

## 摘 要

我国电信业实施大规模重组后,电信市场的竞争越来越激烈。同时,客户对电信服务质量的要求越来越高,保证通信网络的运行质量和提供优质服务,成为电信运营企业能否获得优势地位的重要因素。目前通信网络中各种网元设备类型多,网络结构复杂。这些都给通信网络的运行管理增加了一定的难度。各专业网管数据分散在不同系统当中,难以形成统一运营信息视图,历史数据堆积,但目前没有一个有效的分析工具来对这些海量的数据进行分析,缺乏对运营信息的全面掌控,无法对运营决策提供支持。

数据仓库和联机分析 (On-Line Analytical Processing, OLAP)为解决这一问题提供了有效途径。在数据仓库中建立一个统一的数据模型,然后将各专业网管中的数据经过抽取、转换和加载(Extract, Transform and Load, ETL)进行有效集成,根据业务模型建立多维数据集;运用 OLAP 技术弥补数据仓库对多维数据展示方面的不足,通过切片、切块、钻取等操作对多维数据进行灵活深入的分析和多种方式的数据展示,能够满足多层次用户对数据的需要。

本文是基于黑龙江电信公司综合话务分析系统而来的,现在系统已经通过验收,运行良好,为数据分析和运营决策提供有力支持。

**关键词:** 数据仓库; ETL; 联机分析; 综合话务分析

## ABSTRACT

After large-scale reorganization of China's telecoms industry, six companies were collapsed into three, each spanning mobile, fixed and broadband services. There is more intense competition in telecom market. Meanwhile, the quality requirements of telecommunications services are increasing. Good running of communications networks become the important factor to achieve competitive advantages. Current communication networks include many types of network elements and the network structure is complex, thus the certain difficulty of operation management of communication network increased. Data in different network management systems is accumulating, so it is difficult to form a unified view of operational information. There is no an effective analysis tools to analyze these vast amounts of data. Lack of full control of operational information can not support operate decision-making.

Data warehouse and OLAP technology provide an effective way to solve this problem. In the data warehouse we create a unified data model, and integrate the data in the different network systems through ETL(Extract, Transform, Load) process into data warehouse, then build the cube according to the business model. OLAP technology make up for data warehouse deficiency in multidimensional data display. By slice, dice, drilling and other operations on the multidimensional data, flexible in-depth analysis and various ways of data display can meet the multi-level user data needs.

This paper is based on Comprehensive traffic analysis system in Heilongjiang Branch of China Telecom. now system has been through inspection, running well, and provides forceful support for data analysis and operation decision.

**Key words:** Data warehouse; ETL; OLAP; Comprehensive traffic analysis

## 第1章 绪论

### 1.1 选题的背景和意义

随着信息技术的不断推广和应用,许多企业使用数据库系统处理日常业务,企业内部积累了海量的数据。然而,面对大量复杂的数据,传统的数据库系统,即联机事务处理系统(On-Line Transaction Processing, OLTP)统计分析能力不足,大量数据不能转化成有用的信息指导行业决策,越来越不能满足企业发展的需求。为此,人们提出了为业务的统计分析建立一个数据中心,专门为分析统计和决策支持提供应用服务,通过这个数据中心可满足联机分析应用和决策支持。数据仓库(Data Warehouse)和联机分析(On-Line Analytical Processing, OLAP)技术的共同发展,为解决从海量数据中获得对企业决策有用信息的难题提供了理论支持。最终使得企业可以实现用户对数据多视角、多维度的观察,从而能够灵活而深刻的理解数据,进而提供用户在决策方面的支持,最终真正让数据创造价值。

近年来,全球范围内通信发展迅速,电信市场竞争日益加剧,行业发展面临着新的机遇和挑战。2008年,我国对电信业实施了大规模的重组<sup>[1]</sup>。形成了中国移动、中国联通、中国电信三家电信运营商的鼎足之势。三家运营商分别拥有移动和固网业务,各获得一张3G牌照。因此,三家必将在各业务中展开激烈的竞争。与此同时,用户对电信服务质量的要求越来越高,面对营运成本的大幅度增加,如何能够在满足用户需求和提供优质服务的前提下,充分利用现有条件降低成本、提高效益,成为电信运营企业能否获得竞争优势地位的重要因素之一。

要提供优质服务,就必须保证通信网络的稳定高效运行。尤其现在通信网络发展迅速,用户规模越来越庞大。工业和信息化部公布了截止到2010年10月通信业整体运营状况<sup>[2]</sup>,统计数据显示,我国电话用户累计达到11.3亿户,其中移动电话用户数达到8.3亿户。目前通信网络中含有各种网元设备多,报告种类多,设备类型多,系统版本多,涉及厂家多,网络结构复杂。这些都给通信网络的运行管理增加了一定的难度。改组以后,现有网管系统已经难以满足新形势新业务的需求,主要表现在:

- (1)专业网管历史数据堆积,数据价值等待发掘。
- (2)满足一线运维人员需求的同时,还需满足战术层、战略层等各种角色用户的需求。
- (3)网管数据仍然分散在各个不同的子系统当中,难以形成统一运营信息视图,无法为运营分析提供一致的服务,急需集成共享。

(4)缺乏对运营信息的有效分析和全面掌控,无法有效地提供个性化、差异化的服务。

(5)新的商业模式、新的服务模式不断出现,支持系统在技术和功能上也需要进行补充和完善。

因此需要建立一套综合的话务数据分析系统,充分利用运营商在经营管理过程中积累的丰富经验和宝贵数据,通过这个平台对目前网络管理中大量的数据进行集中管理和分析,帮助运营商解决发展中的问题,以提高资源利用率、提供高质量的服务、降低运维成本、保障主导业务的发展。

## 1.2 国内外研究现状

随着信息技术的不断发展,数据仓库技术(Data Warehouse)和联机分析处理(OLAP)成为新兴的研究领域。美国著名信息工程学家 Willian.H.Inmon 博士提出数据仓库的概念,他认为一个数据仓库通常是一个面向主题的、集成的、随时间变化的、稳定的数据集合,用以支持管理决策过程<sup>[3]</sup>。在 Inmon 首次提出数据仓库概念时,人们并没有给予过多关注。但是,随着信息技术及企业全球化的不断发展,人们开始在实践中尝试和研究数据仓库系统的相关技术了。1998 年在美国举行的第一届《ACM 数据仓库和联机分析国际会议》<sup>[4]</sup>到 2009 年在香港举行的《第十二届 ACM 数据仓库和联机分析国际会议》<sup>[5]</sup>,以及 2004 年在德国举行的《数据仓库展望研讨会》<sup>[6]</sup>,众多专家对数据仓库及面临的主要问题做了广泛而又深入地讨论。

国外许多计算机公司在数据仓库和 OLAP 领域的研究与开发也投入很大,在数据库技术当前及今后的发展里程中,数据仓库及基于此技术的商业智能将是大势所趋,必将成为兵家必争之地。IBM 实验室在这方面已经进行了 10 多年的研究,在 1998 年 9 月发布 DB2UDB 5.2 版,除了用于 OLAP 分析的后台服务器 DB2OLAPServer 外,还提供了一系列的相关产品,并且包括了前端展示工具,形成一整套的解决方案<sup>[7]</sup>。IBM 在一个通用的数据库系统中实现 OLTP 和 OLAP 功能,Informix 也是如此,在其动态服务器 IDS (Informix Dynamic Server)中提供了一系列相关的选件,如高级决策支持选件、OLAP 选件、扩展并行选件等,并且认为这种体系结构严谨,索引机制完善,管理方便,并行处理效率更高;而微软则是从 SQLServer7.0 开始集成了 OLAP 服务器;与上述公司不同的是,Sybase 提供了专门用于 OLAP 功能的服务器 Sybase IQ,并与数据仓库其它相关工具打包成 Warehouse Studio。

数据仓库技术在国外应用已较为普遍,并呈现电子化数据积累程度高、业务应用丰富、业务人员 IT 背景强、并且有较为完善的管理和实施等特点。从目前来看,世界 500



强的企业大多数都在建设或者已经建设完成自己的数据仓库系统，处在世界前列的电信运营企业都建有数据仓库系统。

国外电信运营企业数据仓库的建设起始于20世纪90年代中后期，如AT&T wireless，从1997年开始建设数据仓库，西南贝尔的数据仓库建设始于1994年，当时是最大的数据仓库系统。近几年来，随着中国市场竞争越来越激烈和企业信息化建设的需要，国内的数据仓库的建设也得到了快速的发展，但总体来说，国内数据仓库的建设及应用起步相对较晚，与国外相比还有一定的差距。在中国，中国移动和中国联通都是较早引入了数据仓库系统的电信运营商，中国移动2002年开始建设数据仓库，取得了相当不错的成绩，建成了世界最大的数据仓库系统，而且建设性地提出了“分级式数据仓库理论”，为我国移动通信数据仓库的应用积累了宝贵经验<sup>[8]</sup>。

### 1.3 研究的目的

电信运营企业各专业网管中存在着大量的数据，如何能够将这些数据集成并充分利用，是电信运营企业所关心的重大问题。本文的目的就是将先进的数据仓库技术和OLAP技术，应用到电信的话务网管领域。通过对数据仓库和OLAP技术的学习研究，结合电信重组的实际情况，建立一套基于数据仓库技术和OLAP技术的，能够很好满足运营商需求的话务综合分析系统。通过对现有各专业网管的数据整合，并为电信运营商提供各种强有力的分析手段，从海量数据中发现有用的信息，为决策提供支持；发现网络性能的变化趋势，网络业务量的发展趋势以及潜在的网络故障的预兆等，为建设和容量扩充、网络的维护等工作提供支持。从而帮助电信运营商解决发展中遇到的问题，以提高服务质量和资源利用率，使运营商在重组后日益激烈的竞争环境中保持优势。

### 1.4 论文研究内容及章节安排

在论文期间参加了黑龙江电话话务综合分析系统项目的实施工作，对通信系统、电信网络管理以及数据仓库的相关理论知识有了比较全面的了解和掌握。在此期间主要进行了以下工作：

#### 1、需求调研

通过与黑龙江电信网管中心的沟通交流，了解他们在日常网络运行维护、网络优化以及业务决策时遇到的困难，充分了解其各方面需求，进行相关的需求调研，为建立一套真正实用的电话话务综合分析系统打下坚实的基础。

#### 2、理论学习

认真学习数据库、电信网络管理、数据仓库、OLAP 等相关方面的知识；了解相关技术及最新发展动态，了解国内外电话话务网管数据分析系统的发展状况及应用前景，为论文的顺利进行提供理论保证。

### 3、系统建设

参与亿阳信通承建的黑龙江电话话务综合分析系统建设的全过程，参与系统总体设计，部分承担数据仓库建模和数据的 ETL 设计工作。目前该系统已经成功通过初验，其在一定程度上解决了电信黑龙江分公司日常运维、网络优化和决策支持等方面的相关问题，具有较强的实际应用价值。

在以上工作基础上完成了本论文，论文的章节安排如下：

#### 第 1 章 绪论

本章主要讨论课题的研究背景和意义、国内外的研究现状、研究目的以及论文的主要内容；

#### 第 2 章 数据仓库与 OLAP 技术概述

本章分别介绍了数据仓库和 OLAP 技术的基本概念和基本理论

#### 第 3 章 话务综合分析系统设计

本章重点介绍了系统设计中的几个关键环节，包括系统需求分析，分析专题设计，数据模型设计，以及 ETL 平台的设计。

#### 第 4 章 话务综合分析系统实现

本章首先介绍了系统部署的硬件和软件环境，然后说明了数据采集的实现过程，最后重点描述了分析功能和展现情况，展示了系统建设的成果。

## 第2章 数据仓库与 OLAP 技术概述

### 2.1 数据仓库技术

数据仓库是以计算机应用为基础的信息系统，用来支持决策分析。数据仓库从很多不同的联机事务处理数据源抽取数据，并通过一系列的处理、汇总计算吧数据组织成易于分析的形式，为企业提供一个统一的信息平台，为管理人员和决策者快速提取信息提供有力支持。数据仓库技术是从数据库技术发展而来，为了满足企业对决策支持的需要，并逐步独立的一系列新的应用技术。

#### 2.1.1 数据仓库的定义

数据仓库的概念首先由 W.Inmon 在 1991 年提出，他把数据仓库描述为一个“面向主题的、综合的、稳定的、随时间变化的数据集合，用于支持决策管理”<sup>[9]</sup>。数据仓库有以下特点<sup>[10]</sup>：

##### （1）面向主题的

在操作型数据库中，数据组织面向事务处理任务，各个业务系统之间相互分离，而数据仓库中的数据是按照一定的主题域进行组织存储的。

##### （2）综合的

为了正确决策，需要将所有不同系统中的数据整合到一起。数据仓库中的数据是通过将原有分散的数据库中的数据抽取、清理的基础上经过汇总和整理得到的，必须消除源数据中的不一致性，以保证数据仓库内的信息是关于整个企业的一致性的全局信息。

##### （3）稳定的

数据仓库的数据主要用于企业决策分析之用，涉及的操作主要是数据查询，一旦数据进入数据仓库后，一般都将被长期保留，也就是说数据仓库中会有大量的查询操作，但修改和删除操作很少，通常只需要定期的加载新的数据。

##### （4）反映历史变化

数据仓库的目的决定了它除了包含当前数据之外，还必须包含历史数据，系统记录了企业从过去某一时间点开始到目前的各个阶段的信息，通过对这些信息分析处理，可以对企业的发展历程总结分析以及对未来趋势做出预测。

数据仓库技术是一种数据整合、知识管理的有效方法。数据仓库能够集成来自不同数据源的数据。这意味着可以通过数据仓库以一个统一的视图来呈现不同

系统内的数据，将企业的业务数据整合并转化为适合战略决策的信息，用以支持经营管理中的决策制定过程。数据仓库中的数据是面向主题的，与传统数据库面向应用相对应。

### 2.1.2 数据仓库与传统数据库的区别

数据仓库是在传统数据库基础上发展起来的，但它并不是对传统数据库的彻底抛弃，而是弥补了传统数据库在数据分析方面的不足，提供良好的大规模数据分析能力，为管理决策提供有效的技术支持<sup>[11]</sup>。数据仓库与传统的数据库系统相比，数据库是一种通用平台，建立于严格的数学模型之上，用来管理企业数据，进行事务处理，完成相关业务。而数据仓库没有严格的数据理论，更偏向于工程，它不是花钱就可购买到的成品，而是企业一个日积月累的建立过程，它的应用对象是不同层次的管理者，它的数据源是多种数据源，库中数据很少修改和删除数据，主要是大规模查询和分析，因此要求具备大量的历史数据和汇总数据<sup>[12]</sup>。和传统数据库相比，在总体特征、面向用户、功能目标、数据规模等方面，都有着很大的差异（参见表 2-1）。正是由于这些差异的存在，实现了数据仓库在分析能力的突破<sup>[13]</sup>。

表 2-1 数据仓库与数据库的差异

比较项目	传统数据库	数据仓库
总体特征	高效的事务处理	提供决策支持
面向用户	业务处理人员	决策管理人员
存储内容	以当前数据为主	主要是历史的、存档的归纳数据
功能目标	面向业务操作，注重实时性	面向主题，注意分析功能
使用频率	很高	较低
数据规模	较小	较大
数据访问量	每次事务处理访问量较少	每次分析处理访问大量数据
响应要求	很高的实时性	对实时性要求不高

### 2.1.3 数据仓库的体系结构

在有关数据仓库体系结构的多种理论中，着眼于体系部件功能的“三层结构”理论<sup>[14]</sup>得到了最广泛的接受从各部件的功能分析，数据仓库在逻辑上分为三个层次，即数据获取管理层、数据存储层和数据分析应用层<sup>[15]</sup>（如图 2.1）。

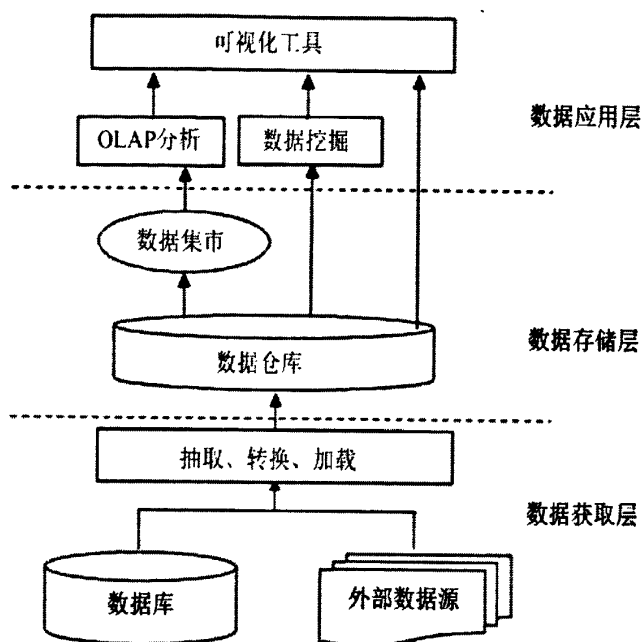


图 2.1 数据仓库的结构

### 1. 数据获取管理层

数据仓库中保存的业务数据来自多个数据源，这些数据源提供的数据并不是理想状态的，存在各种缺陷，必须经过适当的处理后，才能导入到数据仓库；同时数据仓库中的数据也必须进行维护，以保证系统正常运行。归纳起来主要负责以下工作：

（1）数据仓库的定义与修改。定义数据仓库的数据来源、数据的组织方式，并根据决策需要与环境变化对这些内容进行修改，同时对数据仓库的元数据进行管理。

（2）数据获取。根据已制定的规则，定期从数据源抽取数据，并对数据进行清洗、转换和整合，对数据仓库中的内容定期检查，清除失效数据。

（3）数据仓库系统的管理。根据既定的规则和实际情况，对数据仓库进行维护、管理、备份和日志记录等工作。

### 2. 数据存储层

数据存储层是数据仓库的主体，主要有三部分数据：其一是从外部数据源抽取，经清洗、转换处理，并按主体组织存放的业务数据，其二是数据仓库的元数据。其三是针对不同的数据挖掘和分析主题而生成的数据集市。

### 3. 数据分析应用层

数据仓库的数据分析应用层，面向系统的一般用户满足用户的查询、分析的需要，并以适当方式展示查询分析的结果。其主要包括以下功能：

(1) 查询统计服务。为用户提供常规的查询检索、简单统计和报表等服务。

(2) OLAP 服务。对于以多维数据库方式进行数据组织的数据仓库，OLAP 是一种有效的分析方法，通过上卷、下钻、切片、切块和旋转等操作，对数据进行深入分析。

(3) 数据挖掘服务。数据挖掘服务是数据仓库系统的深层次应用，是从大量的数据中，提取隐含其中的、不为人所知的但又十分有用的信息和知识的过程。

#### 2.1.4 数据仓库的建模技术

数据仓库建模是数据仓库创建过程正式开始的第一步。目前，主流的数据仓库逻辑建模技术为维度建模(Dimension Modeling)，包括星型模式和雪花模式等<sup>[16][17]</sup>。实体关系建模在创建联机事务处理(OLTP)系统方面很有效；而维度建模则是创建有针对性的分析模型，成为创建数据仓库典型的逻辑结构<sup>[18]</sup>。

##### (1) 星型模式

星型模式是由一个事实表以及一组维度表组成。每个维度表都通过主键与事实表直接相关联。事实表是数据仓库中的主要数据表，维表则是用来描述事实表中的数据。事实表位于星型中央，而各维度表分布于星型的各个角上。

##### (2) 雪花模式

雪花模式是星型模式的一个扩展，将星型模式的维度表规范化就会得到一个以事实表为中心的雪花模式。在雪花架构中，主维度与事实表相连，其他维度表则连接到主维度表上。雪花结构与星型结构的本质是相同的，主要区别在于雪花模式将维度表进行了规范化处理。

#### 2.1.5 数据集成

随着信息技术的不断发展，企业内部建立了众多的信息系统，但是随着各种信息系统的增加，各自孤立工作的系统就会造成越来越多的数据冗余以及重复劳动。数据集成能够使更多的人员共享数据资源，减少重复劳动，提高工作效率。数据集成是把不同来源、不同格式的数据，在物理上或者逻辑上有机地集中，从而提供综合的数据共享。ETL

(Extract、Transform、Load)是企业数据集成的主要解决方案<sup>[19]</sup>。一个简单的 ETL 体系结构如图 2.2 所示:

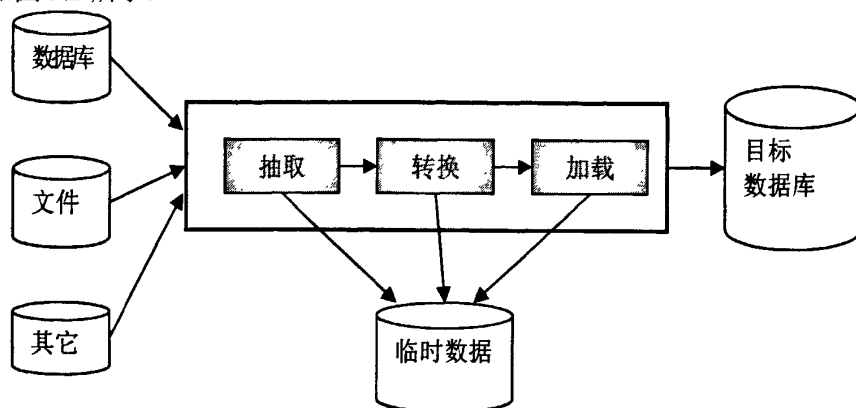


图 2.2 ETL 体系结构

ETL 技术能够按照既定规则集成数据,是负责完成将数据从数据源向数据仓库转化的过程,是创建数据仓库的一个重要步骤<sup>[20]</sup>。在整个数据仓库项目中 ETL 的设计以及实施是工作量最大的。ETL 的关键是变更数据捕获和数据更新<sup>[21]</sup>。ETL 作为构建数据仓库中的一个重要环节,起着承前启后的重要作用,负责将各种异构数据源中的海量数据抽取到临时数据库中进行清洗、转换、集成,最后加载到数据仓库中<sup>[22]</sup>。

### 数据抽取

数据抽取是从业务系统数据库中抽取所需要的数据,经过一系列处理后输入数据仓库。源数据库中的所有数据并不是对于数据仓库中的主题都是有用的,必须根据需要,从源数据库中抽取相关的数据到数据仓库中。数据抽取在工程上要根据数据源的不同而采取不同的抽取方式。从数据库中抽取数据,一般分为以下几种方式:

#### 1. 全量抽取

全量抽取类似数据迁移或数据复制,它将数据源中的所需要的数据原全部从源数据库中抽取出来,并转换成数据仓库中的数据格式。全量抽取的原理比较简单。

#### 2. 增量抽取

增量抽取只抽取源数据库中自上次抽取数据以来新增或修改的数据。在数据抽取过程中,增量抽取相比全量抽取应用也更为广泛。如何能够准确捕获发生变化的数据是进行增量抽取的关键。实际过程中一般有两点要求:不能影响现有业务;能够将业务系统中变化的数据准确地捕获到。增量数据抽取中常用的方式有:

##### (1) 触发器方式

在要抽取数据的表上建立需要的触发器,一般有三种触发器:插入、修改和删除。

当源表中的数据发生变化时，相应的触发器将发生变化的数据存到一个临时表里，然后从临时表中抽取数据，并标记或删除临时表中被抽取过的数据。

优点：加载规则比较简单，不需要修改原系统的表结构，数据抽取的性能高，能够实现数据的递增抽取。

缺点：要建立触发器，对原业务系统会有一定的影响。

### （2）时间戳方式

通过在源表上加入一个时间戳字段，当系统修改或更新数据的时候，同时修改时间戳的值。通过与时间戳的时间对比来决定抽取哪些数据。

优点：时间戳方式的抽取性能也比较好，设计清晰，原理简单，能够实现数据的递增抽取。

缺点：时间戳维护必须由业务系统来完成，对没有时间戳的业务系统有较大的侵入性，对没有自动更新时间戳的数据库系统来说，更新时间戳操作工作量大，改动风险大。

### （3）日志表方式

在原系统中增加系统日志表，当数据发生变化时，更新日志表内容。抽取数据时，通过日志表中的数据来决定加载那些数据。

优点：不需要更改原系统的表结构，数据抽取速度较快。能够实现数据的递增抽取。

缺点：日志表的维护需要由原来系统完成，需要对操作程序作修改，记录日志信息。对原有系统有较大的影响。改动较大，工作量较大，存在一定的风险。

### （4）全表比对方式

抽取所有需要的源数据，并按照相应规则进行数据转换，然后与目标数据逐条对比。通过主键值来判定进行更新与插入操作。如果目标表已存在该主键值，表明该记录已存在，然后进行其余字段的比对，如有不同，进行 Update 操作；如目标表不存在该主键值，则进行 Insert 操作。

优点：对原系统表结构不产生影响，所有操作均由 ETL 完成，管理维护方便，对原系统没有风险。可以实现数据的递增抽取。

缺点：比对过程较复杂，设计较为麻烦，速度较慢。

### （5）全表删除插入方式

每次 ETL 操作均删除目标表数据，全新加载数据。

优点：加载规则简单。

缺点：随着数据的不断增加，数据的更新速度会越来越慢。不能实现数据的递增加载。



### 数据转换

数据转换是指把从不同系统中抽取到的数据，根据数据仓库的要求，进行数据的转换、清洗、拆分、汇总等一系列处理，保证来自不同系统的数据，具有完整性和一致性。

数据转换主要完成由以下原因造成的数据不一致问题：

- (1) 源数据系统与数据仓库系统在数据模型上的差异。
- (2) 源数据结构不一致。
- (3) 源数据定义不规范导致的错误数据。
- (4) 逻辑上关联不一致。
- (5) 不同数据平台命名规范不统一。
- (6) 存在重复数据。

### 数据加载

数据装载就是将从数据源系统中抽取的数据经过转换、清洗后，装载到数据仓库系统中。数据的装载策略根据数据的抽取策略以及业务规则确定的，一般有三种类型：

- (1) 直接追加

每次装载时直接将数据追加到目标表中。

- (2) 全部覆盖

如果抽取数据本身已经包括数据当前和所有历史状况，可对目标表采用全部覆盖的方式。

- (3) 更新追加

对于需要连续记录业务的状态变化，并用当前最新状态同历史状态数据进行对比的情况，可采用更新追加的方式。

在系统实施时，具体采用何种方式，要综合考虑效率和业务实现等因素。

## 2.2 联机分析 (OLAP) 技术

数据仓库创建以后，企业的决策者和分析员就开始对数据仓库进行一些复杂的数据查询，这些查询必须是多角度的、能够下钻上卷的、可旋转的，才能对决策和分析提供强有力的支持。只靠拥有海量数据的数据仓库是不够的，必须通过一定的方式把所要查询的数据丰富多彩的展示出来。因此，一种新的基于多维分析的联机分析处理 (On-Line Analytical Processing, OALP) 技术应运而生。

### 2.2.1 联机分析的定义

1993 年, 关系数据库之父 E.F.Codd 最早提出了联机分析处理的概念<sup>[23]</sup>。Codd 认为传统的联机事务处理 (OLTP) 无法满足用户对数据库复杂查询分析的要求, 通过 SQL 语言对大数据库的简单查询, 不能满足用户分析的需求。用户的决策分析需要进行大量的统计计算才能得到结果, 而查询结果也不能满足分析决策的需求。因此, Codd 提出了 OLAP 的概念。

OLAP 委员会对联机分析处理的定义为: 使分析人员、管理人员或执行人员能够从多种角度对从原始数据中转化出来的、能够真正为用户所理解的、并真实反映企业维特性的信息进行快速、一致、交互地存取, 从而获得对数据的更深入了解的一类软件技术<sup>[24]</sup>。OLAP 的目标是为了满足决策支持, 实现多维环境中特定的查询。其技术核心是“维”的概念, 因此也可以说是多维数据分析工具的集合。

联机分析面向的主要用户是企业中的管理决策者及专业分析人员, 他们在分析数据时, 很自然的会从不同的角度来观察各项业务指标。联机分析处理就是仿照用户的多角度思考模式, 预先为用户建立多维的数据模型。在这里“维”指的是用户观察数据的角度。多维数据模型一旦建立完成, 用户就可以从各个不同的角度快速地获取数据, 也能在不同观察角度之间切换, 或者进行多角度的综合分析<sup>[25]</sup>。因此具有很强的分析灵活性。这也是在近年来联机分析处理被广泛关注的根本原因。

### 2.2.2 联机分析的主要特性

随着人们对 OLAP 理解的深入, 逐渐概括出了几个基本特性, 认为 OLAP 应该具有共享多维信息的快速分析 FASMI(Fast Analysis of Shared Multidimensional Information)的特点<sup>[26]</sup>, FASMI 已经成为设计人员用来判断一个 OLAP 设计的重要准则。

#### 1、快速性

用户对联机分析的反应能力有很高的要求。系统应能够在 5 秒钟内对用户的大部分查询要求做出响应。

#### 2、可分析性

联机分析系统能够处理与应用有关的任何统计分析和逻辑分析。用户无需编程即可实现新的专门计算, 并能够以用户满意的方式给出报告。

#### 3.可视性

联机分析系统应能提供可视化的界面, 能够把信息直观简洁的展示给用户。

#### 4、多维性

多维性是联机分析的关键属性。系统必须提供对数据的多维分析。实际上，多维分析是分析数据最有效的方法，也是联机分析的灵魂。

#### 5、信息性

不论数据量有多大以及数据存储在哪里，联机分析系统应能够及时地获得信息，并且管理大量信息。

联机分析的基本思想是使企业的决策者能灵活地操纵数据，以多维的形式来观察企业的状态并了解企业的发展趋势。利用 OLAP 分析，企业的分析人员、管理人员和执行人员能够从多个角度对某一主题进行分析，从而快速、交互地得出分析结论。

### 2.2.3 联机分析中的基本概念和操作

#### 1. OLAP 概念体系中的基本概念

联机分析的概念体系的基本概念主要有度量、维度、维度的层次、维度的成员、多维数据集等<sup>[27]</sup>。

##### (1) 度量(measure)

度量是用户所要观察和分析的指标数据，典型的度量有产量、成本、销量等。在多维数据集中存在很多度量值。多维数据集是由其所包含的度量和维度来共同描述的。

##### (2) 维度 (Dimension)

维度是指人们在观察数据时的特定角度，是用于描述多维数据集中度量值一组属性。同一个问题，可以通过不同维度进行观察和分析。

##### (3) 维度的层次 (Level)

我们从某个特定角度来观察数据，这一角度可以存在不同的细节程度（如时间维：日、月、季度、年），这就是维的层次。当人们通过维度的不同细节程度（即维层次）去观察数据，可以得到多种不同描述方法。

##### (4) 维度的成员 (Member)

维度成员是指某个维度上的某个具体取值。如果维度具有多个层次，维度成员则由在该维度各层次上的取值组合而成的。如“2009年6月5日”是时间维的一个的维度成员。

(5) 多维数据集 (Cube)

多维数据集是联机分析中的重要对象，多维数据集通过其度量值和维度定义。多维数据集中的度量值和维度派生自数据源视图中的表和视图，数据源视图是多维数据集基于的逻辑视图。

2. OLAP 多维分析中基本操作

OLAP 多维分析包括切片、切块、钻取、旋转等基本操作手段<sup>[28]</sup>，能够对以多维形式组织的数据进行深入观察，从而使用户达到从多个角度、多个细节分析数据的目的。

(1) 切片 (Slice)

对于三个维度的多维数据，切片操作是指在某一维度选定一组成员，取出其余两个维度所构成的平面的过程。而对于大于三维的多维数据，切片是指选定两个维度，而将其它维度都取定一个维度成员后，所得到的在原多维数据在选定的这两个维度上的二维子集。既然维度可理解为观察角度，那么切片作用也可理解为在对数据进行观察时，将注意力集中在某两个角度上，而舍弃其他角度<sup>[29]</sup>。通过切片的操作可以降低多维数据集的维度，将注意力集中在较少的维度上进行观察数据，使人们能够更好地了解多维数据集。如图 2.3 为 CDMA 网络性能指标中系统接通率的切片示意图。

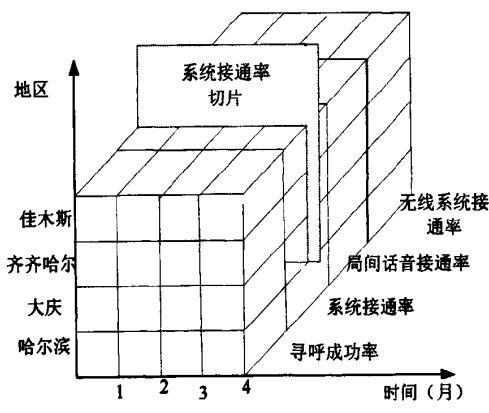


图 2.3 C 网性能指标系统接通率切片示意图

(2) 切块 (Dice)

切块操作就是从多维数据中，取出一个包含三个维度的数据立方的过程。也就是只从选定的三个维度来观察数据，而不考虑其它维度。

### (3) 钻取(Drill)

钻取是指改变维度的层次,变换分析的粒度。包括向向上钻取(Drill-up)和下钻取(Drill-down)。上钻是在某一维上将较低层次的细节数据聚集到高层次的汇总数据;而下钻则相反,从汇总数据深入到细节数据来进行观察。

### (4) 数据旋转

旋转是指改变展示数据时维度的位置关系,使用户可以从不同的视角来观察多维数据。例如,交换行维度和列维度的位置,或是把某一个行维度移到列维度中去。

## 2.2.4 联机分析系统的分类

现代联机分析系统一般以数据仓库为基础,从数据仓库中抽取出数据的一个子集,经过必要的聚集存储到联机分析数据库中,以供前端分析工具读取。联机分析系统按照其数据存储模式的不同可以分为 Relational OLAP(简称 ROLAP)、Multidimensional OLAP(简称 MOLAP)和 Hybrid OLAP(简称 HOLAP)三种类型[30]。

### (1) ROLAP

ROLAP 是基于关系数据库实现的,将分析用的多维数据以二维表格的形式存储在关系数据库中。对用户可能的查询,优先利用事先计算好的实视图来产生查询结果,以提高查询效率。同时现有关系型数据库也对 OLAP 作相应的优化,比如并行查询、并行数据管理、并行存储、位图索引和 SQL 的 OLAP 扩展等。ROLAP 的特点是细节数据、聚合后的数据都保存在关系型的数据库中。这种方式查询效率最低,但是如果数据量比较小的话,用这种方法未尝不可。

### (2) MOLAP

MOLAP 是基于多维数据组织的 OLAP 实现形式。将 OLAP 分析所用到的多维数据物理上存储为多维数组的形式,多维数据将形成“立方块 Cube”的结构。在 MOLAP 中,对“立方块”的切片、切块、旋转是分析多维数据的重要手段。由于 MOLAP 采用的是新的存储结构,物理层存储多维数据,又称为物理 OLAP;而 ROLAP 主要通过一些中间软件实现,数据仍存储在关系数据库中,因此也称为虚拟 OLAP。

### (3) HOLAP

HOLAP 结合了 MOLAP 和 ROLAP 各自的优点,将细节数据保留在关系型数据库的中,但是聚合后的数据保存在多维数据集中。目前, HOLAP 还没有一个正式

的定义，但很明显 HOLAP 结构不应该是 MOLAP 与 ROLAP 结构的简单组合，而是这两种结构技术优点的有机结合，能满足用户各种复杂的分析请求。

## 2.3 本章小结

本章主要介绍了数据仓库和 OLAP 技术的基本理论，对数据仓库的定义，体系结构，建模技术以及 OLAP 技术的定义、特点、基本操作和分类等作了简要介绍。

## 第3章 综合分析系统的设计

电话务综合分析系统是构架于电信各个专业网管之上的综合业务分析系统，主要从各个网管系统中采集相关业务数据，创建数据仓库，建立 OLAP 多维分析模型，从各种角度、多层次地对 C 网、固网和信令等关键性能指标进行深入分析，找出各专题关键性能指标的变化规律和趋势，从而发现网络运行存在的问题；分析不同业务对网络的占用情况，统筹分配网络资源；分析通信网络的运行状况，对各网元的性能指标、处理能力、系统容量及系统故障等进行观察分析，为全网的平稳运行提供保障。

考虑到系统应该具有开放性和可扩展性，本系统采用前端数据采集层、SQL SERVER 数据仓库和用户终端三层体系结构<sup>[31]</sup>。前端数据采集层将各专业网管系统数据库中原始的业务数据，进行 ETL 处理后存放于数据仓库中，在数据仓库中建立多维数据集，用户可以通过 Web 浏览器访问多维数据集，完成所需的各种 OLAP 分析操作。

综合话务分析系统采用先进数据仓库和 OLAP 技术，能够满足多层次用户灵活的多视角的分析需求，能辅助领导层进行网络运行维护决策，提高网络运行的管理水平，提升网络运行质量，提高客户满意度。

### 3.1 收集和分析业务需求

随着电信各专业网管系统的不断建设和深入应用，逐渐积累了大量宝贵的网络运行状况数据和性能数据，但目前没有一个有效的分析工具来对这些海量的数据进行分析，没有充分发挥出历史数据的真正作用；同时，移动通信网络发展迅速，网络结构复杂，网络的建设及维护人员也需要一个有效的分析工具，来分析和了解网络运行状况<sup>[32]</sup>。而网络的运行状况是通过网络系统中各种设备的运行数据来反映的，要加强网络的管理就必须利用有效的方法或工具，来全面而准确的分析网络运行状况的各项数据，发现隐含在这些数据中潜在的网络问题，只有对网络运行状况的充分了解，才能合理配置网络资源，让有限的资源发挥最大的作用。

根据电信网络管理方面所遇到的新问题，收集和分析业务需求，明确建设的目的和要求。中国电信黑龙江分公司，现阶段需要接入的网管系统主要有固定电话网管系统，信令网管系统和 C 网网管系统。因此，数据仓库的数据源主要来自这三个系统的数据库。目前网管系统主要存在以下问题：

- 1) 话务数据保存时间短，无法获得较长时间的历史数据

2) 报表的表现形式单一。不支持图形方式, 不能多角度展示数据。

3) 网管数据分散于不同的子系统当中, 难以形成统一的网络运营信息视图, 急需将数据集成共享。

4) 统计、分析功能薄弱, 无法支持上层决策。

5) 使用操作不够简便友好。用户要接受一定的培训外, 还应该具备一定的数据库专业知识。

因此, 迫切要求一个统一的平台来集中和管理各系统的运行数据, 打破各专业系统数据彼此孤立, 与业务体系衔接不紧密的状态, 以日常业务工作与网络运营为服务目标, 是面向多业务、多层次用户的综合性分析系统, 实现了以全局角度进行业务分析与决策支撑。综合分析系统为各个业务层面的用户提供其所需求的数据分析应用, 不同层面的用户关心的数据专题内容不同、数据分析粒度不同, 应用目的与应用方式也存在差异, 综合分析系统要实现面向用户的贴身式服务<sup>[33]</sup>。

经过与省网管中心以及黑龙江电信各级领导的多方交流, 最后形成《黑龙江电信网络业务综合分析系统技术建议书》, 提供给网管中心, 再经过他们的进一步反馈, 在开发中不断完善功能, 满足新的要求。

## 3.2 业务分析模型的构建

从宏观网络运营分析 C 网、固网重要 KPI 指标, 便于用户全方位查看 KPI 指标的波动情况和对比情况。综合话务分析系统的建设是以 OLAP 分析为主, 通过建立灵活多样的业务模型分析, 深入了解网络的运行状况和存在问题, 为网络分析和经营决策提供有力支持。

### 3.2.1 专题分析

专题分析是针对某一特定主题, 如运营考核指标, 话务量、网络负荷等以不同的维度和视角进行整体分析。专题分析是基于网络能力、网络运行、服务质量等核心应用, 运用趋势分析、比较分析、分布分析等多种分析方法来进行综合的统计分析。OLAP 分析能够分析很多的内容, 包括很多的业务分析专题。通过这些分析专题, 可以对网络运营状况提供全面综合的技术指导。

根据用户所关心的内容, 设计各种分析专题。按照业务分为 CDMA 网络专题、固网业务专题和信令监测 CDR 业务专题三大专题。每个大专题又细分出很多小专题, 来详细分析各项指标。



## 1. CDMA 网络业务专题

C 网络专题主要包括考核指标专题、CDMA 话务量专题、交换负荷和资源专题、无线负荷和资源专题、系统性能指标专题、与用户行为有关的指标专题、与接续无关的指标专题。这 7 个分析专题视图是用户常用的分析方法和分析模型的固化。

各专题内容如下：

### 考核指标专题

业务信道负载率分析，系统接通率分析，无线系统接通率分析，寻呼成功率分析，寻呼成功率分析，呼叫建立成功率分析，业务信道掉话率分析，话务掉话比分析，坏小区比例分析。

### 话务量分析专题

A 接口话务量分析，业务信道承载话务量分析，WALSH 码承载话务量分析，第一载频 280 话务量分析，第二载频 201 话务量分析，第三载频 242 话务量分析，第四载频 160 话务量分析，无线容量分析。

### 交换负荷和资源专题

交换机 BHCA 负荷比分析，VLR 利用率分析，HLR 利用率分析，交换机 2M 端口使用率分析，HLR 用户的每户 BHCA 值分析，HLR 用户的每户 BHCA 值分析，HLR 用户每户话务量分析，平均每次呼叫占用时长分析，主叫比例分析，HLR 生成用户开机率分析。

### 无线资源和负荷分析专题

业务信道负载率分析，软切换因子分析，信道单元可用率分析，业务信道拥塞率分析，超忙小区比例分析，忙小区比例分析，超闲小区比例分析，严重溢出基站比例分析，溢出基站比例分析。

### 系统性能指标分析专题

系统接通率分析，交换系统接通率分析，无线系统接通率分析，局间系统接通率分析，局间话音接通率分析，主叫业务信道分配成功率分析，被叫业务信道分配成功率分析，业务信道分配成功率分析，寻呼成功率分析，呼叫建立成功率分析。

### 与用户行为有关的指标分析专题

话音接通率分析，来话接通率分析，用户行为分析，用户早释比例分析，振铃早释比例分析，被叫用户忙比例分析，拨打号码为空比例分析，被叫用户拒接分析，被叫用户久叫不应分析，被叫用户关机分析，被叫用户服务受限分析。

### 与接续无关的指标分析专题

位置更新成功率分析, 基站系统硬件切换成功率分析, 系统软切换成功率分析, 业务信道掉话率分析, 话务掉话比分析, 坏小区比例分析, 短消息 MO 交换机转发成功率分析, 短消息 MT 发送成功率分析。

### 2. 固网业务专题

网络接通率分析, 来话话务量分析, 去话话务量分析, 目的码话务量分析, 交换机 CPU 负荷分析。

### 3. 信令监测 CDR 业务专题

网间响应时延分析, IP 电话分析, 应答不计费、计费无指示分析, 主叫超短超频、被叫超短超频分析, 网间话务流量流向分析, 中继利用率分析。

专题分析是综合各专业网管系统、各部门运营数据, 针对网络运营过程中的热点问题, 通过各种的分析方法, 加工形成专题报告, 提供有助于解决这些问题的有用信息。通过这些专题分析, 可以为不同部门的人员提供所需的数据, 从而满足不同层次、不同部门的数据需求。

## 3.2.2 网络概况分析

由于将各网管数据进行了集中管理, 可以对公司的整个通信网络概况有更全面的了解。为管理者和分析人员提供统一的信息视图, 通过监控和查看重要的网络业务负荷和网络性能指标来实现对整个网络状况进行监控、观察和分析。可分为网络发展、网络质量和网络负荷三部分。

### 1. 网络发展

通过 TCH 话务量、A 接口话务量、每户话务量、VLR 开机用户数、VLR 登记用户数、HLR 生成用户数、HLR 开机率、平均呼叫时长、来话占用话务量、来话有效话务量、去话占用话务量、去话有效话务量、占用话务量、有效话务量等指标来了解网络的发展现状和发展能力, 对网络的发展方向提供支持。

### 2. 网络质量

通过系统接通率、交换系统接通率、无线系统接通率、寻呼成功率、话务掉话比、TCH 掉话率、TCH 分配成功率、呼叫建立成功率、网络接通率、溢出百分比、应答试呼比、应答占用比、电路闭塞率、电路闭塞率等指标, 能够对全网络的通信质量有一个全面的了解, 及早发现网络存在的问题, 保障通信的顺畅稳定。

### 3. 网络负荷

通过交换机 CP 负荷、VLR 利用率、HLR 利用率、超忙小区比例、TCH 负载率、每线话务量、每线话务量等指标，可以对网络的运行负荷有全面的了解，从而对负荷重的网元进行重点维护或者增加网络承载能力，确保网络安全运行。

3.3 数据仓库的构建

3.3.1 数据仓库主题的确定

主题是指用户使用数据仓库进行决策时所关心的重点方面，是在一个较高的管理层次上对数据进行综合归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，这些集合对分析对象作了比较完整一致的描述，这种描述不仅涉及到数据本身，而且涉及到数据之间的关系。

本系统初期数据来源于三个专业网管，因此可按业务分为 C 网主题、固网主题和 CDR 信令主题。但每个主题下还可以再细分主题来组织数据，以 C 网话务网管数据为例，根据网元可再分为无线与交换主题、HLR 主题和中继群主题等。无线主题是无线与收发基站有关的数据，交换主题是有关移动交换中心的数据。HLR 则是 HLR 网元的数据。

3.3.2 构建数据模型

表 3-1 HLR 主题中各项指标

指标名称	指标算法	备注
HLR 生成用户数		HLR 粒度
HLR 容量		HLR 粒度
网上发展用户数		HLR 粒度
HLR 利用率	$\text{HLR 生成用户数} / \text{HLR 容量} * 100\%$	地区粒度
HLR 用户的每户 BHCA 值	$\text{交换机 BHCA 值} / \text{HLR 生成用户数}$	地区粒度
HLR 用户的每户语音 BHCA 值	$\text{交换机语音 BHCA 值} / \text{HLR 生成用户数}$	地区粒度
HLR 用户每户话务量	$\text{A 接口话务量} / \text{HLR 生成用户数}$	地区粒度
HLR 生成用户的开机率	$\text{VLR 开机用户数} / \text{HLR 生成用户数} * 100\%$	地区粒度
HLR 利用率	$\text{网上发展用户数} / \text{HLR 容量} * 100\%$	地区粒度
每户 BHCA 值	$\text{交换机 BHCA 值} / \text{网上发展用户数}$	地区粒度
每户语音 BHCA 值	$\text{交换机语音 BHCA 值} / \text{网上发展用户数}$	地区粒度
每户话务量	$\text{A 接口话务量} / \text{网上发展用户数}$	地区粒度
开机率	$\text{VLR 开机用户数} / \text{网上发展用户数} * 100\%$	地区粒度

主题确定后，数据仓库的数据逻辑模型也就基本成型了。此时，需要在主题的逻辑关系模型中添加相关的属性。向主题里增加所需要的信息以及能充分代表主题的属性。如在 HLR 主题中的属性有：时间和日期，是否处于忙时，HLR 网元信息，地理属性等。这些属性作为主题的维度。HLR 主题中有很多用户关心的指标（如上表），作为事实表的内容。

其概念模型为

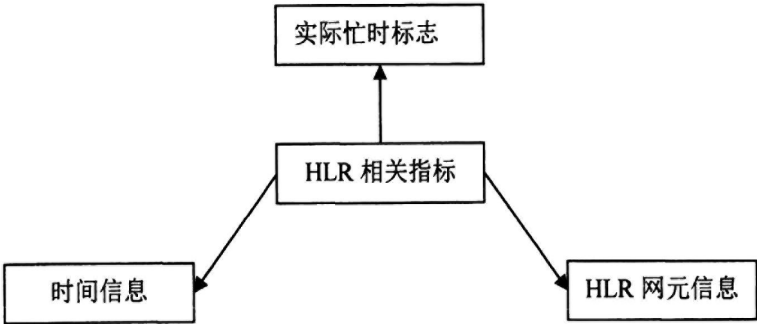


图 3.1 HLR 概念模型

可以看出描述主题的有三方面的属性分别建立三个维度，时间维、实际忙时维和 HLR 网元维，HLR 维又引用了地理维。

其具体逻辑模型为

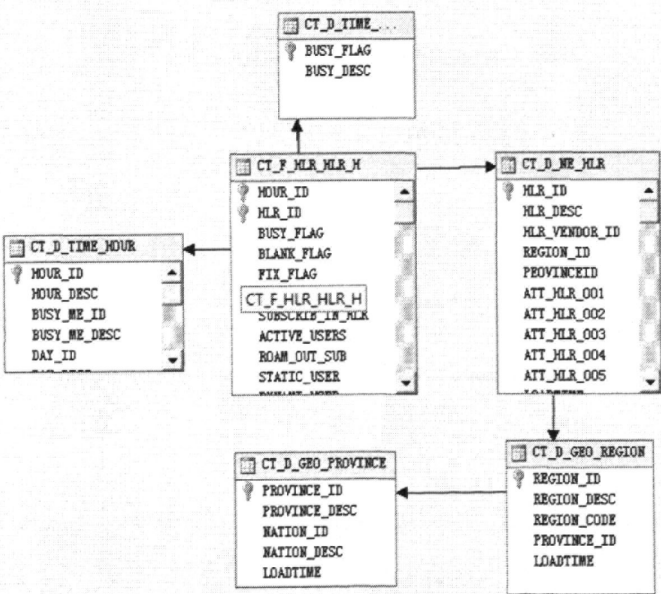


图 3.2 HLR 数据模型

可以看出这是一种的雪花模型，雪花模型将维表进一步规范化，将维度进行层次化处理，在一定程度上减少了存储空间，并使复杂维度的层次结构清晰，同时规范化的结

构更容易更新和维护。由于无线与交换主题和 HLR 主题数据有很多相似的特点，并且存在共享的维度（时间维、地理维等），与所有事实表都相关，并且有些维度之间存在引用关系，实际上数据模式是比雪花模式更复杂的星座模式。如图所示：

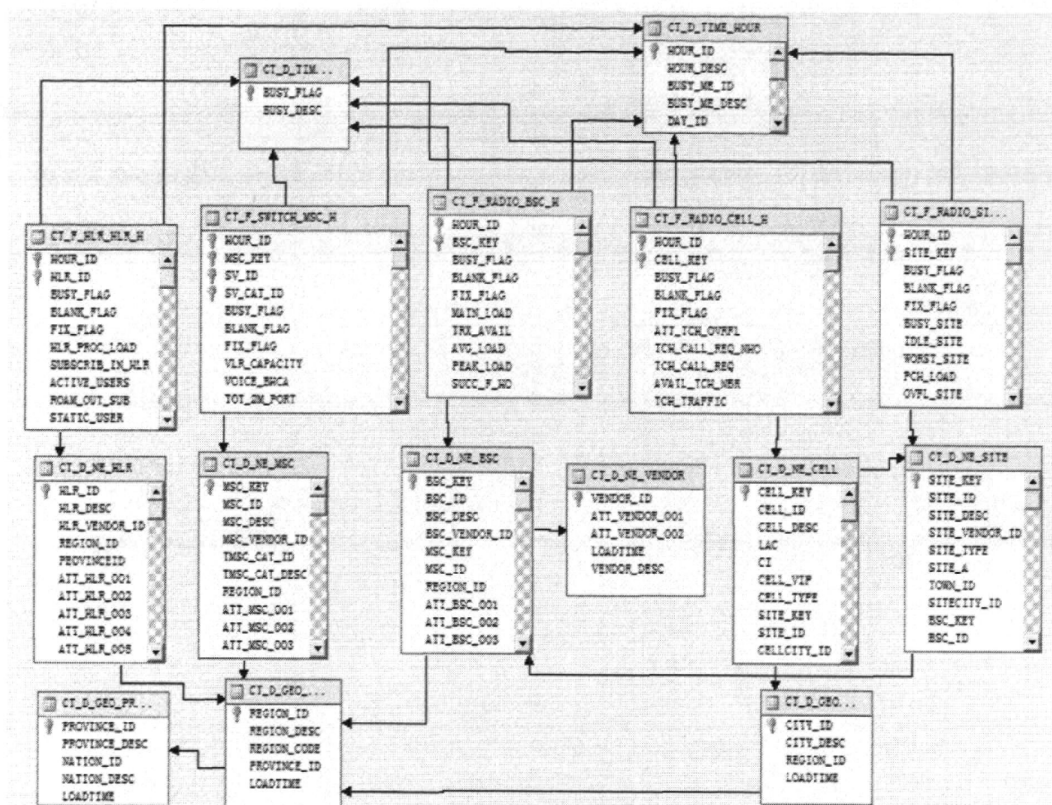


图 3.3 无线与交换主题数据模型

### 3.3.3 维度的设计

在数据仓库中，数据表分为两类：一类是事实数据表，用来存储数据仓库中的实际数据，如 HLR 主题中 CT\_F\_HLR\_HLR\_H 表中存储着与 HLR 相关的各种指标数据；另一类是维度数据表，用来存储数据仓库中的维度数据，如这里的关于时间、网元信息、实际忙时和地理等分析要素的表均为维度表。

事实表是数据仓库中最大的表，事实表中包含了相关主题的详细信息，是对主题进行各种分析的基础。维度是人们在观察数据时的观察角度。例如按照时间或者地理进行分析，那么这里的时间、地理就是相应的维度。通过不同的维度，可以看到数据在各个维度的数据情况，同时也可以通过多种维度来进行交叉分析。因此维度的设计也是至关重要的一步。数据仓库中 C 网主题主要的几个维度设计如下：

#### 时间维度

数据仓库中保留的数据都是历史数据，因此时间维是数据分析最基础的一个维度。时间维度也是各个主题分析中的共享维度，是必不可少的。时间维度的成员包括年、季、月、日、小时、年中月、周、年中周、星期、月中日、日中时、早晚忙时。

其层次结构为：

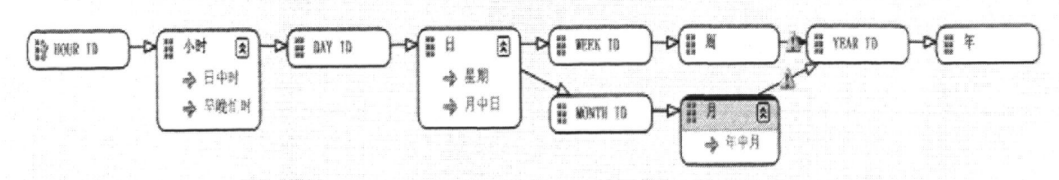


图 3.4 时间维层次结构

地理维度

地理维在数据展现中也是非常重要的，因为我们经常要以地理角度去观察和分析数据，同地理信息结合起来能够使用户更直观的理解数据。地理维成员包括省份、地区。

层次结构：

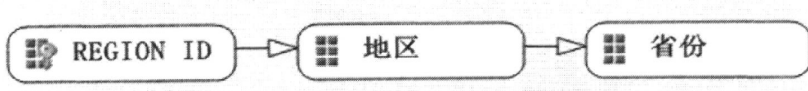


图 3.5 地理维层次结构

各类网元维

网元就是 Network Element，是一个网络系统中的某个网络单元或者节点，该单元能独立完成一种或几种功能的设备。在移动通信网络系统中，一个基站是一个网元，能单独完成一项功能的实体都可以称为一个网元，比如 BTS、BSC、MSC、SGSN、GGSN、HLR、交换机等都是网元。网元之间关系有并行的，也有上下级的关系。如一个本地网的 BSC 网元间的关系是平行关系，而 BSC 网元和 BTS 网元，就是控制与被控制的关系。

上图星座模型图中包含 5 张事实表，分别对应 5 个级别的网元。分为小区维、基站维、BSC 维、MSC 维、HLR 维。其层次结构分别为：

小区维

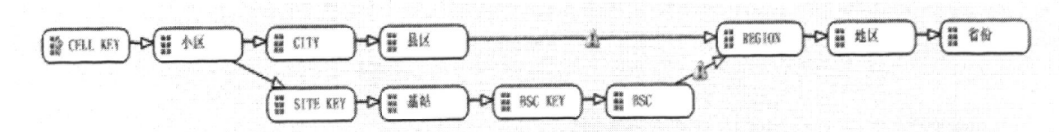


图 3.6 小区维层次结构

### 基站维



图 3.7 基站维层次结构

### BSC 维

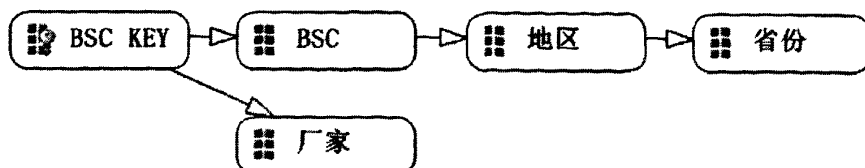


图 3.8 BSC 维层次结构

### MSC 维

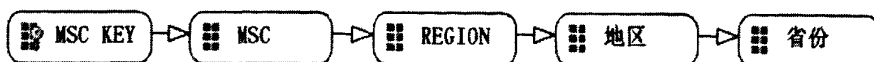


图 3.9 MSC 维层次结构

### HLR 维



图 3.10 HLR 维层次结构

实际上这些网元维都有地理属性，从结构图中我们可以看到，各网元维都引用了地理维。通过引用实现各网元对地理维度的共享，节省存储空间，提高设计效率。

## 3.3.4 创建数据仓库数据库

数据仓库的整个设计过程完成后，就可以在服务器上创建数据仓库了。数据仓库其实也是一种数据库，其日常的管理维护，也是通过数据库管理系统(DBMS)来进行管理的。因此，数据仓库可以像一般的数据库一样进行各种操作。

通过 SQL Server Management Studio，按照建立普通数据库的方法建立一个名为“OBIDB”的数据库作为我们的数据仓库。以下是创建数据库的脚本。

```

USE [MASTER]

GO

/***** CREATE DATABASE [OBIDB] *****/

CREATE DATABASE [OBIDB] ON PRIMARY

( NAME = N'OBIDB', FILENAME = N'D:\OBIDB\OBIDB.MDF', SIZE = 8192MB ,

```

```

MAXSIZE = UNLIMITED, FILEGROWTH = 1024KB ),

/**** DIMENTION'S FILE GROUP *****/

FILEGROUP [DIMFG] ( NAME = N'DIMFG', FILENAME =
N'D:\OBIDB\DIMFG.NDF' , SIZE = 8192MB , MAXSIZE = UNLIMITED, FILEGROWTH
= 1024KB ),

/**** FACT'S FILE GROUP FOR MONTHS *****/

FILEGROUP [FACTFG201001] ( NAME = N'FACTFG201001', FILENAME =
N'D:\OBIDB\FACTFG201001.NDF' , SIZE =8192MB, MAXSIZE = UNLIMITED,
FILEGROWTH = 32768KB ),

... ..

FILEGROUP [FACTFG201212] ( NAME = N'FACTFG201212', FILENAME =
N'D:\OBIDB\FACTFG201212.NDF' , SIZE =8192MB, MAXSIZE = UNLIMITED,
FILEGROWTH = 32768KB ),

/**** FACT'S FILE GROUP FOR YEARS *****/

FILEGROUP [FACTFG2010] ( NAME = N'FACTFG2010', FILENAME =
N'D:\OBIDB\FACTFG2010.NDF' , SIZE = 8192MB , MAXSIZE = UNLIMITED,
FILEGROWTH = 32768KB ),

... ..

FILEGROUP [FACTFG2012] ( NAME = N'FACTFG2012', FILENAME =
N'D:\OBIDB\FACTFG2012.NDF' , SIZE = 8192MB , MAXSIZE = UNLIMITED,
FILEGROWTH = 32768KB ),

LOG ON

(NAME=N'\OBIDB_LOG', FILENAME = N'D:\OBIDB\OBIDB_LOG.LDF' , SIZE =
4096KB , MAXSIZE = UNLIMITED, FILEGROWTH = 10%)

COLLATE CHINESE_PRC_CI_AS

```

数据仓库数据库建完后，将设计好的事实表和维表按照统一的命名规则创建到数据仓库中，数据类型根据原始数据库的各个表中字段的数据类型进行设置。

### 3.4 ETL 平台的构建

创建数据仓库是一个复杂的系统工程，整个系统的成败与每一环节的性能和效率息息相关。在创建数据仓库系统的各个环节中，数据的提取、转换和加载过程(ETL



过程)尤其重要,它是数据仓库系统能否成功的关键。ETL性能的好坏能够直接影响到整个系统运行的效率和使用结果,最重要的是其决定着数据仓库数据的质量。

### 3.4.1 ETL工具简介

SQL Server 2008的数据集成服务(SQL Server Integration Services) 是用于创建企业级数据集成及转换解决方案的平台。SSIS为构建ETL应用程序提供了所需的大量必要功能,并且有非常高的扩展性能。SSIS的作为SQL Server 2008中的一个组件,随SQL Server 2008的安装可以免费获得,因而不需要再购买其额外的许可。

使用SQL Server Integration Services能够用来解决复杂的业务问题。可用来提取和转换来自多种数据源(如平面文件、XML数据文件和关系数据库等)的数据,然后将数据加载到指定的目标数据库中。SSIS中有一组丰富的用于构造包的工,及用于管理和运行包的SSIS服务<sup>[34]</sup>。可以用SSIS图形设计工具创建解决方案,甚至可以不用编写一行代码;也可以通过编程方式来创建包,并编写自定义任务和其它包对象的代码。这些强大的特性都使SSIS成为理想的ETL平台。整个Microsoft SQL Server Integration Services 平台的体系结构如图所示<sup>[35]</sup>。

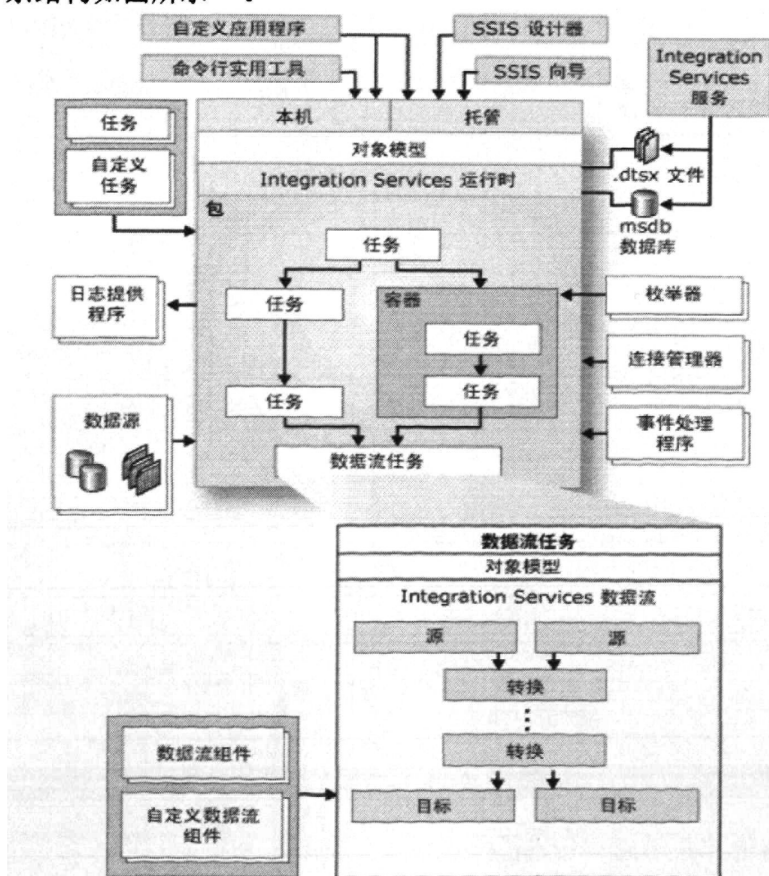


图 3.11 SSIS 体系结构

SSIS 能够支持大数据量的操作，并且可以处理十分复杂的情况。它是面向处理过程的，易于创建、编辑和理解，能够完成从简单到复杂的各种处理过程；SSIS 是基于组件的体系结构，各种操作都可以通过组件来完成，但也支持编程扩展组件功能；支持广泛的源数据格式和多平台工作。因此，SSIS 具有伸缩性、易用性、灵活性、互操作性等特点。

### 3.4.2 数据源分析

在数据进行采集之前，必须对数据源进行分析。源数据分析包括：数据范围及业务含义；数据所在平台，包括系统平台和数据库平台；数据结构；数据更新周期和数据量等。

对话务综合分析的数据仓库来说，数据主要有以下几个来源：

#### 1. C 网话务网管

- 数据范围

- 无线网元 (Cell、Site、BSC)

- 交换网元 (MSC)

- 中继网元 (中继群)

- 数据接口

- 采用 DB 接口

- 抽取周期和频次

- 1 次/天，默认每天 5:00 开始抽取前一天 24 小时的数据。

#### 2. 固网话务网管

- 数据范围

- 中继群数据，目的码数据。

- 数据接口

- 采用 DB 接口

- 抽取周期和频次

- 1 次/天，默认每天 6:00 开始抽取前一天 24 小时的数据。

#### 3. 中兴信令监测系统

- 数据范围

- TUPCDR 数据

- ISUPCDR 数据

中继群配置数据

- 数据接口

文件接口

- 抽取周期和频次

1 次/天，默认每天 6:00 开始抽取前一天 24 小时数据。

特殊要求：中兴信令监测系统按照接口约定将所需数据定期上传到指定文件服务器上。网络业务综合分析系统待数据抽取完毕后，负责对文件进行删除，保证空间正常释放。

### 3.4.3 数据的抽取策略

在数据仓库创建完毕后，一般需要将源数据库中所需要的全部数据装载进来，这时需要对数据进行全量抽取。全量抽取一般采取数据复制或备份的方法来完成。数据全量抽取完后，接下来的操作只需抽取表中自上次抽取后，新增或修改过的数据，也就是数据增量抽取。如何定期向数据仓库中增量追加数据也是十分重要的技术。优秀的增量抽取方案要求能够将源系统中更新的数据按一定的时间准确地捕获到，并且不能对原有系统造成太大的压力，影响原有业务。

数据仓库的数据来自各网管数据库，如何及时准确地追加数据是保证数据仓库数据准确性的关键。由于网管数据库中的数据都包含有时间戳字段，因此我们可以采用时间对比方法来解决数据仓库中数据追加的问题。即通过对时间字段的比较，正确地插入新增数据。为了防止源数据库数据因故未能及时采集数据，出现数据漏采现象，我们在设置采集方案时，设定每天定时采集前三天的数据，重复采集，提高数据的完整性和准确性。

### 3.4.4 渐变维的处理方法

随着通信技术的发展，近年来移动通信用户数量在飞速增长，通信网络不断扩容，这就需要不断接入新的网元。并且对话务容量的需求波动性较大。为了扩容或者优化网络，网元割接会经常出现。较为常见的割接情况有 BSC 与 MSC 或 BTS 与 BSC 之间的所属关系发生变化，或者有新的网元接入。数据仓库中存在网元维表，网元的不断割接给网元维表的维护带来了很大问题。因此，有效地解决这个问题是保证数据仓库数据质量的一个重要方面。维表中的数据变化相对于事实表来说是缓慢的，所以称为渐变维。对于处理渐变维，一般有三种解决方案<sup>[36]</sup>：

### 1. 覆盖当前记录，不保留历史记录。

通过这种方法，用户能够得到维表的最新信息。其特点是比较方便，容易实现。但是对于主要用于保存大量的历史数据的数据仓库来说，不保存属性的变化历史，没有维表的历史信息，不能对历史数据进行任何的维护，不能跟踪维度属性的变化。

### 2. 保留历史记录，增加新的记录。

通过这种方法，用户既可以得到维表的当前的信息也能够使用历史信息，可以准确地跟踪维度属性的变化过程。为了保证分析的准确性，在记录中加入时间字段和标志位字段，用于说明维表信息发生变化的时间以及当前有效记录。

### 3. 保留旧记录，增加新字段记录变化值。

通过这种方法，用户保存维度的原有记录，通过增加新的列来记录新的值和发生时间。这样用户可以得到上次改变的记录和当前记录。但当维度信息发生多次改变时，并不能记录所有改变。

根据网管数据的特点，采用第二种解决方案，即保留原来的信息，将割接后的网元的所属关系作为一个新的记录插入到维表中；并在维表中加入开始时间字段和结束时间字段，开始时间表明网元的启用时间，结束时间表明网元的停用时间，结束时间为空则说明网元仍在使用的。

## 3.4.5 包的设计

在 SSIS 中“包”是一个有组织的集合，其中包括连接、控制流元素、数据流元素、变量和配置、事件处理程序。可以使用 SQL Server Integration Services 的图形设计器或以编程方式将这些对象组合到包中，然后将完成的包保存到 SSIS 包存储区或文件系统中。“包”一般是由一个控制流及一个或多个数据流组成。SSIS 提供了三种不同类型的控制流元素：提供包中结构的容器、提供功能的任务和优先约束；提供三种不同类型的数据流组件：源、转换及目标。

数据 ETL 过程实际上都是通过执行包来实现的，因此包的设计非常重要。主要有两类包，一类包用于维表的采集，一类用于对事实表的采集。我们选取几个重要的包来说明包的设计过程。

### 1. 维表的采集

#### (1) 时间维的生成

时间维度是不变的，时间维表建好后，可以通过手工编写脚本插入时间维数据。

#### (2) 地理维的采集

通过包 re\_ct\_d\_geo.dtsx 来实现对地理维的采集。地理维的采集按照地理维的结构层次由上到下顺次采集，先进行省份维表的采集，然后是地区和县市。如图是控制流的设计图，包括三个数据流任务。

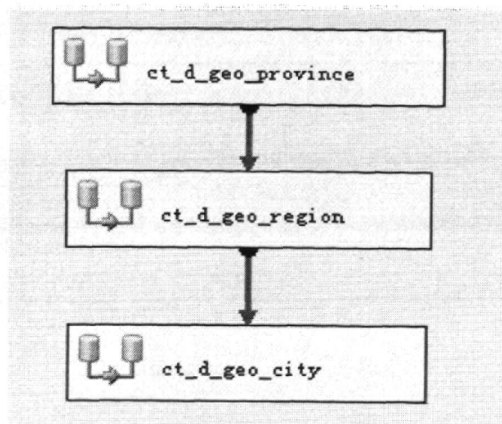


图 3.12 地理维控制流的设计图

由于我们的综合话务分析系统用于省内，因此省份维表中只有黑龙江一条数据。对地区维表的采集其数据流设计图为：

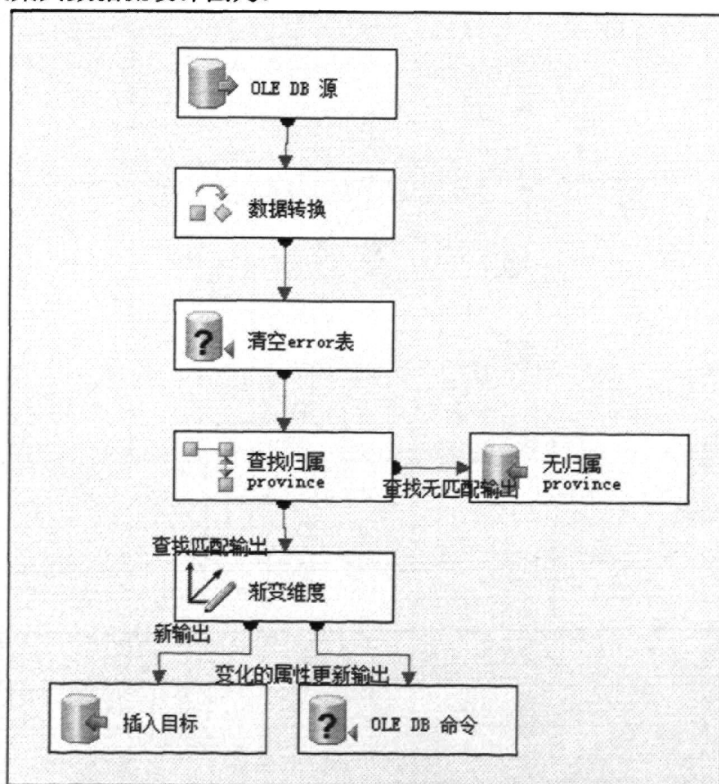


图 3.13 地理维数据流的设计图

OLE DB 源通过使用数据库表、视图或 SQL 命令，从各种兼容 OLE DB 的关系数据库中提取数据。在此使用 SQL 命令从数据库 nrmdb 网元配置数据库中抽取所需要的

字段。

在数据转换组件中可以将输入列中的数据转换为其他数据类型，然后将其复制到新的输出列。在此组件中我们将源数据库中数据类型不符合要求的转换为合适的类型。

在查找组件中，我们查找所采集的地区数据是否属于黑龙江省，如果信息不匹配说明出现错误，将不匹配的数据存入表 CT\_D\_GEO\_REGION\_ERR 中，以便维护人员查看，处理相关问题。为了能够自动清理错误表中的数据，在查找前添加一个组件清除数据。如果信息匹配，则将采集的数据进行渐变维转换。如地区维度没有变化，则不对维表进行改动。如有变化，比如黑龙江省划出一个新的地市，或者地市改变名称等，就要用到渐变维度的转换。通过命令，将维度更改后的信息覆盖现有信息，存入表 CT\_D\_GEO\_REGION 中。县市级的维表采集类似。

(3) 无线与交换网元维的采集

通过包 re\_ct\_d\_ne.dtsx 来实现对无线与交换各网元维表的采集。根据网元的所属关系从上到下采集，控制流设计如图所示，包括四个数据流任务，分别对 MSC 维、BSC 维、基站维和小区维的数据采集。如下图所示。

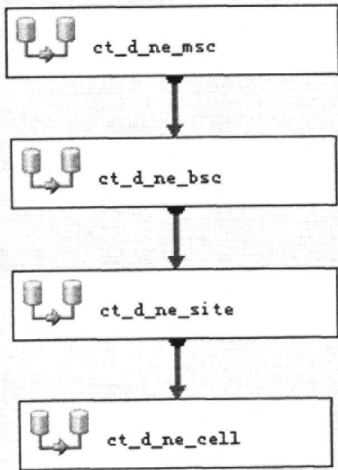


图 3.14 无线与交换维控制流的设计图

在对各网元维的采集过程中尤其要注意的是对渐变维度的处理。因为随着通信用户的不断增加，网络扩容是不可避免的，这就要接入新的网元；还有可能进行网元的调整，比如 MSC 归属地的改变，BSC 与 MSC 之间的所属关系等，都会引起网元维度的缓慢变化。

数据经过渐变维度变换有三种输出。如下图是 MSC 维中渐变维度转换，变化的属性更改覆盖现有记录，渐变维度转换将这些行定向到名为“变化的属性更新输出”的输

出。OLE DB 命令转换使用 UPDATE 语句更新记录。变化的属性更新输出主要用于接入新的网元，直接更新网元维表即可。历史属性更改创建新记录而不更新现有记录。现有记录中允许的唯一更改是对指示记录是当前记录还是过期记录的列的更新。渐变维度转换将这些行定向到两个输出：“历史属性插入输出”和“新输出”。

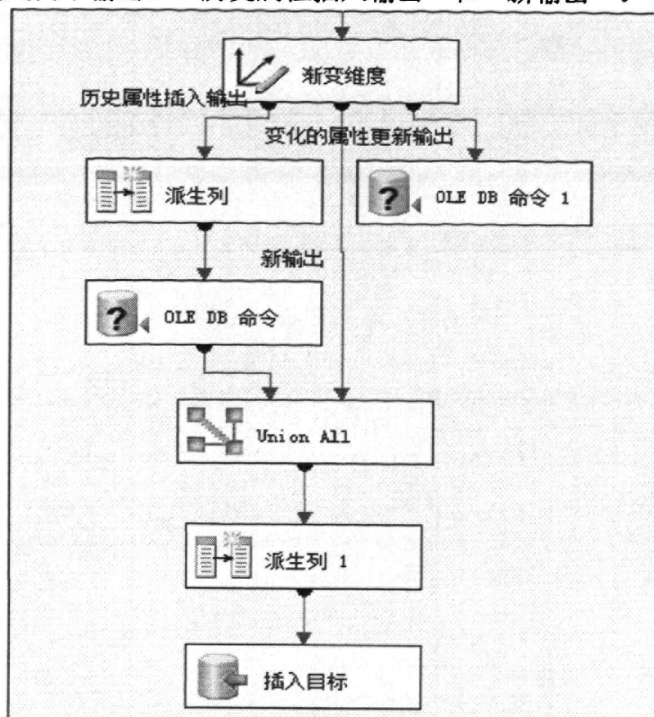


图 3.15 渐变维设计图

这主要用于网元的割接，比如网元归属地的变化。如下图，一个 MSC 网元经过两次割接，归属地由哈尔滨割接到大庆，然后到绥化地区。End\_Time 为 NULL 说明该网元信息仍在使

结果 消息					
	MSC_ID	MSC_DESC	REGION_ID	START_TIME	END_TIME
1	21706586	HRB-MSCE-SOFT	-257096336	2009-05-05 16:39:00	2009-05-26 05:05:00
2	21706586	HRB-MSCE-SOFT	-1651670625	2009-05-26 05:05:00	2009-12-21 06:00:00
3	21706586	HRB-MSCE-SOFT	-1568693242	2009-12-21 06:00:00	NULL

结果 消息				
	REGION_ID	REGION_DESC	REGION_CODE	PROVINCE_ID
1	-1651670625	大庆市	NULL	153199949
2	-1568693242	绥化地区	NULL	153199949
3	-257096336	哈尔滨市	NULL	153199949

图 3.16 一个 MSC 网元割接情况

## 2.事实表的采集

通过设计四个包(re\_ct\_f\_ne.dtsx、re\_ct\_f\_gw.dtsx、re\_ct\_f\_cdr.dtsx、re\_ct\_f\_tkqp.dtsx)

分别对无线与交换、中继群、固网和 CDR 信令的事实表进行采集。

下面以包 re\_ct\_f\_ne.dtsx 为例，说明一下设计过程。调用包 re\_ct\_f\_ne.dtsx 对无线与交换的事实表进行采集，其控制流设计图如下图：

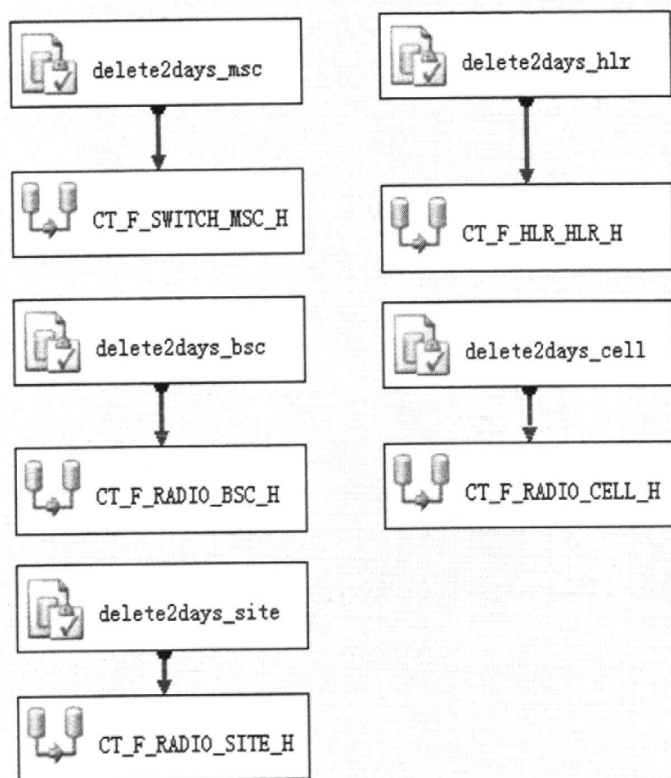


图 3.17 无线与交换事实表采集控制流设计图

上图是控制流设计图，包括五部分，分别对五种网元事实表进行采集，由于我们的采集策略是对前三天的数据进行采集，因此每天采集的时候，要把前一天采集的三天数据中的前面两天的数据删除，在上图中我们可以看到在每个事实表采集开始时都有一个执行 SQL 任务，来删除两天的数据，其脚本为：

```
delete from obidb.CT_F_HLR_HLR_H
where
    HOUR_ID >= convert(int, CONVERT(varchar(8), dateadd("day",-3,GETDATE()), 112))*100
and
    HOUR_ID < convert(int, CONVERT(varchar(8), dateadd("day",0,GETDATE()), 112))*100
```

以 BSC 数据采集为例，删除两天数据后进行数据采集，其数据流设计图如下：



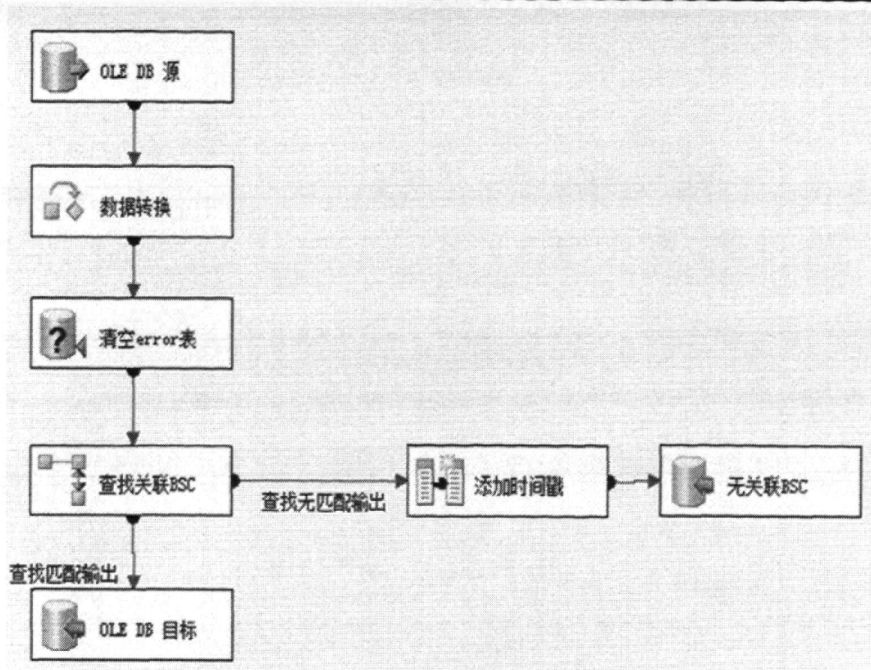


图 3.18 BSC 数据采集控制流设计图

OLE DB 源通过使用数据库表、视图或 SQL 命令，从各种兼容 OLE DB 的关系数据库中提取数据。从数据库 npmdb 中通过 sql 命令抽取所需要的指标字段。

在 sql 命令文本中，列出抽取的字段，并且通过下列 where 语句限定时间范围，抽取前三天数据。

where

`bsc.first_result >= to_date(to_char(sysdate-3,'yyyy-mm-dd'),'yyyy-mm-dd hh24:mi:ss')`

and

`bsc.first_result < to_date(to_char(sysdate,'yyyy-mm-dd'),'yyyy-mm-dd hh24:mi:ss')`

在数据转换组件中可以将输入列中的数据转换为其他数据类型，然后将其复制到新的输出列。在此组件中我们将源数据库中数据类型不符合要求的转换为合适的类型，为数据进入数据仓库做好准备。

在查找组件中，我们查找所采集的 BSC 数据在 BSC 网元维中是否有匹配的网元信息，如果没有，说明可能有新接入的网元，在网元维中还没有及时正确的更新，添加时间戳后，将不匹配的数据存入表 CT\_D\_NE\_BSC\_ERR 中，以便维护人员查看，处理相关问题。为了能够自动清理错误表中的数据，在查找前添加一个组件，删除 30 天前的数据，只保留一个月的错误信息以备查询。如果网元维中有匹配的信息，则将采集的数据存入事实表 CT\_F\_RADIO\_BSC\_H 中。

## 3.5 联机分析设计

### 3.5.1 建立多维数据集

原始的业务数据经过转换装载到数据仓库中，这就为分析决策打下了基础。用户在系统中的分析操作都是基于这些数据而来。但是，我们在进行数据的多维分析时，却不是直接基于数据仓库中所有数据的，而是从中提取相关的子集，建立多维数据集 Cube。因此，在具体分析数据时，通常要创建多维数据集。

在数学中用 X、Y、Z 三个坐标轴来表示 3 个方向，创建一个空间立方体，用类似的方法，我们可以用不同的观察角度做为维度，创建一个分析用的数据立方体。多维数据集（Cube）是二维表格的多维扩展，我们可以把三维 Cube 看作是一组相互叠加在一起的二维表格。实际上多维数据集的维度并不止是三维的。多维数据集中的数据是经过处理的并聚成立方的形式，因此基本不需要在多维数据集中进行计算。这就是为什么我们在访问多维数据集是能够快速得到相应的原因。

#### 1. 定义数据源

Microsoft SQL Server Analysis Services 数据源是一个对象，提供连接到数据源所需的信息，用于建立话务数据分析的多维数据集。首先要在 Business Intelligence Development Studio 中创建一个新的 Analysis Services 项目，我们将其命名为 HLJCT\_Cube\_SqlServer\_2008。创建了项目之后，在解决方案资源管理器中，通过数据源向导将创建一个“数据源”，将其与数据仓库 OBIDB 建立连接。并将数据源命名为“综合分析”。

#### 2. 定义数据源视图

数据源视图是 SSAS 中的一个对象，它包含来自所选数据源对象的元数据。数据源视图提供了一种优良的机制，能够避免再设计过程中应用程序直接使用底层的数据库，尤其在数据源中包含大量的表，但其中只有较少的表用于应用程序中；再就是 Analysis services 项目需要来自多个数据源的数据时，数据源视图是很方便的。数据源只能提供与源数据库的简单连接，而更多的功能，如添加关系、设置属性等要通过数据源视图来完成。

我们以创建 C 网的数据源视图为例，说明创建过程。

（1）在解决方案资源管理器中，右键单击“数据源视图”，再单击“新建数据源视图”。打开数据源视图向导。

（2）设置名称匹配。由于没有为数据仓库中的各表设置外键及其关系，Analysis

services 服务在需要匹配的列上建立逻辑关系时找不到外键，在此处选择“与主键同名”。

### (3) 选择表和视图

选择与 C 网有关的事实表和维度表，其它固网及 CDR 的相关表则不选。建立一个 C 网相关的数据源视图。将这个数据源视图命名为 C 网专题。

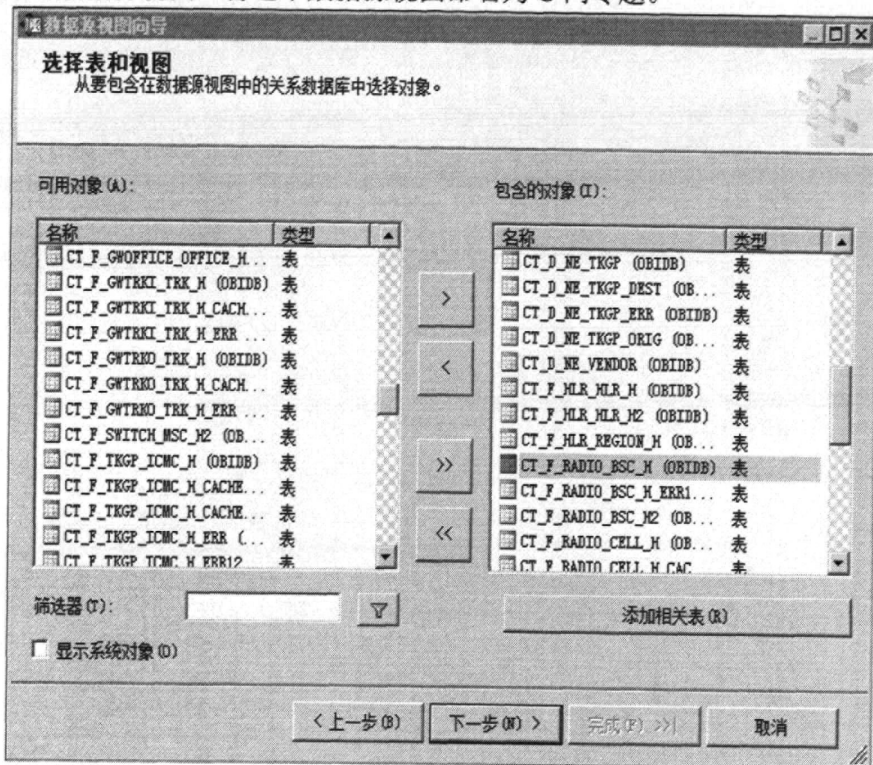


图 3.19 选择表和视图

### (4) 创建关系

对每个表设置好主键和外键后，然后设定事实表和维度表之间的关系。方法是拖动事实表中定义的外键字段到维度表中的同名字段就可建立关系。如果是组合外键那么要将多个字段同时选中拖到目标表。为保证其主外键关系符合业务中的逻辑关系，一定要把事实表作为源，维度表作为目标。由于表比较多，关系比较复杂，直接拖动可能比较混乱，可以通过在数据源视图设计面板的左侧表名的列表中进行设置。在所要设置的表名单击右键，点击‘新建关系’，然后再选择源表和目标表的关联字段。有些维表之间也存在主外键的关系，用同样的方法进行设置。

数据源视图设计器会基于数据库中的数据关系自动定义添加到数据源视图中的各表之间的关系，或基于指定名称的匹配条件。并且在数据源视图中进行手动定义逻辑关系，以便对自动定义的关系进行补充。

如下图是设置完后的关系图：

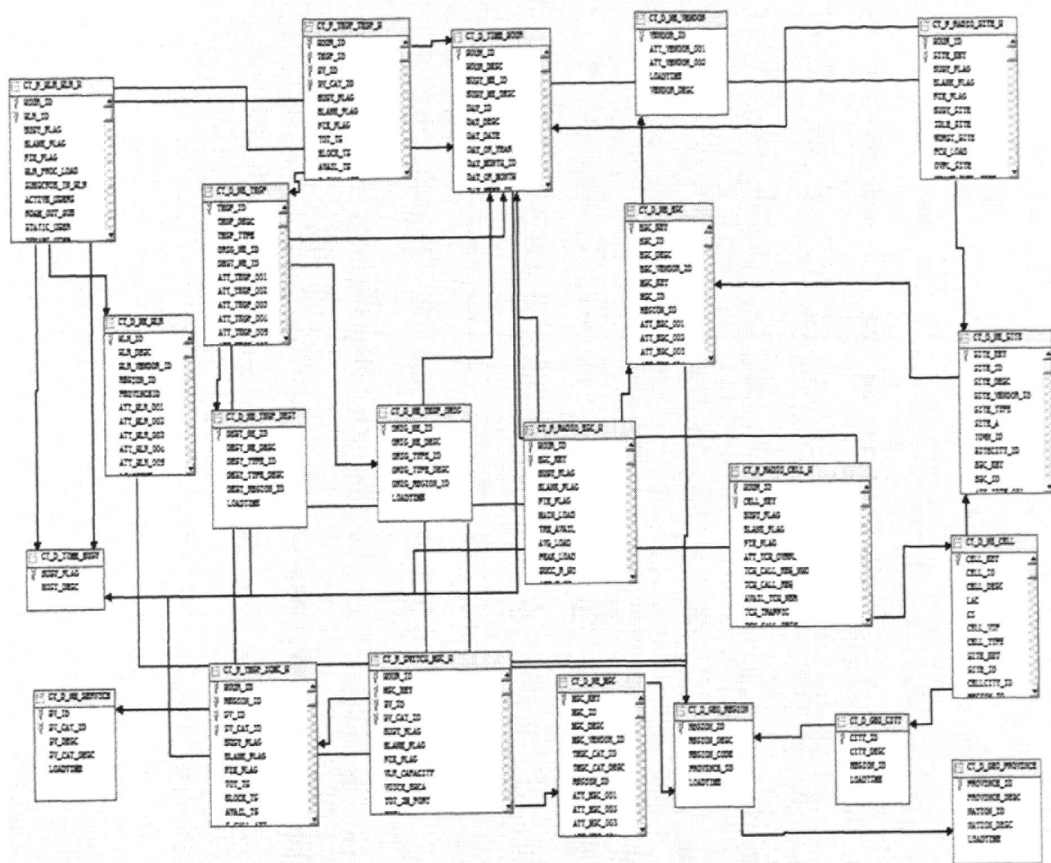


图 3.20 C 网数据源视图各表相互关系

### 3. 生成多维数据集

创建了 C 网专题的数据源视图之后，便可以右键单击“解决方案资源管理器”中的“多维数据集”图标，然后点击“新建多维数据集”，来创建多维数据集。我们利用 C 网专题的数据源视图，根据业务分类建成两个多维数据集，C 网无线和交换专题多维数据集和 C 网中继专题多维数据集。下面以 C 网无线和交换专题多维数据集为例说明创建过程。

(1) 解决方案资源管理器中右键单击“多维数据集”，并点击“新建多维数据集”。  
打开多维数据集向导。

(2) 选择生成多维数据集的方式。选择“使用现有表”，即通过数据源视图中所包含的表自动生成多维数据集。

### (3) 标识事实表和维度表

选择 C 网专题数据源视图为多维数据集的数据基础, 根据前面设计好的关系设置事实表以及维度表, 将事实表中的字段就是多维数据集中的度量值, 维表用于创建多维数据集中的维度。

#### (4) 设置维度的层次结构

根据前面的维度设计的层次结构将多维数据集中用到的维度进行设置。通过设置维度的层次结构，用户就可以轻松地按照维度的层次结构进行上钻、下钻等操作。

多维数据集创建完后，在数据源视图框中，展现的是多维数据集中各表之间的相互关系，用黄色标记事实表，蓝色标记维度表。完成创建多维数据集向导后，还可以通过多维数据集设计器来添加、删除和修改各表间的相互关系，配置维度属性以及层次结构。多维数据集建好后，经过处理将其部署到 Analysis Services 数据库中。如图是 C 网无线和交换专题 Cube 的数据源视图。

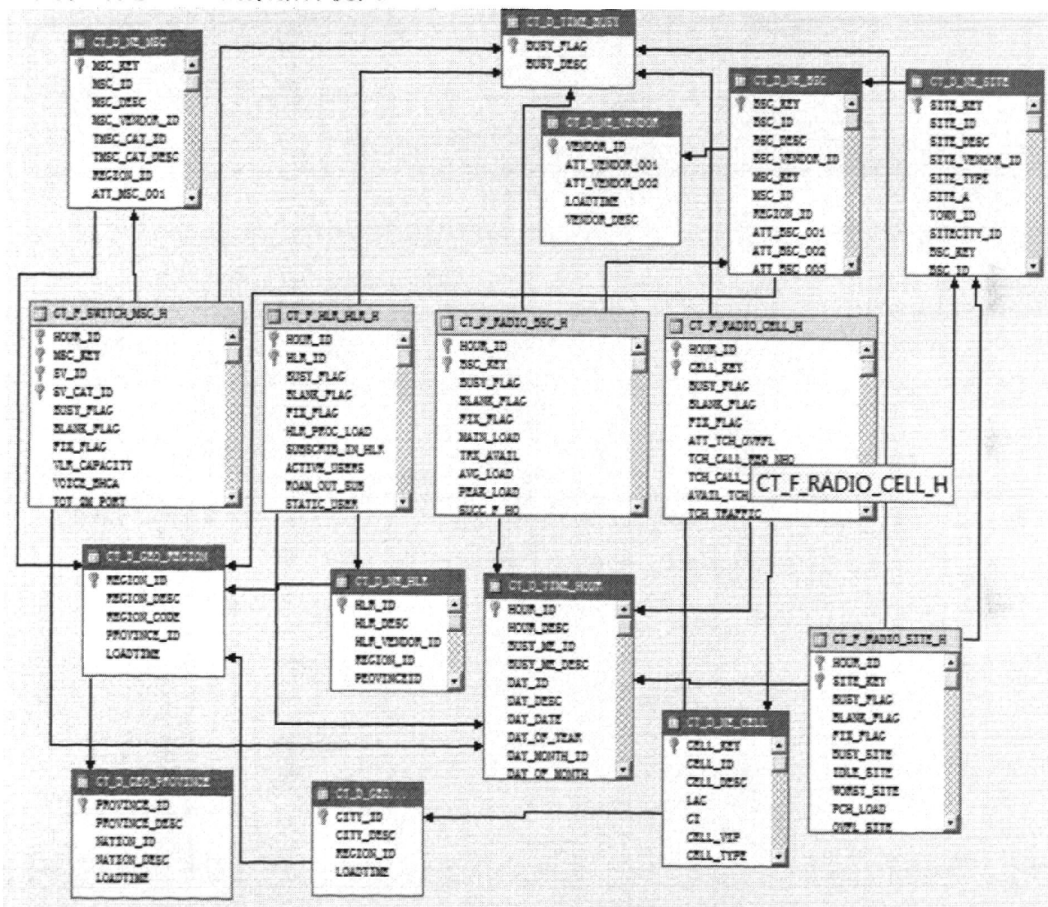


图 3.21 C 网无线和交换专题 Cube

C 网中继专题 Cube 的创建过程类似。下图是 C 网中继专题的数据源视图。用同样的方法进行创建固网专题和 CDR 专题的数据立方体。

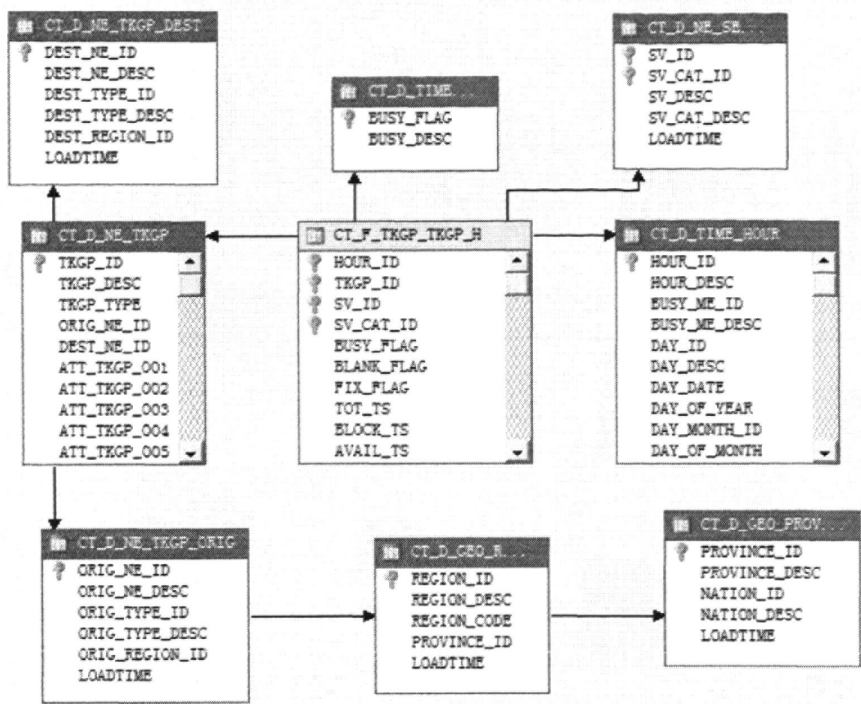


图 3.22 C 网中继专题 Cube

3.5.2 OLAP数据展示

OLAP(On-Line Analytical Processing)是近年来兴起的广泛应用于数据仓库中的一种多维数据分析技术。随着商业智能技术的不断发展，以 OLAP 技术为基础的前端展示工具也成为支持企业信息决策必不可少的工具<sup>[37]</sup>。

B/S 结构（Browser/Server，浏览器/服务器模式），是 WEB 兴起后的一种网络结构模式，WEB 浏览器是客户端最主要的应用软件。这种模式统一了客户端，将系统功能实现的核心部分集中到服务器上，简化了系统的开发、维护和使用。客户机上只要安装一个浏览器，服务器通过 Web Server 同数据库进行数据交互。B/S 最大的优点就是可以在任何地方进行操作而不用安装任何专门的软件，只要有一台能上网的电脑就能使用，客户端零维护,系统的扩展非常容易<sup>[38]</sup>。

用户在Web浏览器上发出请求，通过HTTP链接至Web服务器，Web服务器则将请求解析成MDX语句，并通过ADOMD.NET和OLAP服务器建立连接。通过查询多维数据集将结果返回给Web服务器，最终由Web服务器传送到Web浏览器上。用户访问基于Web的数据仓库可以是跨部门，跨区域的，不同的用户权限会有所不同，所能浏览的信息也不同。系统在Web上的总体设计如图所示：



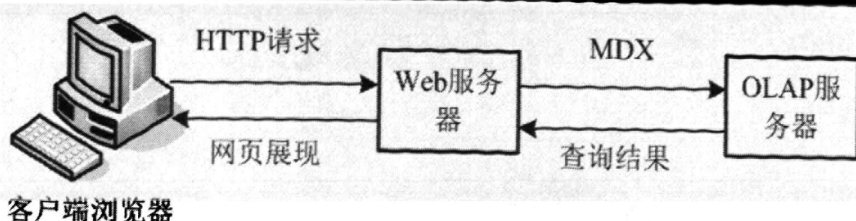


图 3.23 Web 结构图

从上图看出分为三个层次，第一层是 OLAP 服务器，它实现与数据仓库系统的连接，将数据仓库中的数据聚合成多维数据集；第二层是 WEB 服务器，将客户端的请求分解成 OLAP 分析的各种操作；第三层是前端展现工具，用于将处理得到的结果以多维报表、柱状图、趋势图形等直观的方式展现给用户。这种三层体系结构使数据、应用逻辑和客户应用分离开，有利于系统的维护和升级。当系统需要修改功能或者增加功能时，可以只修改其中的某些部分，而不需要向两层的客户/服务器体系那样做整体的改动。

### 3.6 本章小结

本章主要介绍了综合分析系统整个设计过程，重点介绍了系统设计中的几个关键环节，包括系统需求分析，分析专题设计，数据模型设计，以及 ETL 平台的设计等。

## 第 4 章 综合分析系统的实现

### 4.1 系统配置

#### 4.1.1 硬件结构

综合话务分析系统硬件拓扑图如图 4.1 所示：

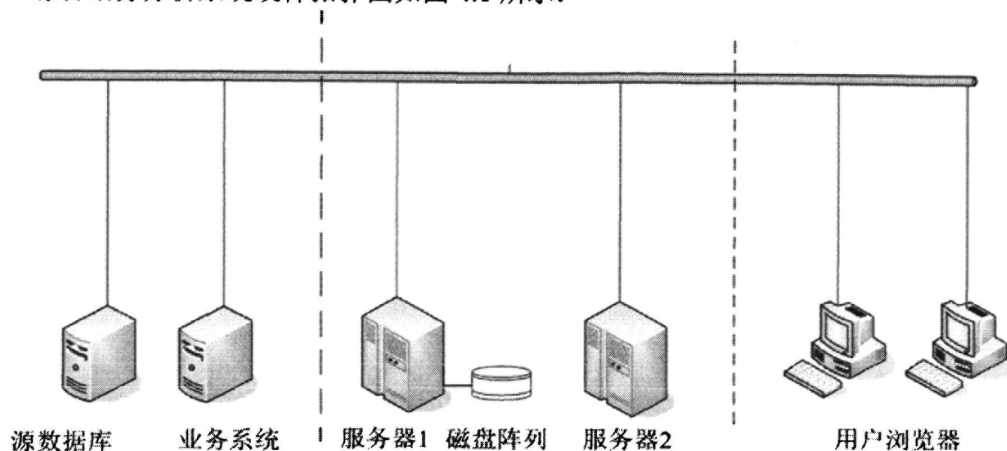


图 4.1 系统硬件拓扑图

PC 服务器 1：ETL 数据采集程序，OLAP 服务器，运行在这台机器上。

PC 服务器 2：Web 服务器运行在这台机器上。

ETL 数据采集程序每天凌晨开始调度执行，抽取前三天的数据，然后进行处理。

OLAP 服务器和 Web 服务器主要响应用户查询，负荷高峰一般发生在白天。

磁盘阵列：为大数据量提供数据存储空间。

浏览器：提供用户数据查询和分析界面，展现查询和分析结果。

#### 4.1.2 开发环境

服务器操作系统采用 MS Windows Server 2008

数据库软件采用 MS SQL Server 2008 企业版

### 4.2 数据预处理

#### 4.2.1 数据采集

数据采集通过构建的 ETL 平台来进行，主要调用各采集包来实现。由于各采集包比较多，单个调用比较麻烦，因此我们另外设计包，将相关的包放到一起，通过这个包调用其他包。

##### 1. 维表的采集



通过一个代理包 `agent_etl_d.dtsx` 来实现各维表的采集, 在这个包内调用其他各维表的采集包进行各维表的采集。其控制流设计如图 4.2 所示。每天执行这个包, 在包内调用其他各个维表的采集包。因此每天只要执行这个包就可将各维表进行更新。

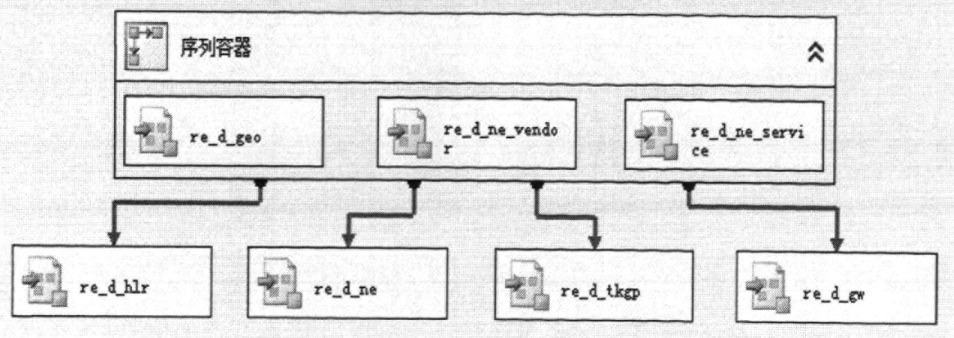


图 4.2 维表采集包的控制流设计图

## 2. 事实表的采集

通过一个代理包 `agent_etl_f.dtsx` 来实现各事实表的采集, 在这个包内调用其他各事实表的采集包进行各事实表的采集。其控制流设计如图 4.3 所示:

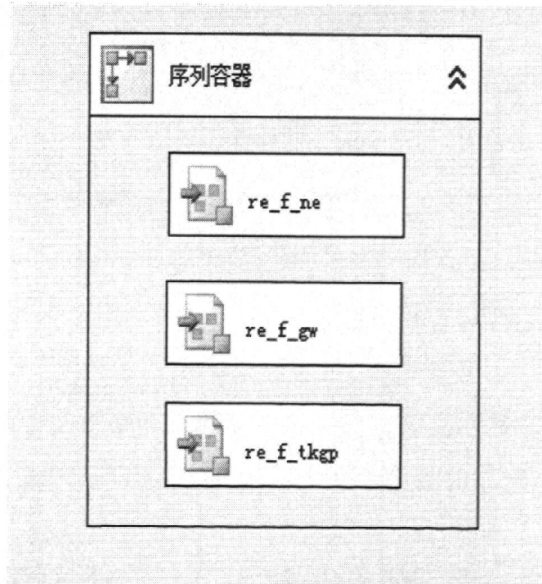


图 4.3 事实表采集包的控制流设计图

事实表的采集采用更新机制, 通过对比时间, 每天采集前三天的数据, 通过重复采集提高数据的完整性和准确率。

## 3. CDR 数据的采集

由于 CDR 数据来源于文本文件, 因此我们单独设计了一个代理包 `agent_etl_cdr.dtsx` 用来调用各采集包, 来实现对 CDR 相关维表和事实表的采集。其控制流设计图 4.4 为:

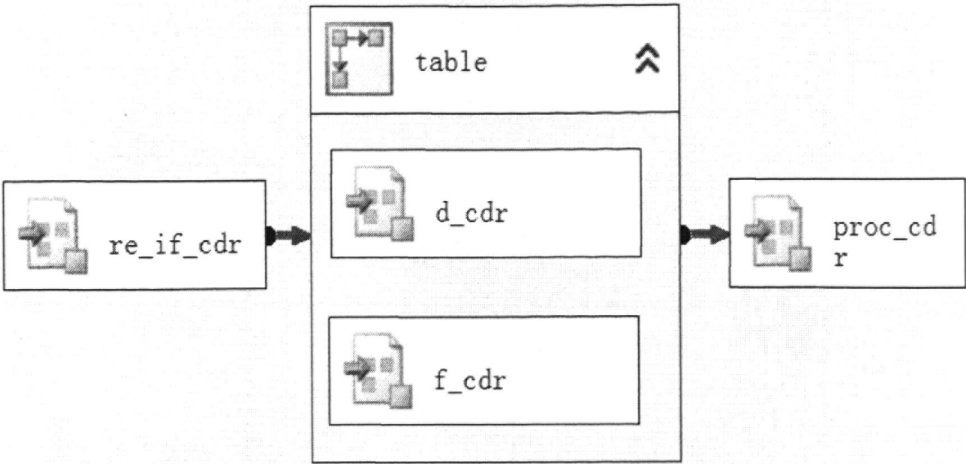


图 4.4 CDR 采集包的控制流设计图

4.2.2 定时调度的实现

每天对数据的采集和处理是通过 SQL Server 代理作业实现的。作业是一系列由 SQL Server 代理按一定的顺序执行的指定操作。一个作业可以执行多种操作，包括运行 Integration Services 包、Transact-SQL 脚本、查询或复制任务和 Analysis Services 命令等。作业能够运行重复的任务，并可以通过生成警报自动通知用户作业的状态，从而地简化了 SQL Server 的管理。

要使作业能够自动运行，必须要启动 SQL Server 代理服务，该服务是用于执行作业、监视 SQL Server 和激发警报的 Windows 服务。因此必须启动 SQL Server 代理服务，才能使本地或者服务器管理任务自动运行。

首先，我们在 SQL Server 配置管理器中启动 SQL Server Agent 服务，然后通过 SQL Server Management Studio 的对象资源管理器，创建 SQL Server 代理作业。我们创建两个作业：

1. 综合分析数据抽取处理调度任务

设定作业计划为每天 6 点执行，此作业包括有四个步骤，如下图 4.5 所示：

作业步骤列表 (J):				
步骤	名称	类型	成功时	失败时
1	维度表数据抽取	SQL Server Integration Services 包	转到下一步	退出报告失..
2	事实表数据抽取	SQL Server Integration Services 包	转到下一步	退出报告失..
3	维度数据处理	SQL Server Integration Services 包	转到下一步	退出报告失..
4	Cube数据处理	SQL Server Integration Services 包	退出报告成...	退出报告失..

图 4.5 综合分析数据抽取处理调度任务的作业步骤

## 作业步骤

步骤 1. 维表数据的采集, 调用包 agent\_etl\_d.dtsx。

步骤 2. 事实表数据的采集, 调用包 agent\_etl\_f.dtsx。

步骤 3. 维度数据的处理, 调用包 agent\_proc\_d.dtsx。

步骤 4. Cube 数据的处理, 调用包 agent\_proc\_f.dtsx。

## 2. 综合分析 CDR 抽取处理调度任务

设定作业计划为每天 7 点执行, 调用 agent\_etl\_cdr.dtsx 包, 实现对 CDR 相关数据的采集和处理。

多维数据集从数据仓库中获得数据, 需要经过处理才能生成各维度汇总的数据。数据仓库定期增量抽取数据源数据进行存储, 多维数据集也要定期进行增量处理, 才能将数据汇总成多维数据。

## 4.3 功能展现

整个系统的功能框架如图 4.6 所示:

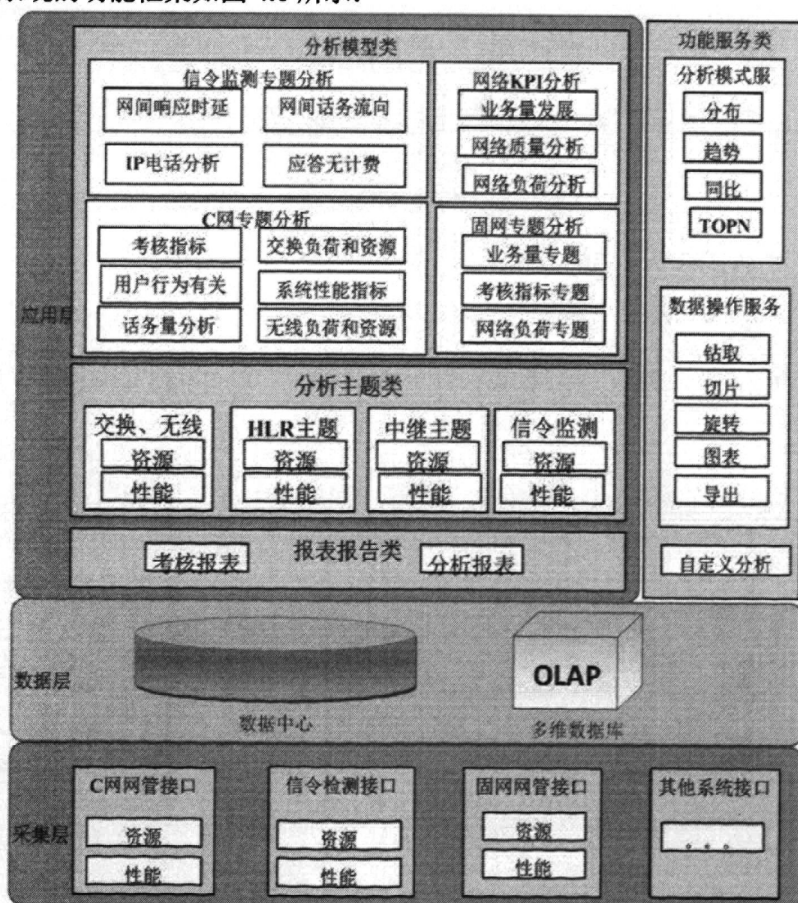


图 4.6 系统功能框架图

4.3.1 系统首页

在浏览器中输入电信综合话务分析系统服务器的登陆地址，进入登陆界面输入用户名密码，链接后自动呈现首页。首页呈现网络运行状况总体信息，可以通过配置文件来配置要呈现的信息。可以通过在主程序菜单中点击【首页】来呈现。系统主页如图 4.7 所示：

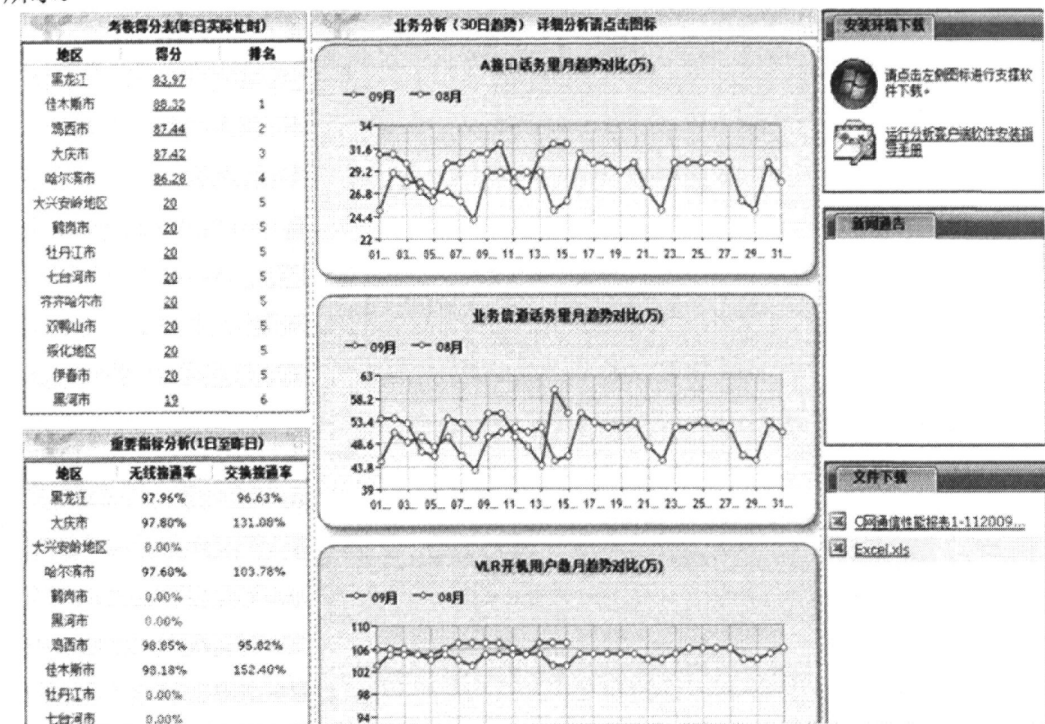


图 4.7 系统主页

信息发布首页为管理人员和网络维护人员提供统一的信息视图，比较全面了解网络发展和运行状况信息；首页通过 WEB 页面的方式向省/地市分公司管理层、规划部门、市场部门、客服部门等相关部门发布前一天移动网络运行的主要信息，让他们及时了解整个网络的整体运行情况，更好的起到网管系统的支撑作用。

通过首页能反映整个网络的整体运行情况，提供网络运行的概要信息，包括：考核得分表及详细列表、重要指标分析、业务趋势分析、专题分析以及公告新闻等信息。综合信息呈现提供对网络运行概要信息的进一步分析功能，并通过首页链接进入，如通过点击文字、图片直接进入相关模块。

4.3.2 报表分析

为报表分析人员、网优人员、监控人员、维护人员等提供有关 C 网和固网的相关指

标报表。以导航树图方式,按照分类列出报表,点击导航树图中的报表名称可以浏览每个报表。并且用户可以根据想要查看的内容设置报表查询条件:可以选择时间(日、小时、早晚忙时),可以选择地域条件(省和地区),可以选择网元粒度(地区, MSC, BSC)等,设定好后点查询按钮,即可呈现所要查询的数据。数据可以导出 Excel, 点保存按钮, 可以把打开并查询出的报表结果保存在 Excel 中; 总体呈现样式如下图:

代码	网元粒度	时间	C3.1 业务信号成功率 %	C3.2 掉话因子 %	C3.3 信道单元利用率 %	业务信
	黑龙江省	2010年08月18日	7.39%	76.25%	100.58%	
	大庆市	2010年08月18日	11.31%	74.23%	98.52%	
	大兴安岭地区	2010年08月18日	4.40%	51.71%	97.55%	
	哈尔滨市	2010年08月18日	8.91%	80.20%	99.63%	
	鹤岗市	2010年08月18日	8.32%	73.05%	99.24%	
	黑河市	2010年08月18日	5.62%	84.72%	98.61%	
	鸡西市	2010年08月18日	7.59%	83.99%	127.12%	
	佳木斯市	2010年08月18日	7.43%	78.69%	102.00%	
	牡丹江市	2010年08月18日	5.54%	73.99%	100.36%	
	七台河市	2010年08月18日	9.82%	84.80%	100.00%	
	齐齐哈尔市	2010年08月18日	5.14%	83.23%	99.37%	
	双鸭山市	2010年08月18日	6.23%	60.33%	100.27%	
	绥化地区	2010年08月18日	5.10%	78.34%	99.98%	
	伊春市	2010年08月18日	4.85%	47.21%	99.62%	

图 4.8 报表分析界面

### 4.3.3 网络状况视图

从宏观运营网络分析C网、固网重要KPI指标,便于用户全方位查看KPI指标的波动情况对比情况。以KPI指标为单位,能够实现不同时间粒度不同地域粒度KPI指标的同比分析、分布分析、环比分析、24小时趋势分析。根据分析的角度不同,分为日概、况月概括和详细分析,默认为日概况。具体如下图是A接口话务量的日概况分析。

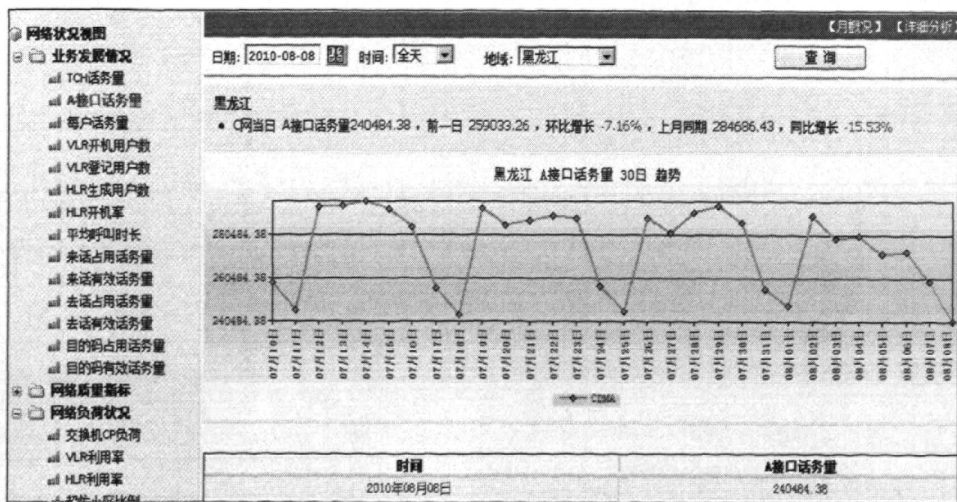




图 4.9 网络状况视图界面

日概况和月概况分析均支持指标环比，同比分析。环比是指与前一天做对比，同比指与上个月同一天作对比。显示结果以文字描述，文字显示模板如上图所示。支持指标趋势分析，以折线图显示近 30 日指标波动趋势。

4.3.4 专题分析

专题分析是我们根据用户的要求，将用户关心的KPI指标分成各个分析专题，便于分析人员根据通信网络的某些指标，有针对性的对网络的运营情况进行观察分析，找出网络存在的问题和隐患，为网络优化和发展提供决策支持。分析模式包括：分布分析、趋势分析、同比分析、TOPN分析。每次只能选择一种分析。如图4.10所示在左侧导航树图中是我们提前设计好的分析专题，从中选择分析专题，然后选择分析模式。

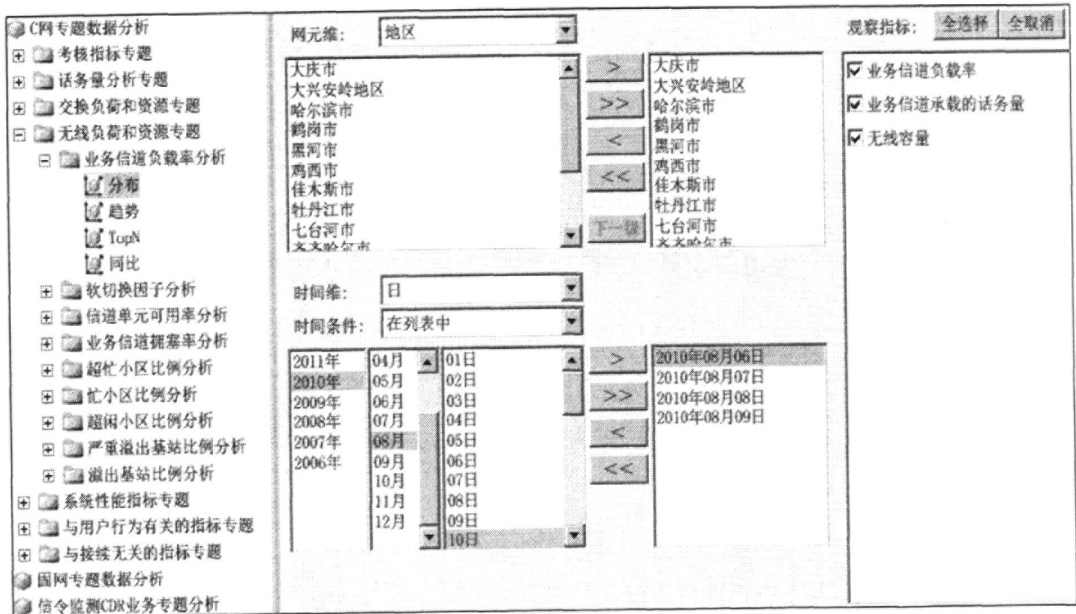


图 4.10 专题分析界面

分布分析

如上图4.10所示，分布分析的条件初始化。网元维选择‘地区’，然后选择全省的地区；时间维确定一种分析时间粒度，支持小时、日、周、月、年等各种粒度；选择‘日’粒度，选取四个时间点进行观察；观察选取的指标，单击‘执行查询’。

查询结果如下图4.11所示，以柱图和表格的形式直观地呈现所选指标在地理维度上的分布情况。在表格中选取业务信道负载率这一列，则柱状图呈现出各市业务信道负载率的分布情况。如在这一列中选择某一地区，在柱状图中会以红色区分其它地区。

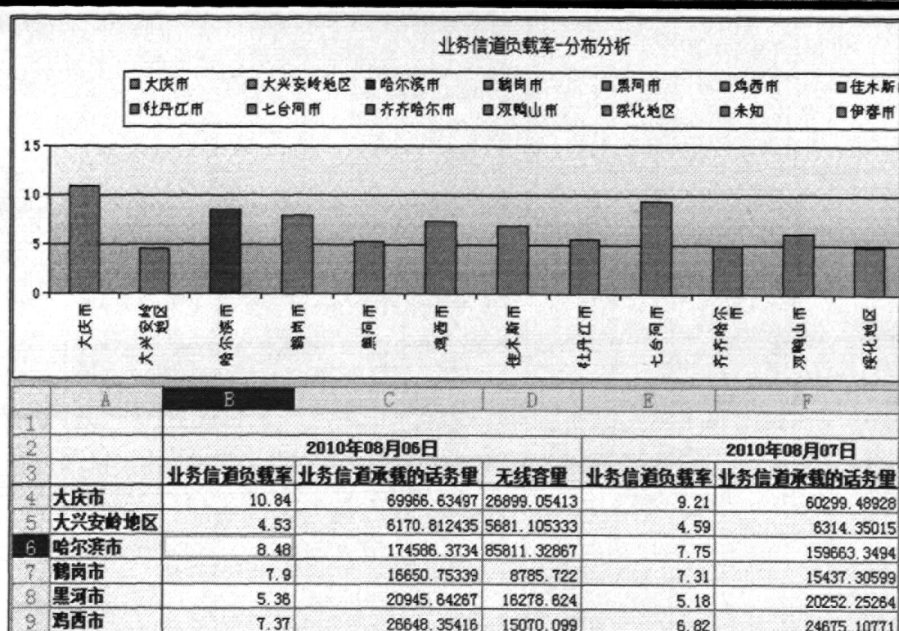


图 4.11 分布分析查询结果

### 趋势分析

趋势分析的查询设置如下图所示。网元维选择‘地区’；然后选择大庆和哈尔滨两个地区；时间维选择‘日’，时间条件选‘在范围内’，然后选择开始时间和结束时间；观察指标选取主题中的所有指标；单击‘执行查询’。

网元维: 地区 观察指标: 全选择 全取消

大庆市  
大兴安岭地区  
哈尔滨市  
鹤岗市  
黑河市  
鸡西市  
佳木斯市  
牡丹江市  
七台河市  
大兴安岭地区

>  
>>  
<  
<<  
下一级

大庆市  
哈尔滨市

时间维: 日

时间条件: 在范围内

开始时间: 2010年 04月 01日

结束时间: 2010年 04月 15日

☒ 业务信道负载率

☒ 业务信道承载的话务量

☒ 无线容量

图4.12 趋势分析查询条件设置

查询结果如下图。以曲线图的形式直观呈现指标在时间维上的趋势情况。图中我们选择了大庆市业务信道承载的话务量指标进行观察，曲线呈现出大庆市在2010年4月1号到15号的趋势。从曲线图上看4月3号到5号和10号、11号明显比其它天低，因为前三天是正好是清明放假，后两天是周末双休日。周末节假日业务信道负载降低。

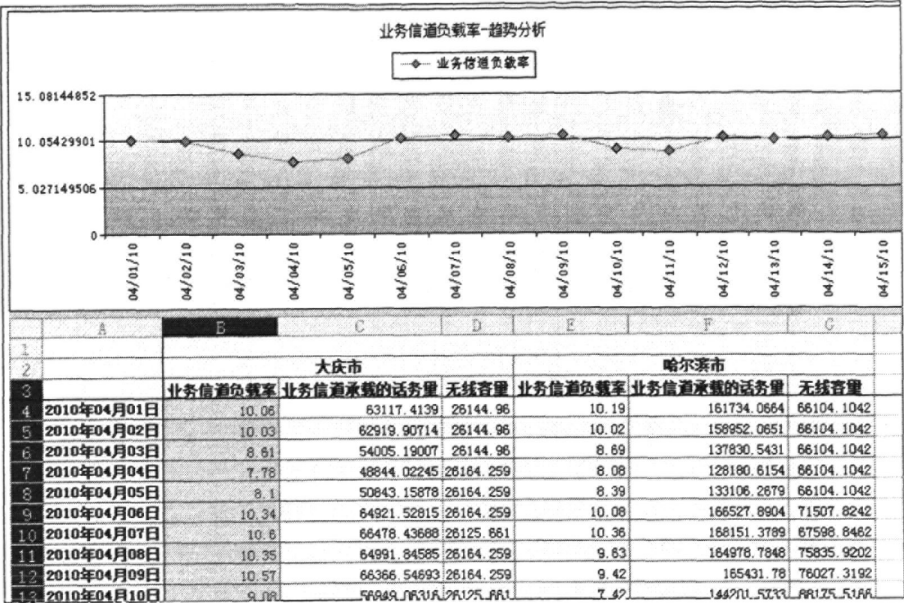


图 4.13 趋势分析查询结果

TopN分析

TopN分析的设置如图上图所示，网元维度我们选择‘地区’，然后选取所有地区；时间维选择‘日’，时间条件选‘在列表中’，然后选取其中一个小时；观察指标选取业务信道负载率，TopN个数选最大的10个，单击‘执行查询’。

网元维: 地区

观察指标: 全选择 全取消

大庆市  
大兴安岭地区  
哈尔滨市  
鹤岗市  
黑河市  
鸡西市  
佳木斯市  
牡丹江市  
七台河市  
双鸭山市

时间维: 日

时间条件: 在列表中

2011年 04月 21日  
2010年 05月 22日  
2009年 06月 23日  
2008年 07月 24日  
2007年 08月 25日  
2006年 09月 26日  
10月 27日  
11月 28日  
12月 29日  
30日

2010年08月28日

业务信道负载率  
☒ 业务信道承载的话务量  
☐ 无线容量

TopN个数: 最大 10

图 4.14 TopN 分析查询条件设置

查询结果如下图。以柱状分布图和表格的形式呈现业务信道承载话务量排在前十



位的地市。

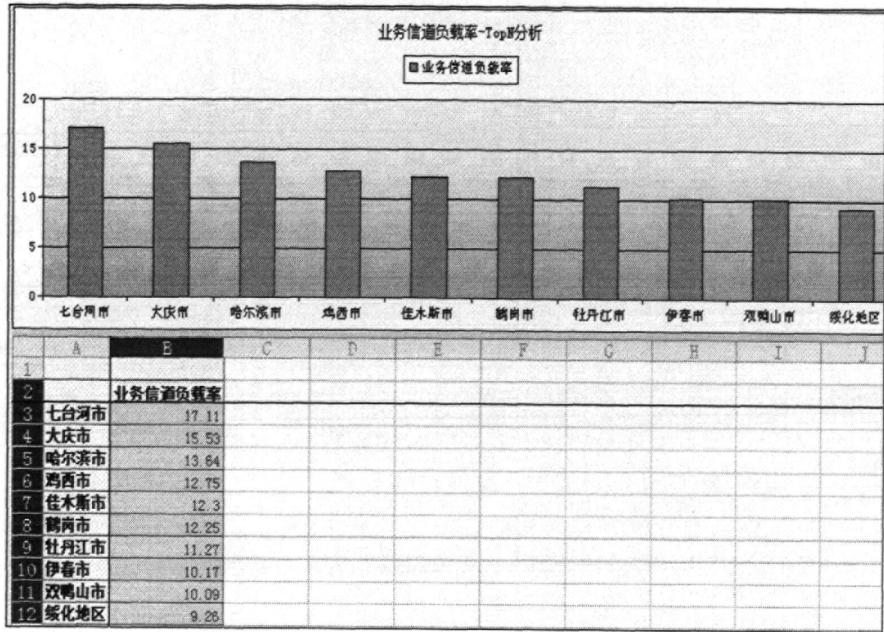


图 4.15 TopN 分析查询结果

同比分析

同比分析的分析条件如上图所示，网元维度选择‘地区’，然后选择在此维度下具体的城市；时间维选‘日’，然后选取两天作为同比日期；观察指标选取了业务信道负载率，执行查询。

网元维: 地区

观察指标: 全选择 全取消

大庆市  
大兴安岭地区  
哈尔滨市  
鹤岗市  
黑河市  
鸡西市  
佳木斯市  
牡丹江市  
七台河市  
双鸭山市

>  
>>  
<  
<<  
下一组

哈尔滨市

☒ 业务信道负载率  
☐ 业务信道承载的话务量  
☐ 无线容量

时间维: 日

同比日期: 2010年 08月 22日

同比日期: 2010年 08月 23日

图 4.16 同比分析查询条件设置

查询结果如下图所示，以曲线图的形式直观地呈现所选指标相对于前一天同一个时间点的对比情况。在此选取了哈尔滨市业务信道负载率来进行观察，以曲线图呈现哈尔滨市业务信道负载率在所选两天中的趋势，同时以柱状图来表示对应各时段的增幅情况。

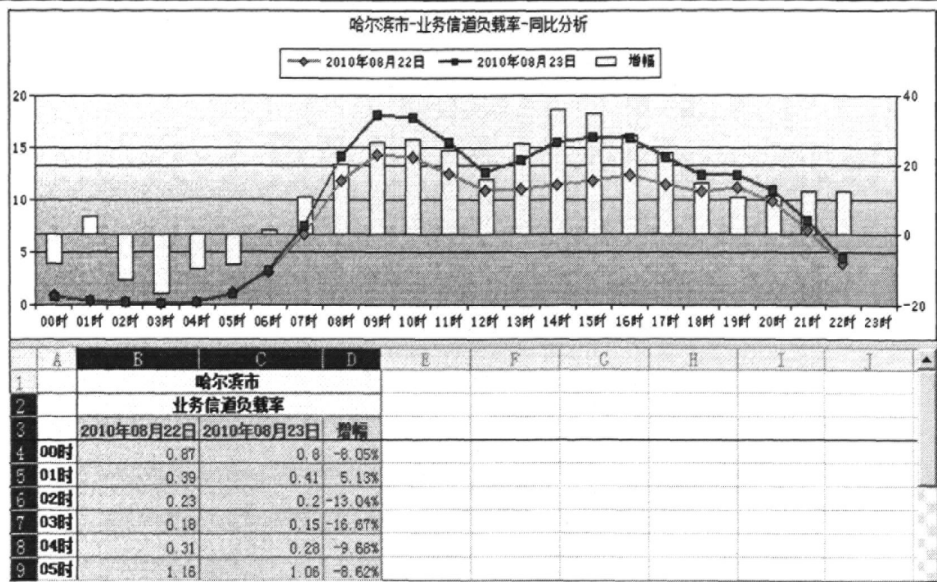


图 4.17 同比分析查询结果

4.3.5 自定义分析

在首页菜单中点击‘自定义分析’即可进入自定义分析模块。自定义分析功能模块界面由三部分构成：

- 1) 报表树和报表管理区
- 2) 专题元数据区
- 3) 分析报表定义区

界面如图所示：



图 4.18 自定义分析界面

报表树和报表管理区

报表树和报表管理区位于主界面的最左边；树图中列出了所有定义好的报表，包括永久保存在服务器上的报表模板和临时分析报表模板，同时提供新建报表，查看定义，执行报表，删除报表，保存报表等功能；

专题元数据区

专题元数据区位于主界面的中间部分，由一个下拉列表框和一个树图组成；下拉列表框中列出所有分析专题，可供选择；树图中列出分析专题相关的所有元数据，包括分析指标和分析维度信息，分析指标按类别放在不同的目录下。如上图所示

分析报表定义区

分析报表定义区用来定义报表的分析方式，分析内容，分析维度，过滤条件等。如下图所示。

请先选择分析模式

分布分析

▼

分析模式

分布分析：以柱图等方式呈现指标在地理等维度上的分布情况

请将度量字段拖到此框中

话务掉话比

度量框

请将行字段拖到此框中

省份

行轴属性框

请将列字段拖到此框中

列轴属性框

请将切片字段拖到此框中

切片属性框

请将维度属性拖至此窗口中来定义条件

条件定义框

执行查询

执行查询按钮

图 4.19 分析报表定义区

在自定义分析中，可以灵活地选择所要分析的指标以及分析维度，并且可以根据用户的需要，定义维度是作为行轴还是列轴。使用户能够灵活的从不同角度以不同形式来观察数据，从而更容易地发现网络中出现的问题。如下图是分布分析的查询结果。

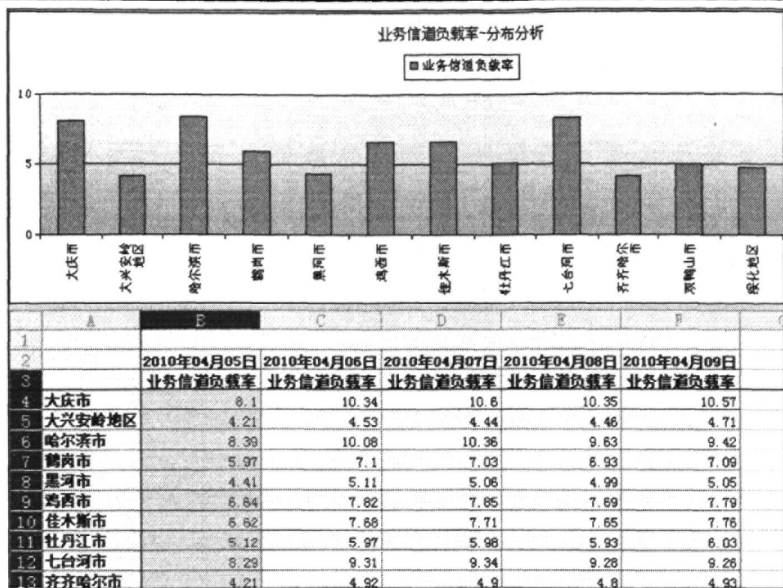


图 4.20 自定义分布分析的查询结果

默认情况下分布图呈现的是第一列的数据，各地区作为柱状图的横轴，当选中不同的列 柱状图也就呈现不同的数据。

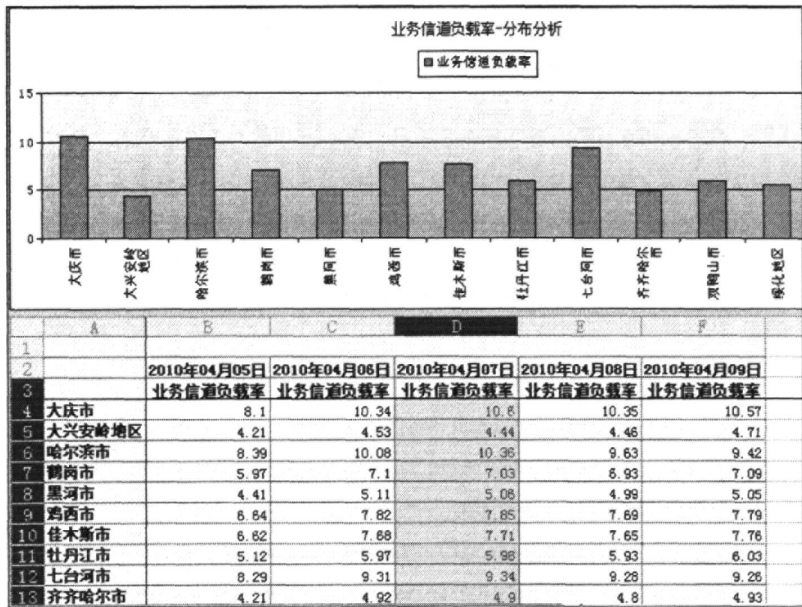


图 4.21 自定义分布分析的查询结果

自定义分析给用户极大的灵活性，观察指标可以根据需要随意拖拽，行字段和列字段也可以用户自行设定，可以帮助用户从不同的角度跟灵活观察数据

4.3.6 CDR报表分析

CDR(calling detail records)即呼叫详细记录，它描述了呼叫接续的全过程。在CDR中记录的参数来自于原始的信令消息数据，通过对记录中的一些重要参数进一步的分析

和处理，可以为固定电话网或移动电话网业务提供更全面的分析。

将鼠标放置CDR报表上即会出现下拉菜单，从下拉菜单中选择分析角度，然后设定查询条件，点击查询按钮就会呈现查询结果。如图是PCM统计分析结果图。



图 4.22 CDR 报表分析界面

### 4.3.7 系统管理

系统管理包括信息发布和用户管理。信息发布用来在首页上发布新闻通告和发布文档。用户管理主要是对用户进行管理，通过对用户设置不同的角色来限定用户的权限<sup>[39]</sup>。

角色管理的主要内容是对用户角色的权限进行设置，对于不同的角色可以看到不同的功能模块。管理员用户可以具有超级权限，对整个系统可以进行所有操作。其功能如下图：

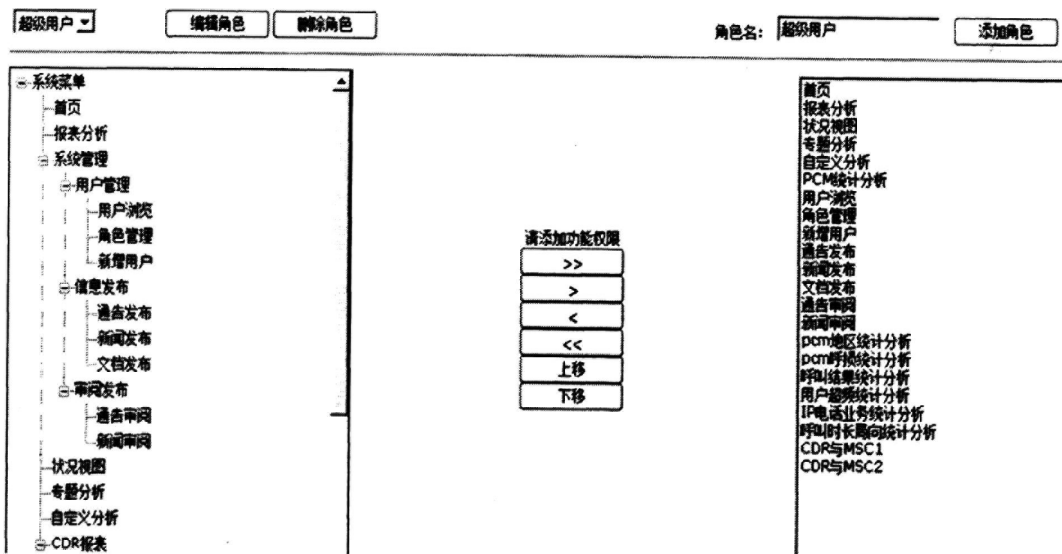


图4.23 用户角色的权限设置

通过设置用户角色的权限，不同级别的用户具有不同的权限，能够浏览的内容不同，

保证数据的安全。这对于系统中涉及到的一些重要的企业内部数据是非常重要的，限定数据的知晓范围，维护企业的商业利益<sup>[40]</sup>。

4.4 OLAP 分析的基本操作

1. 钻取

在同一个维度上，可以从一个层次上钻到更高的层次，查看上一层的汇总数据；或下钻到更低的层次，查看更细节的数据。通过上钻下钻操作，改变观察数据的层次，能够跟踪数据，定位数据，能够快速准确的找出问题所在的位置，为运维提供依据。如图所示是进行上钻和下钻的操作和结果

	2010年04月08日		2010年04月09日	
	交换机实际BHCA值	交换机主处理器负荷	交换机实际BHCA值	交换机主处理器负荷
大庆市	954,204.50	13	952,088.21	13.5
哈尔滨市	1,209,647.54	29	1,110,386.08	30
鸡西市	132,779.75	12	134,295	11.5
佳木斯	703,460.25	8	704,813.08	8
绥化地区	520,205.21	11.5	663,123.92	11.5

下钻结果

	2010年04月08日		2010年04月09日	
	交换机实际BHCA值	交换机主处理器负荷	交换机实际BHCA值	交换机主处理器负荷
HRBMSC1	593,683.25	27	600,773.96	28
HRBMSC2	289,018.17	29	290,987.08	30
HRBMSC3	326,946.13	12	218,625.04	9

图4.24 下钻操作及下钻结果

	2010年04月08日		2010年04月09日	
	交换机实际BHCA值	交换机主处理器负荷	交换机实际BHCA值	交换机主处理器负荷
大庆市	954,204.50	13	952,088.21	13.5
哈尔滨市	1,209,647.54	29	1,110,386.08	30
鸡西	132,779.75	12	134,295	11.5
佳木	703,460.25	8	704,813.08	8
绥化地区	520,205.21	11.5	663,123.92	11.5

上钻结果


	2010年04月08日		2010年04月09日	
	交换机实际BHCA值	交换机主处理器负荷	交换机实际BHCA值	交换机主处理器负荷
黑龙江	3,520,297.25	29	3,564,706.29	30


图4.25 上钻操作及上钻结果


2. 切片

支持灵活的切片操作，能够很方便地从不同的维度来观察数据，通过切片操作能够

滤除其它角度的影响，只在用户关心的维度上进行观察数据，从而更准确地分析数据。

日期: 2010-07-16 

时间: 全天 

地域: 哈尔滨市 


粒度: 地区 

图4.26 各种切片选项

3. 旋转

旋转分析通过改变维度的相对位置，改变数据的呈现样式。数据旋转可以把相同的数据以不同的方式呈现出来，在一张报表中重新安排行和列，这样可以达到从不同的角度观察数据。如图4. 26所示：

	2010年07月15日	2010年07月16日	2010年07月17日	2010年07月18日	2010年07月19日
	A接口话务量	A接口话务量	A接口话务量	A接口话务量	A接口话务量
大庆市	75202 71501	75669 51223	67685 85556	64738 44028	76087 0011
哈尔滨市	11929 19444	9350 972222	8659 444444	5391 333333	3501 638889
鸡西市	13083 86111	13077 80556	12022 55556	11657 55556	13347 02778
佳木斯市	66702 59861	67508 73195	62803 02611	60863 01832	69046 83834
绥化地区	100309 0922	103881 0194	92112 4625	89487 07167	108778 91

	大庆市	哈尔滨市	鸡西市	佳木斯市	绥化地区
	A接口话务量	A接口话务量	A接口话务量	A接口话务量	A接口话务量
2010年07月15日	75202 71501	11929 19444	13083 86111	66702 59861	100309 0922
2010年07月16日	75669 51223	9350 972222	13077 80556	67508 73195	103881 0194
2010年07月17日	67685 85556	8659 444444	12022 55556	62803 02611	92112 4625
2010年07月18日	64738 44028	5391 333333	11657 55556	60863 01832	89487 07167
2010年07月19日	76087 0011	3501 638889	13347 02778	69046 83834	108778 91

图4.26 数据的旋转显示

通过这些操作，能够让用户对数据进行多个角度的灵活的查询和分析，从而更深入的理解数据。

4.5 本章小结

本章主要介绍了综合分析系统的实现过程。简要介绍了系统配置和数据预处理的过程，然后说明了数据采集的实现过程，最后重点描述了分析功能和展现情况，展示了系统建设的成果。

## 结 论

随着通信技术的发展和通信规模的扩大,以及行业间越来越激烈的竞争,运营商需提高网络维护管理水平,这对通信网管系统提出了更高的要求。随着电信业竞争的加剧,电信运维必将走向集中监控、集中维护、集中管理的格局。逐步实现对各专业网络进行集中监控、综合分析等,使得网络管理机构相对集中,减少管理层次,实现网络快速的指挥调度。

话务综合分析系统运用数据仓库技术将各专业网管的数据有效集成和管理,解决了数据分散和历史数据堆积的问题,打破了各专业系统数据彼此孤立;利用OLAP技术进行灵活深入的查询分析,并通过WEB应用连接多维数据集将查询结果展现出来,满足多层次用户灵活的多视角的业务分析需求。本系统是以日常业务工作与网络运营为服务目标,面向多业务、多层次用户的综合性分析系统,辅助领导层进行网络维护决策,提高网络业务管理水平,提升网络运行质量,提高客户满意度,实现了以全局角度进行业务分析与决策支撑。

目前话务综合分析系统已经通过验收,并且运行良好。较好的满足了企业的现阶段需求,为数据分析和运维决策提供了有力支持。

该系统目前只提供了数据分析的功能,随着系统的二期建设,会增加数据挖掘的功能,使系统具备一定的发现数据潜在规律的能力;同时会接入更多的网管系统,像传输网管,电子运维系统等。从而形成有关整个通信网络的更全面的数据基础,更好的为电信决策服务。



## 参考文献

- [1] 百度百科. 中国电信产业重组 [OL]. <http://baike.baidu.com/view/1506352.htm?fromenter=%B5%E7%D0%C5%D6%D8%D7%E9>,2010.
- [2] 中国新闻网. 中国电话用户数达 11.3 亿[OL].<http://news.21cn.com/domestic/yaowen/2010/10/227893340.shtml>,2010.
- [3] W H Inmon.Building the Data Warehouse[M].John Wiley&Sons.Inc.,1993:10-20 .
- [4] Il-Yeol Song.ACM First International Workshop on Data Warehousing and OLAP [C]. New York, USA: ACM Press,1998.
- [5] Esteban Zimanyi.Proceeding of the ACM twelfth International Workshop on Data Warehousing and OLAP[C].New York, USA: ACM Press, 2009.
- [6] Joachim Hammer,Markus Schneider,Timos Sellis.Data Warehousing at the Crossroads [C]. Manifesto of a Dagstuhl Perspectives Seminar,Dagstuhl,2004.
- [7] Seungrahn Hahn, Ashvin Amin, Klenmens Dickhover.Getting Started with DB2OLAP Server for OS/390.International Business Machines Corporation, 2000: 1-18.
- [8] 金芳.浅谈数据仓库技术在电信行业中的应用[J].长春大学学报, 2009,19(12):68-69.
- [9] 陈志柏. 数据仓库与数据挖掘[M].清华大学出版社, 2009: 3-5.
- [10] 林宇.数据仓库原理与实践[M].人民邮电出版社, 2003:16-23.
- [11] 祖巧红, 高海耀, 王慧. 基于数据仓库的在线分析及其多维可视化研究[J].武汉理工大学学报, 2009, 31 (18): 108-111.
- [12] 齐红霞. 中国移动一级经营分析系统的 ETL 设计与优化[D].北京交通大学硕士论文, 2010.
- [13] 陈京民. 数据仓库与数据挖掘技术[M]. 电子工业出版社, 2002: 4-5.
- [14] Thilini, Hugh.J.Watson. Which Data Warehouse Architecture Is Most Successful[J]. Business Intelligence Journal,2006,11(1).
- [15] MatteoGolfarelli,stefanoRizzi.数据仓库设计: 现代原理与方法[M].清华大学出版社, 2010: 6-11
- [16] RalphKimball, MargyRoss.数据仓库工具箱:维度建模的完全指南[M].电子工业出版社, 2003.

- [17] 池太崴.数据仓库结构设计与实施[M].电子工业出版社, 2009: 46-49.
- [18] 李盛恩, 王珊.多维数据模型[J].计算机学报, 2005, 28(12): 2059-2067.
- [19] 张旭峰.ETL 若干关键技术研究[D].复旦大学博士学位论文, 2006.
- [20] Simitsis,Vassiliadis. A method for the mapping of conceptual designs to logical blueprints for ETL processes[J].Decision Support Systems,2008,45(1):22-40.
- [21] W.Liang,M.E.Orlowska,X.Yu.Optimizing multiple dimensional queries simultaneously in multidimensional databases[C].VLDB, 2000.
- [22] 张勇, 杨昆锦, 王文杰.移动经营分析系统中 ETL 的分析与设计[J].计算机工程与应用, 2006, 42(3):202-204.
- [23] E.F.Codd,S.B.Codd,C.T.Salley. Providing OLAP to User-Analysts: An IT Mandate[C]. White Paper. Arbor Software Corporation,1993.
- [24] 王珊. 数据仓库技术与联机分析处理[M]. 科学出版社, 1998: 87-89.
- [25] 朱建秋. OLAP 解决方案:创建多维信息系统[M].电子工业出版社, 2004.
- [26] T.LoukoPoulos, I.Allmad. Policies for Caching OLAP Queries in Internet Proxies. Digital Object Identifier.Issue,2006, 17:143.
- [27] 姚家奕.多维数据分析原理与应用[M].清华大学出版社, 2004.
- [28] 薛红,王敏. 基于 DW+OLAP+DM的超市销售决策支持系统[J].计算机工程, 2007: 64-67.
- [29] 樊同科.OLAP 在电信数据仓库中的设计与实现[J].电子设计工程, 2009, 17 (10) 114-115.
- [30] 姚军.数据仓库和 OLAP 技术在电信领域中的研究与实现[D].西北大学硕士论文, 2009.
- [31] 张曙明.论数据仓库的数据架构设计[J].信息通信技术, 2009 (06): 11-16.
- [32] 郑岩.中国电信领域商业智能的研究与展望[J].北京邮电大学学报, 2009, 11(6): 50 -55.
- [33] 徐媛媛. 构建高效的 3G 价值链[J].中国通信业, 2009, 103(07):67-69.
- [34] 余明辉, 胡耀民.基于 SQL Server 2008 决策支持系统模型的研究和应用[J].微计算机信息, 2010, 26 (2): 178-190.
- [35] Microsoft Developer Network. SQLServer2008 联机丛书[OL]. [http://msdn.microsoft.com/zh-cn/library/bb522498\(v=SQL.100\).aspx](http://msdn.microsoft.com/zh-cn/library/bb522498(v=SQL.100).aspx),2010.

- [36] 刘旭东. 数据仓库的 OLAP 技术研究[J].电脑知识与技术, 2005,(17).
- [37] Steve Goddard, Shifeng Zhang. A software architecture and framework for Web-based distributed Decision Support Systems[J]. Decision Support Systems, 2007, 43(4): 1133-1150.
- [38] 韩绍泽, 陈云秋, 常兴, 李军. 一种基于 B\_S 结构的监测系统研究[J]. 计算机与数字工程, 2010(1): 81-83.
- [39] 张明杰. 基于 Web 的数据仓库安全性研究[J]. 计算机与数字工程, 2010, 38 (7): 120-124.
- [40] 康懿. SQLSERVER 数据仓库的安全性[J]. 科技资讯, 2010, 33: 27.

## 攻读硕士学位期间发表的论文

- [1]. 姜延双, 和应民, 周春楠.数据仓库和 OLAP 技术在话务综合分析中的应用[J].  
信息技术, 2011.(已录用)

## 致 谢

首先,我要对我的导师和应民教授表示衷心的感谢。在研究生学习期间,和老师对我的学习给予了极大的鼓励和帮助。他朴素的工作作风,实事求是的态度给了我深深的影响。我从中学到了很多东西,更重要的是做人的哲理。这一切将使我受益终生。同时还要感谢学院全体老师对我的教导。

其次,感谢黑龙江亿阳信通的周总和钟总在实习期间给予我的关心和帮助;感谢辛巍以及公司的许多同事在技术上的帮助和指导,使我有了很大的进步。

最后,我要感谢我的父母以及亲戚朋友,在我的求学生涯中,他们给予了无数的支持和鼓励。