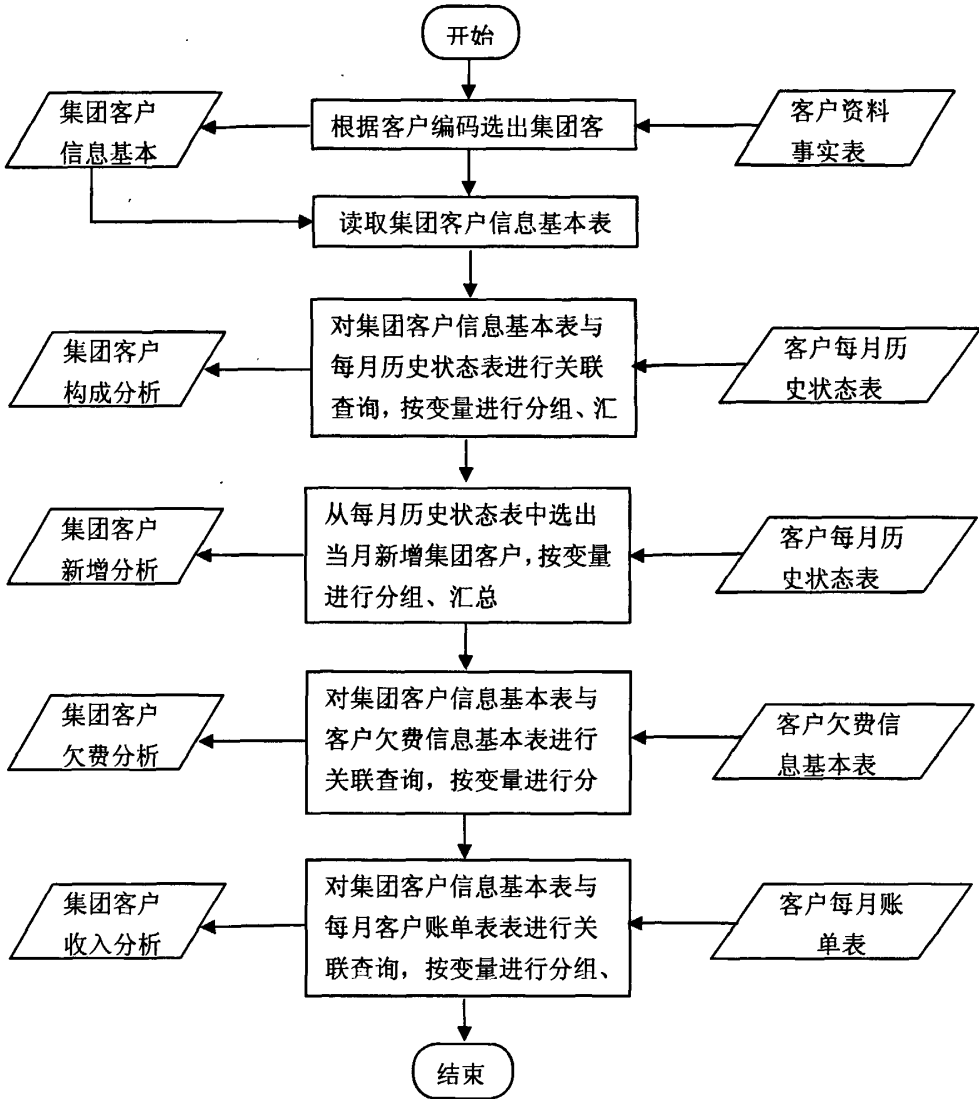


图 4.13 集团客户分析模型实现流程图

Fig. 4.13 Implementation flow chart of group customer analysis model

个人客户分析功能结构与集团客户分析功能结构基本相似，本文不再描述。  
 所需数据源：客户每月历史状态表、客户资料基本表、客户每月帐单表等。  
 定义函数名：single\_customer\_analy。

#### 4.2.4 竞争分析功能设计

功能描述：根据通话详单表信息，分析本企业各类客户群与其他运用商客户群之间通话客户总数的变化情况。主要包括了其他运用商客户总数分析、其他运用商新增客户分析、其他运用商流失客户分析等。

所需数据源：通话详单表。

定义函数名：oppose\_cus\_analy，实现流程图如图 4.14 所示：

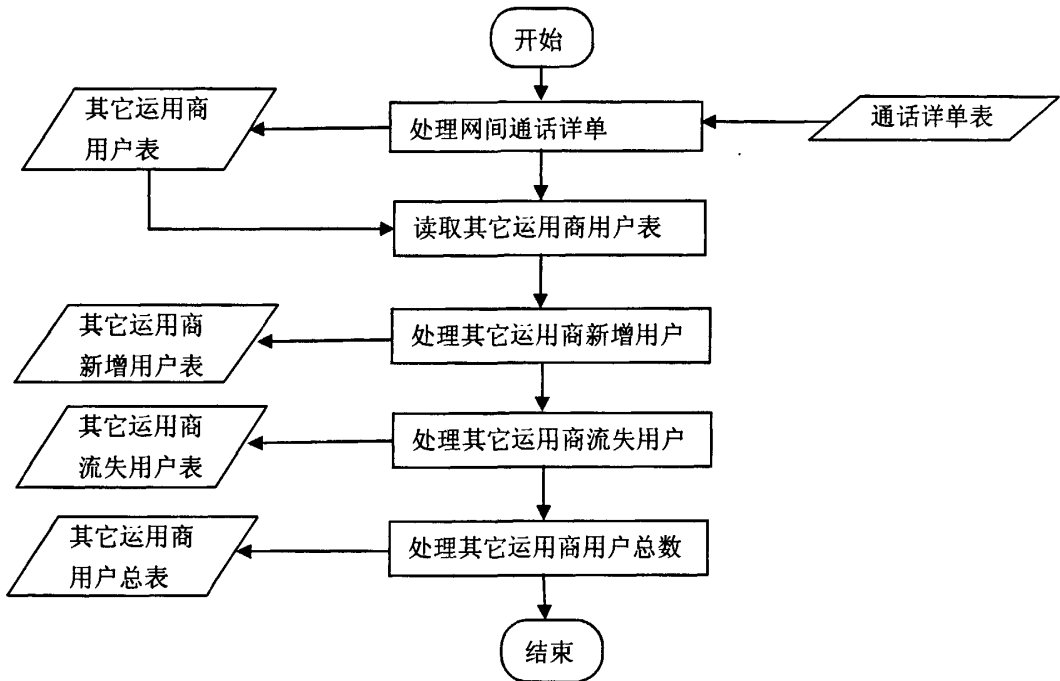


图 4.14 竞争分析模型实现流程图

Fig. 4.14 Implementation flow chart of competitor analysis model

#### 4.2.5 客户流失分析功能设计

功能描述：客户流失主要因素包括拨叫异网(电信、移动)次数、欠费情况、话费突变等，以用户年龄、用户性别等属性去考察，从流失用户的开户点、套餐、等属性去分析客户流失的数量。具体分析功能结构如图 4.15 所示。



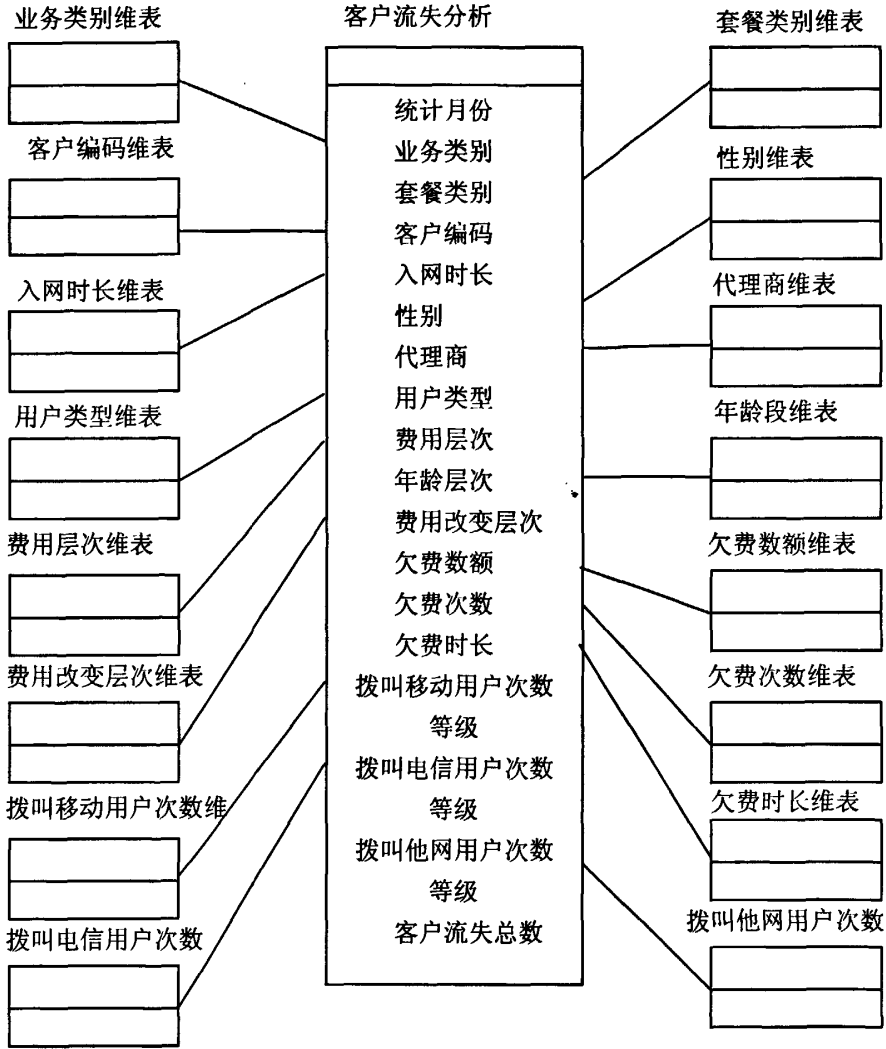


图 4.15 客户流失分析功能结构图

Fig. 4.15 Diagram of customer churn analysis function

所需数据源：客户每月历史状态表、客户资料基本表。

定义函数名：leave\_sum\_analy，具体实现流程如图 4-16 所示。

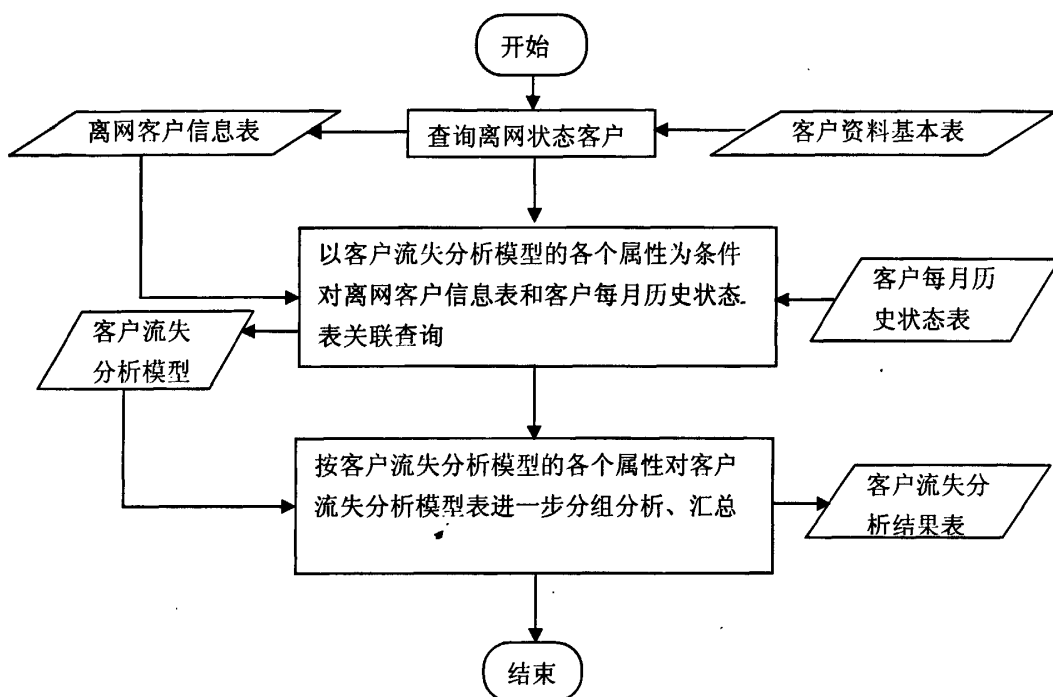


图 4.16 客户流失分析模型实现流程图

Fig. 4.16 Implementation flow chart of customer churn analysis model

#### 4.2.6 收入分析功能设计

功能描述：每月根据套餐类型、业务类别、业务品牌、用户状态、性别等属性对各项通信费用指标进行分析，加强收入监控。具体分析功能结构见图 4.17。

所需数据源：客户帐单表、客户资料基本表。

实现该模型的函数名：income\_sum\_analy，具体实现流程见图 4.18。

### 4.3 应用层子系统功能设计

#### 4.3.1 应用层功能及数据展现方式

经营分析系统应用层子系统的主要功能是使经营分析系统使用人员通过图形和报表的方式，简便、快捷地访问经营分析系统中的各种数据资源并进行各种分析预测的操作。应用层应提供多种数据分析工具，主要包括预定义报表、即席查询、多维动态分析和数据挖掘等。

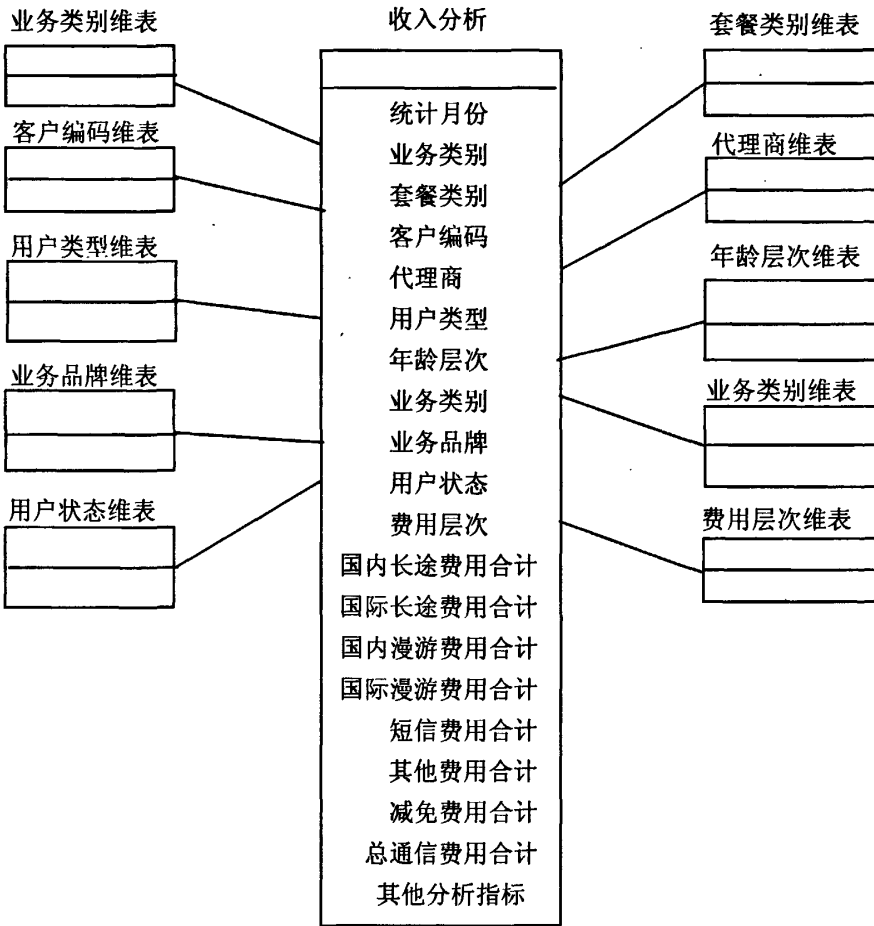


图 4.17 收入分析功能结构图

Fig. 4.17 Diagram of income analysis function

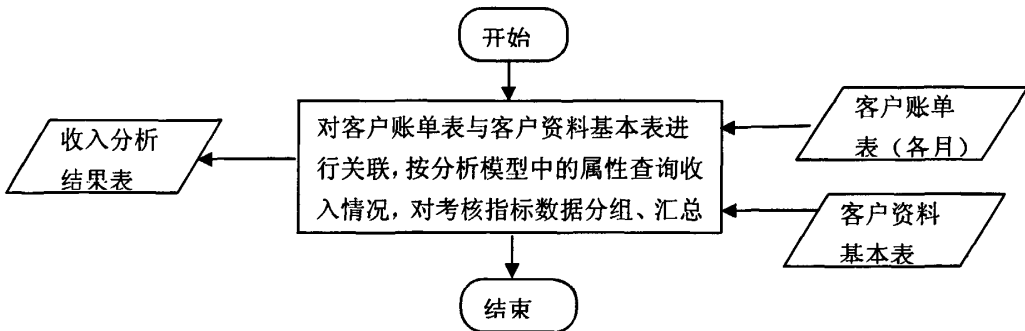


图 4.18 收入分析模型实现流程图

Fig. 4.18 Implementation flow chart of income analysis model

系统应提供一个统一的门户入口,实现预定义报表、即席查询、多维动态分析和数据挖掘等分析工具的紧密连接,并提供集成化的认证、信息发布和管理环境,使经营分析系统使用人员无需关心具体的技术实现途径,即可实现对经营分析系统数据的访问和分析。实际上,对大多数人来说,数据展现就是数据集市。他们不明白也不关心藏在界面之后的时间、功能和资源,他们只想得到答案。此外门户网站还可根据不同分析和决策人员的需求,对所需访问和分析的内容进行方便、简捷的定制,以满足个性化信息服务需求。

对于数据展现,本系统采用的是 Struts + Spring +Hibernate/JDBC 3 层结构,即将表示层、业务逻辑层、数据访问层分离,减少了各个模块间的耦合性,易于开发、调试。采用 Spring 解决了多样的定制属性文件的需要,用一致的配置操作贯穿整个应用,提供数据持久框架,采用 JDBC 或者 Hibernate 的 O/R map-ping 产品实现数据存取。在对后台表进行修改时,可以简单地通过修改映射表来满足要求。模块化清晰,易于调试、复用。随着 WEB 应用和发展,经营分析展现方式也向 WEB 方式发展,由于 WEB 同时具备跨平台、客户端只需浏览器即可方便提供信息展示等特点,因此在本系统中也采用 WEB 行式发布经营分析报表和联机分析结果。

#### 4.3.2 前端分析工具设计

**预定义报表:**经营分析人员在使用系统的过程中,根据需求和有关分析结果进行预定义后、格式内容相对固定的报表展现方式<sup>[24]</sup>。

**即席查询:**经营分析系统使用人员可自由定义内容,允许使用者控制数据的访问方法,并对查询结果提供多种可选择的展现方式。

**多维动态分析:**基于多维数据模型的分析方法,用于支持复杂的分析和预测,包括趋势分析、What-if 分析等。OLAP 是多维数据分析工具的集合,具备上钻下钻的功能,可以满足经营分析人员的所有需求。通过钻取、切片、切面、旋转等操作,经营分析人员可以从不同的角度来全面审视业务发展状况<sup>[25]</sup>。

**数据挖掘:**从海量的数据中提取有用的数据,供决策者使用。在经营分析系统中,数据挖掘主要用于客户流失分析、市场推广分析、客户消费模式分析等。例如为了分析客户流失情况,可以根据已具备的客户流失数据建立服务属性、客户属性和客户消费数据与客户流失可能性关联的数学模型,找出服务属性、客户属性和客户消费数据与客户流失的最终状态的关联关系,并得出明确的数学公式。以后分析中只要具备服务属性、客户属性和客户消费数据,就可以测算客户流失的可能性,以及时地推出挽留措施。

#### 4.4 本章小结

本章主要从数据仓库的数据获取子系统、经营分析子系统、应用层子系统三个方面介绍了经营分析系统的详细设计和实现过程。对数据获取层设计和实现的描述主要包括数据获取层的设计要求、ETL 各模块处理流程及详细设计,重点介绍基本资料读取程序、话费帐单读取程序、原始计费收集程序、通话详单采集程序、计费资料解译程序等模块的流程图。在本章中主要对经营分析子系统的分析主题做出的细化和具体的工作进行了介绍,包括了用户数分析、集团客户分析、个人客户分析、竞争分析、客户流失分析、收入分析模型的功能设计与实现流程;最后本章还对应用层的功能设计和前端分析工具进行了简单的介绍。

## 5 部署实现及界面测试

### 5.1 系统部署实现

#### 5.2.1 系统安装

##### (1) 操作系统的安装

###### ① 软件分区：

SWAP 区：用于内存交换用。

/：用于操作系统本身的安装，大小 20GB。

/boot：用于操作系统本身的安装，大小 800MB。

home/：用于系统的各用户(oracle 用户除外)的登录目录，同时提供个人调测程序、开发程序使用，大小 40GB。

platapp/：用于安装应用系统所需要的各类平台软件和应用软件，大小 20GB；该目录下安装 Apache. Tomcat、实时业务分析系统(OSE)前端软件。

oracle/：用于安装 oracle 数据库软件，大小 20GB。

###### ② 数据文件分区：

CDR/：存放 CDR 文件，150GB。

StatData/：用于存放统计后的文件，150GB；以上两个分区在同一块盘上。

BDR2DB/：存放 BDR 入库用文件，单独一块盘，150GB。

BDR2S/：存放给统计使用的 BDR 文件，单独一块盘，150GB。

StaticRef/：专用于数据统计程序使用的区，单独一块盘，150GB。

###### ③ 数据库分区：

UserData[1-N]/：用户数据表空间，N 块盘。

UserIndex[1-M]/：用户索引表空间，M 块盘。

SysData/：系统数据表空间，大小 20GB。

TmpData/：临时数据表空间，大小 16GB。

UndoData/：数据回滚表空间，大小 20GB。

##### (2) 数据库的安装部署：

创建用户和目录。

设置内核参数。

设置环境变量。

安装 Oracle 9.2.0.4。

安装补丁。

安装完 oracle 数据库后调整数据库参数和在线日志组成员。

运行数据库脚本，创建表空间、用户和建表。

运行 oracle 的存储过程。

### (3) WEB 服务器的部署

#### ① 前提条件：

安装环境：Red Hat Enterprise Linux ES 中文操作系统。

辅助工具：NetTerm4.2 (Telnet 终端软件)、CuteFTP Professional 6.0 (FTP 工具)。

系统安装目录：/platfapp/apache/tomcat5/下。

#### ② 安装 JDK。

#### ③ 安装 WEB 应用程序。

#### ④ 配置系统参数。

### (4) Cognos Powerplay Enterprise Server 的安装

① 安装 NetscapeDirectoryServer；安装 CognosPowerplayEnterpriseServer；安装 CognosPowerplayTransformer。

#### ② 设置虚拟目录。

#### ③ 运行 CognosConfigurationManager，完成配置。

#### ④ 部署 Portal 模版和分析模型。

#### ⑤ 运行建模程序。

#### ⑥ 设定刷新进程的作业调度。

#### (5) 各分析模型的测试和调整

## 5.2.2 系统优化

(1) 采用 Oracle 数据库提供的表内分区技术，通过按照表内某一属性存放的不同值划分不同的存储区域，访问表可以同步检索，可提高查询速度。

(2) 做大量的分析优化，如 SQL 语句的优化，表结构的优化，建立多极索引等，可以大幅提升系统性能。

(3) 使用预先统计的方式创建一些汇总表，将基本表的数据按照不同业务规则和维度层次进行统计，可以在即席查询和报表生成中减少数据统计汇总，提高系统响应速度。

(4) 定期清理历史数据，保证系统运转空间，保证系统稳定

## 5.2 系统部分界面测试

### 5.2.1 系统主界面

登陆后进入系统主界面，如图 5.1 所示，用户通过左边的功能选项进入相应的功能界面，如图 5.2 所示。



图 5.1 系统主界面

Fig. 5.1 Main interface of system

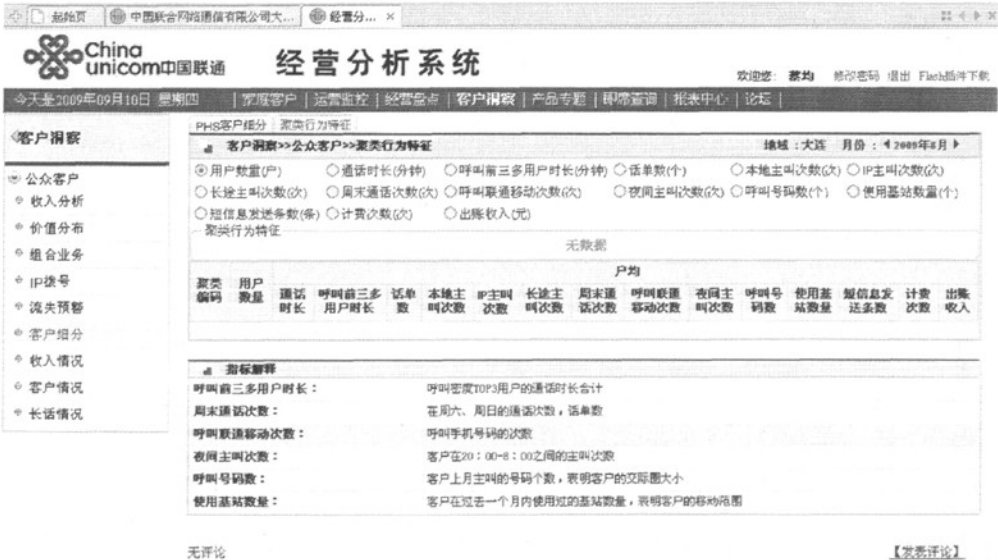


图 5.2 系统功能界面

Fig. 5.2 Function interface of system



导航栏中除按客户分类进行门户分组外，还分别列出运营监控、经营盘点、客户洞察、产品专题、即席查询、报表中心等主要分析工具。

### 5.2.2 用户数分析相关测试

#### (1) 用户数发展指标。

通过单击左上区的指标名称，而左下角的上月与本月的的发展趋势图也会随之变化，例如“固话到达用户”指标如图 5.3 所示，“G 网新增用户”指标如图 5.4 所示。

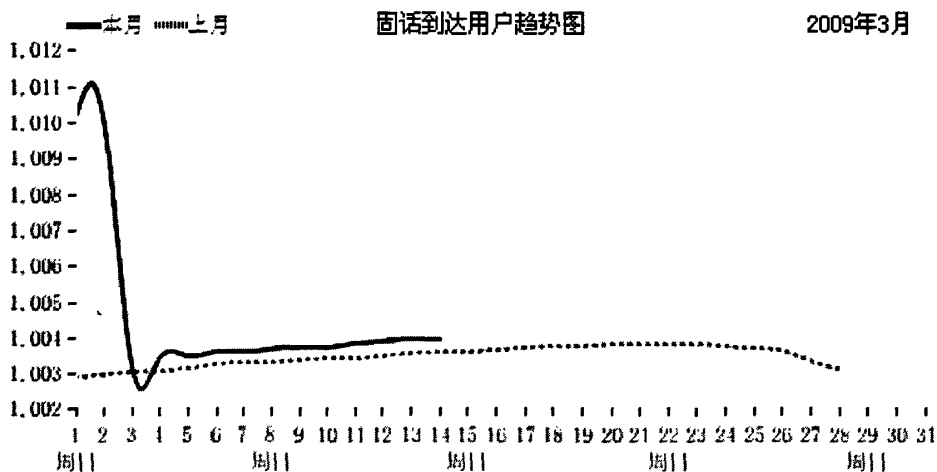


图 5.3 固话用户期末到达指标图

Fig. 5.3 Tendency chart of fixed phone

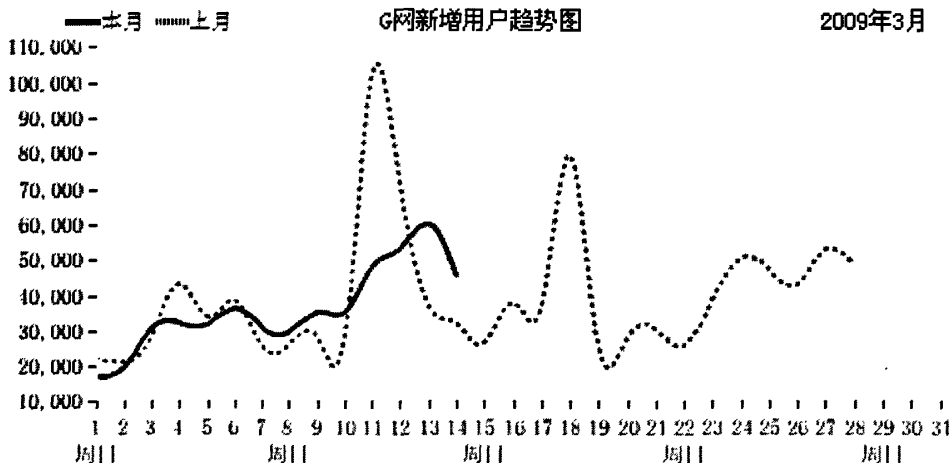


图 5.4 G 网新增用户指标图

Fig. 5.4 Tendency chart of increased customer in G net

### (2) 客户细分分析。

根据客户聚类行为特征对客户细分，如图 5.5 所示。常用的聚类行为特征包括：呼叫前三多用户时长、周末通话次数、呼叫联通移动次数、夜间主叫次数、呼叫号码数、使用基站数量等。

呼叫前三多用户时长：呼叫密度 TOP3 用户的通话时长合计；

周末通话次数：在周六、周日的通话次数，话单数；

呼叫联通移动次数：呼叫手机号码的次数；

夜间主叫次数：客户在 20：00-8：00 之间的主叫次数；

呼叫号码数：客户上月主叫的号码个数，表明客户的交际圈大小；

使用基站数量：客户在过去一个月内使用过的基站数量，表明客户移动范围。



图 5.5 客户细分界面

Fig. 5.5 Interface of customer subdivision

### (3) 客户流失分析界面

反复应用决策树得出用户流失的预测模型，主要因素包括上月欠停用户、资料虚假用户、话费突增用户、话费突减用户等，如图 5.6 所示。

上月欠停用户：指用户资料中本月和上月的用户服务状态为欠费单停、欠费暂停/双停(非语音业务都采用该编码，不采用单停,在 VAS 中表示停机收费)和关联欠费暂停的用户合计；

资料虚假用户：指客户资料中身份证号码验证为非正常的用户合计；

话费突增用户：指本月出帐费用/上月出帐费用 $>3$  的用户合计；

话费突减用户：指本月话费/上月话费 $<1/3$  的用户合计；

流失用户：预测即将流失的用户数量；

损失：即将流失用户如果流失对运营商带来的损失，一般以过去 3 个月的 ARPU 来统计。

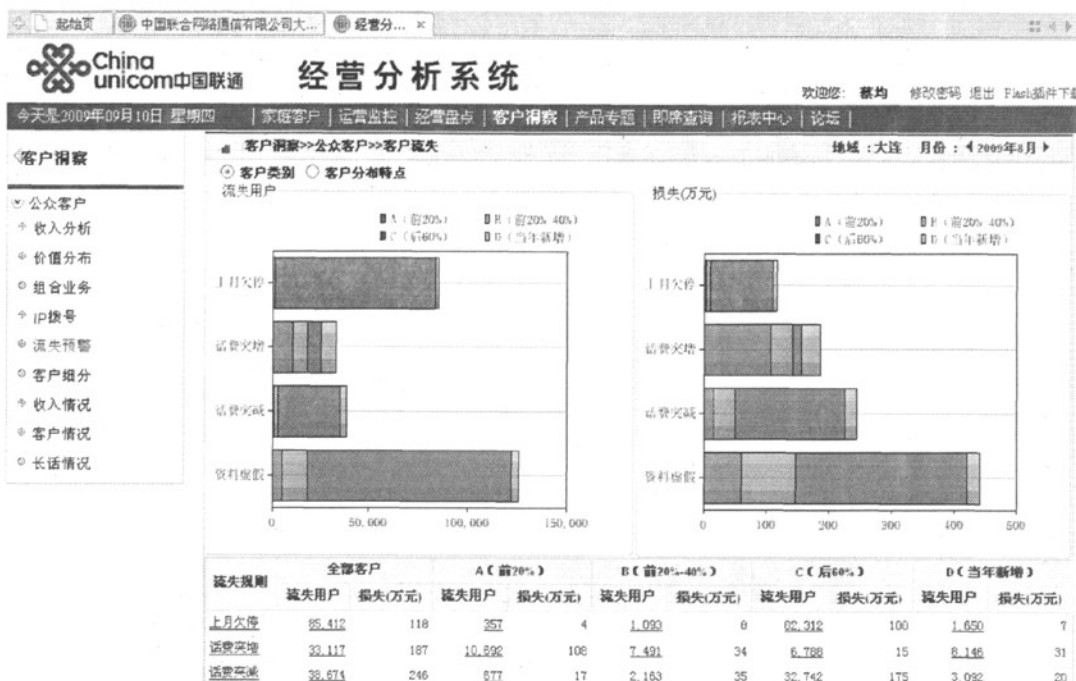


图 5.6 用户流失分析界面

Fig. 5.6 Analysis interface of customer churn

#### (4) 客户流失清单表

点击客户流失分析界面带下划线的数字，展示某类指标的流失清单表，如图 5.7 所示。

流失规则	全部客户		A(前20%)		B(前20%-40%)		C(后60%)		D(当年新增)	
	流失用户	损失(万元)	流失用户	损失(万元)	流失用户	损失(万元)	流失用户	损失(万元)	流失用户	损失(万元)
上月欠费停机用户	237,337	79.22	84	1.31	272	1.78	210,142	55.85	26,839	20.48
沉默用户	43,380	42.59	—	—	—	—	32,704	25.25	10,676	17.34
话费突增	35,047	142.07	3,664	45.13	6,552	28.95	13,047	22.09	11,784	45.90
话费突减	43,552	159.38	96	4.75	378	13.91	35,218	106.98	7,856	33.74
资料虚报用户	403,953	926.42	4,283	67.33	18,302	112.21	379,751	735.91	3,617	10.97

流失清单

■ 公众客户流失预警清单

客户名称	地域	设备号码	业务类型	在网时长	过去3个月ARPU
宋金才	沈阳	81597810	无线市话	27	0.00
陈国峰	沈阳	81390707	无线市话	32	0.00
沈阳市强迅通...	沈阳	81314902	无线市话	13	0.00
朱国华	沈阳市强迅通讯器材经销处	24829259	固定电话	19	0.00
沈阳华茂通...		22937093	互联网	30	2.67
蔡玉迎	沈阳	88033758	固定电话	29	5.33
艾新	沈阳	81252278	无线市话	23	0.00
李欣	沈阳	81942521	无线市话	36	0.00
杜古春	沈阳	81656084	无线市话	36	0.00
吕树元	沈阳	81252098	无线市话	27	3.33
陈守义	沈阳	81366899	无线市话	26	0.00

图 5.7 用户流失清单表界面

Fig. 5.7 Interface of customer churn list

(5) 竞争分析界面

竞争分析是对各运营商发展类指标的一个对比。

点击不同竞争对手的各个发展指标，如图 5.8 所示，左下区会显示与之对应的当前指标本月及上月的发展指标图，如图 5.9 所示。

电信：总用户 203万户 新增 3,971户 手机新增 3,405户 固网新增 566户  
 移动：总用户 2,442万户 新增 45,159户 手机新增 44,504户 固网新增 655户

图 5.8 竞争对手各个发展指标

Fig. 5.8 Development indicators of competitors

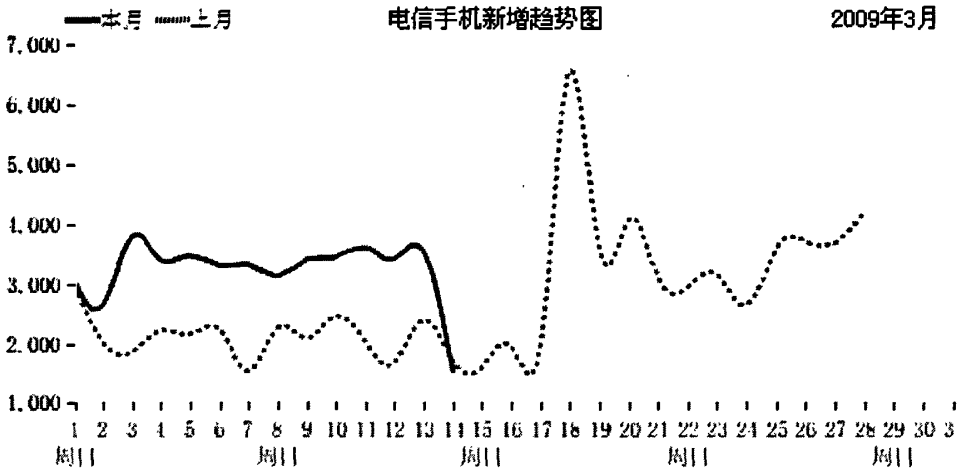


图 5.9 电信手机用户新增指标图

Fig. 5.9 Tendency chart of increased Telecom cell phone customers

### 5.2.3 费用项分析相关测试

#### (1) 收入分析

收入是对当前指标本月的合计值以及其环比及各费用项在收入中的占比情况。分析人员点击【费用项分析】，系统在费用项分析结论性标题下展开空间，此分析只有收入类独有，如图 5.10 为“固话计费收入”指标的费用项分析信息：

费用项分析		固话计费收入(万元)					
国内长途费上升影响较大		本月累计：28,106,380万元 环比：31.40%					
费用项	当月累计收入	环比	用户数	环比	当月累计用户量	环比	
月租费	4,405,039	4.07%	8,697,175	100.00%	0	100.00%	
本地费	35,228,922	3.12%	8,463,131	100.00%	564,438,715	1.71%	
国内长途费	28,106,380	31.40%	3,784,666	100.00%	55,458,655	30.48%	
国际长途费	3,069,359	64.90%	36,085	100.00%	630,489	65.31%	
台港澳长途费	150,122	72.35%	7,962	100.00%	85,221	70.75%	
增值业务费	1,631,322	8.63%	6,751,056	100.00%	0	100.00%	
其他费	-66,000	100.00%	933,296	100.00%	0	100.00%	

图 5.10 固话收入分析界面

Fig. 5.10 Analysis interface of fixed line income

## (2) 客户价值分布

通过此模块可以查看客户价值分布的具体情况，如图 5.11 所示。点击下划线，如客户数中的 1,456，则可展示客户数月趋势图、地域对比图及相应客户明细。

客户数占比：该分档客户数/十个分档客户数之和；

平均在网：该分档客户在网时长之和/该分档客户数；

ARPU：收入/用户数。

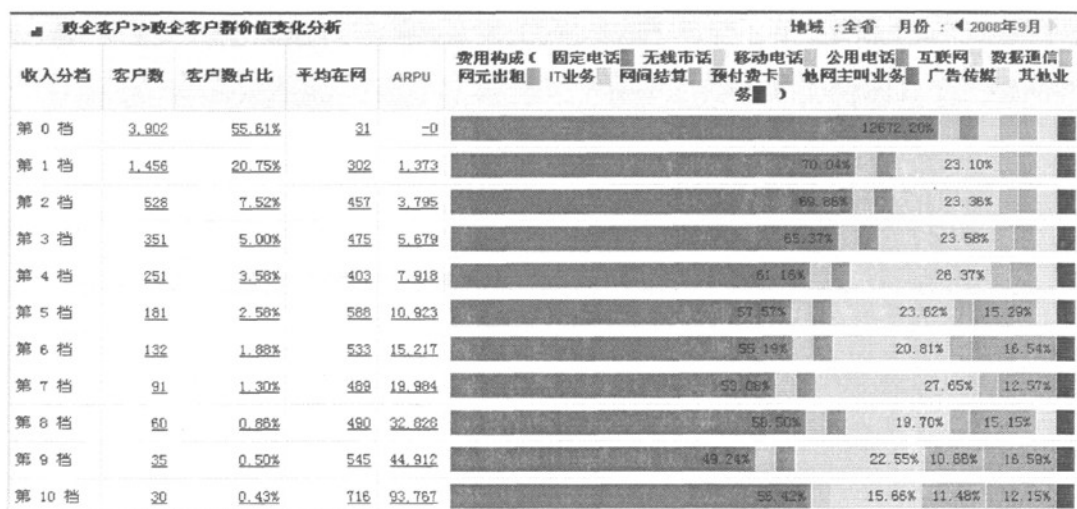


图 5.11 客户价值分布界面

Fig. 5.11 Analysis interface of customer value

## 5.3 本章小结

该经营分析系统已经测试使用一段时间，从目前运行状况看，该系统可以满足企业经营分析的基本要求。在本章中，首先介绍了经营分析系统的安装调试过程，给出系统优化的建议；重点对系统界面测试情况给以描述，例举系统主界面、用户数分析相关测试、费用项分析相关测试的详细说明。

## 结 论

本文将数据仓库技术应用到经营分析系统的设计与实现当中,结合电信企业的经营状况提出了以数据仓库为基础的经营分析系统的主题模型。模型的实现分别按数据获取层子系统、经营分析子系统、应用层子系统三个层次对各主题模型设计与实现进行说明。

为了减少系统在数据获取层中对现有业务系统的影响,该项目根据实际需要设计了一个数据接口区,帐务系统、客户系统、详单等所有业务数据需在接口数据区内进行清洗、转换、整合后装载进数据仓库,大大提高了数据处理的速度和效率。在分析模型设计中,先从各种数据源中获得各种数据基本表,然后在数据基本表的基础上再经过维表整理、加工形成各种分析主题模型,同直接从各种数据源获取数据相比提高了灵活性和扩展性。根据业务主题划分和目前系统实际需求,该论文设计并实现了经营分析主题中用户数分析、集团客户分析、个人客户分析、竞争分析、客户流失分析、收入分析等分析模型。给出了系统组建部署和测试的具体结果。

该系统目前也存在许多不足:数据的集成、清洗和数据纬度的建立还需要进一步研究,数据仓库的建立模型还需要进一步完善,目前各种数据采集还较为繁琐,适应性需要提高。系统的功能还比较单一,不能使用多种挖掘算法对企业数据进行网络资源分析、套餐优化分析等其它主题的应用,不能对数据的相异度进行计算,系统的健壮性还有待改进,这也是下一步的研究方向。

## 参 考 文 献

- [1] 张岩,巫峡. 3G 业务发展趋势及平台建设思考[J]. 邮电设计技术. 2009, 56(5):1-3.
- [2] 马凌,潘伟静. 透视电信运营商的全业务发展布局[J]. 通信管理与技术. 2009, 31(4):28-32.
- [3] 唐磊. 电信本地计费帐务系统开发[D]:(硕士学位论文). 重庆:重庆大学, 2007.
- [4] 周维. 基于数据仓库的管理信息系统的研究[D]:(硕士学位论文). 昆明:昆明理工大学, 2002.
- [5] 谭军朋. 基于下一代网络电信运营支撑系统方案设计[D]:(硕士学位论文). 广州:广东工业大学, 2007.
- [6] 杨卫民,谭骏珊,汪斌. 数据仓库和数据挖掘技术在 DSS 中的应用研究[J]. 计算机工程与设计. 2004, 25(10):1695-1697.
- [7] 高潮. 建立中国电信数据仓库走成功经营管理的道路[J]. 湖北邮电技术. 2001, 15(2):34-37
- [8] INMON W H. Building the data warehouse[M]. NewYork: John Wiley&Sons, 1996.
- [9] 邵贵平. 商务智能系统[J]. 现代计算机. 2003, 20(2):46-48.
- [10] MORIARTY T. What is metadata[J]. Database Programming and Design. 1997, 10(7):57-59.
- [11] RALPH K, M ARRAY R. The data warehouse toolkit: the complete guide to dimensional model[M]. Beijing: Publishing House of Electronics Industry, 2003.
- [12] 何朝红. 数据仓库在我国企业的应用现状及实施策略分析[J]. 广西科学院学报. 2006, 22(4): 375-377.
- [13] 刘丽婷. 试析用数据仓库发展中国消费信贷[J]. 北方经贸. 2002, 22(5):74-75.
- [14] RAHM E, DO H H. Data cleaning:problems and current approaches[J]. IEEE Data Engineering Bulletin. 2000, 23(4):3-13.
- [15] LIU Z, GUO M. A proposal of high performance data mining system[J]. Lecture Notes in Computer Science. 2003, 8(8):145-177.
- [16] 丹尼斯·吉德瑞,陈志刚. OLAP 的构建及其应用研究[J]. 企业技术开发. 2004, 23(4): 12-13+24.
- [17] 肖昭媛. 多维联机数据分析模型和系统设计方法[J]. 上海海运学院学报. 2003, 25(4): 368-372 +377.
- [18] 李波. 数据仓库与联机分析处理(OLAP)技术[J]. 中国传媒大学学报(自然科学版). 2005, 12(04):66-69
- [19] 胡侃,夏绍玮. 基于大型数据仓库的数据挖掘:研究综述[J]. 软件学报. 1998, 9(1): 54-64
- [20] Houtsma M, Swami A. Set-oriented Data Mining in Relational Databases[J]. Data & Knowledge Engineering. 1995, 17(1):245-262.
- [21] 曾华实. 浅谈元数据技术在数据仓库项目中的应用[J]. 信息技术. 2009, 33(1): 79-82.
- [22] 刘耀. 电子商务的营销决策支持系统研究[J]. 深圳大学学报(人文社会科学版). 2002, 19(2):43-49
- [23] 金娟. 数据仓库建设:商业银行应审慎选择[J]. 上海金融. 2004, 25(2):42-44+38.



[24] 张昊. 数据仓库技术及其在企业报表环境中的应用[J]. 当代化工. 2004, 33(6):367-369.

[25] WU L, MILLER L, NILAKANTA S. Design of data warehouse using metadata[J]. Information and Software Technology. 2001, 13(1):109-119.

## 附录 A 通话详单表

表 A.1 通话详单表

Tab.A.1 Calling detail status

字段名	类型描述	说明
YHBH	NUMBER(10)	用户编号
XQ	VARCHAR2(2)	行政区
MONTH	VARCHAR2(6)	月份
DAY	NUMBER(2)	天
SERVICE_TYPE	VARCHAR2(3)	业务类型
SERVICE_CODE	VARCHAR2(4)	业务代码
ROAM_TYPE	VARCHAR2(1)	漫游类型
MSISDN	VARCHAR2(12)	联通用户移动号码
CALL_TYPE	VARCHAR2(2)	呼叫类型
CDR_TYPE	VARCHAR2(1)	本方用户网络类型
HOME_AREA_CODE	VARCHAR2(5)	本方归属地区
CITY_CODE	VARCHAR2(5)	本方归属地城市
USER_TYPE	VARCHAR2(1)	本方用户类型
VISIT_AREA_CODE	VARCHAR2(5)	本方通话地区
MOB_CITY_CODE	VARCHAR2(5)	本方通话地城市
OTHER_PARTY	VARCHAR2(20)	对方号码
OPP_AREA_CODE	VARCHAR2(5)	对方归属地
IS_LOCAL	VARCHAR2(1)	对方位置标志
CALLER_TYPE	VARCHAR2(1)	对方用户网络类型
CALLER_USER_TYPE	VARCHAR2(1)	对方用户类型
CALLED_CODE	VARCHAR2(5)	对方通话地区号
THIRD_PARTY	VARCHAR2(18)	第三方电话号码
THIRD_USER_TYPE	VARCHAR2(1)	第三方用户类型
FLUX	NUMBER(10)	流量
IMSI_NUMBER	VARCHAR2(15)	国际识别码
MSRN	VARCHAR2(11)	动态漫游号
MSC	VARCHAR2(11)	交换机代码
LAC	VARCHAR2(5)	小区代码

表 A.2 通话详单表(续表)  
Tab.A.2 Calling detail status

字段名	类型描述	说明
CELL_ID	VARCHAR2 (5)	基站代码
OTHER_CELL_ID	VARCHAR2 (5)	对方基站
START_DATE	VARC H 人 R2 (8)	呼叫日期
START_TIME	VARCHAR2 (6)	呼叫时间
START_HOUR	VARCHAR2 (2)	开始时段
CALL_DURATION	NUMBER (6)	通话时长
CFEE_TIMES	NUMBER (6)	移动计费次数
CFEEADD_TIMES	NUMBER (6)	移动附加计费次
CFEEADD	NUMBER (6)	移动附加费
LFEE_TIMES	NUMBER (6)	话计费次数
LFEE	NUMBER (8, 2)	长话费
LFEEADD_TIMES	NUMBER (6)	长话附加计费次数
LFEEADD_IN	UMBER (8, 2)	长途附加费
CFEE	NUMBER (8, 2)	移动费
IFEE_TIMES	NUMBER (6)	信息计费次数
IFEE	NUMBER (8, 2)	信息费
JMS_FEE	NUMBER (8, 2)	减免费
ADD_FEE	NUMBER (9, 3)	附加费合计
SUM_FEE	NUMBER (9, 3)	费用总额
TRUNK_GROUPOUT	VARCHAR2 (7)	出中继代码
TRUNK_GROUPIN	VARCHAR2 (7)	入中继代码
OUT_PASS_TYPE	VARCHAR2 (1)	出中继类型
IN_PASS_TYPE	VARCHAR2 (1)	入中继类型
ESN	VARCHAR2 (I1)	电子串号 PHS 才有的
FEE_TYPE	VARCHAR2 (1)	费用类型
STATUS	VARCHAR2 (1)	月末状态
KIND	VARCHAR2 (1)	用户类型
AGE_LEVEL	VARCHAR2 (1)	年龄层次
FEE_LEVEL	VARCHAR2 (2)	费用层次
FEE_CHANGE_LEVEL	VARCHAR2 (1)	话费改变层次
SEX	VARCHAR2 (1)	性别
INNET_TYPE	VARCHAR2 (5)	套餐
INNET_TIME	VARCHAR2 (1)	入网时长

表 A.3 通话详单表(续表)  
Tab.A.3 Calling detail status

字段名	类型描述	说明
INPUT_MONTH	VARCHAR2(6)	月份
AGENT	VARCHAR2(6)	代理商
KHQ	VARCHAR2(6)	集团编码
OPP_INNET_TYPE	VARCHAR2(5)	对方套餐
OPP_KHQ	VARCHAR2(6)	对方集团编码
OPP_AGENT	VARCHAR2(6)	对方代理商
OPP_FEE_LEVEL	VARCHAR2(2)	对方费用层次
ARREAGE_TIME	VARCHAR2(1)	欠费时长
ARREAGE_SUM	VARCHAR2(1)	欠费总数
ARREAGE_COUNT	VARCHAR2(1)	欠费次数

## 攻读硕士学位期间发表学术论文情况

- 1 蔡均. 新一代 BOSS 系统的特点及发展策论分析. 2002 年全国企业管理控制一体化技术研讨会论文集, 2002 年: 304-307. 主办单位: 中国自动化学会, 中国金属学会。
- 2 蔡均. 实现电信普遍服务的模式探讨. 中国通信市场与企业经营管理学术研讨会论文集, 2004 年: 246-248. 主办单位: 中国通信学会通信管理委员会。

## 致 谢

在此论文完成之际，我要衷心地感谢我的导师王爱民教授在我的课题研究以及论文写作过程中所给予的悉心指导和帮助。在研究生学位论文的选题、开题、修改、定稿和论文的形成过程中，王老师倾注了大量心血，我的每一点进步都凝聚了王老师的辛勤劳动。

研究生学业的顺利完成离不开大连理工大学领导与老师的关心与帮助，在这里我向曾经给予过我关心、帮助和教导的各位老师表示衷心的感谢。

最后，感谢我的家人，谢谢你们对我的支持，感谢所有给予我帮助和支持的人，衷心的感谢你们！