

运用数据仓库技术建设银行的 MIS

王晓林 (中国银行广东省分行信息科技处)

摘要:数据仓库理论是管理信息系统建设的方法论,中国银行省、市两级 FMIS 示范系统建设采用了数据仓库理论和多维数据库技术,本文通过对数据采集、数据仓库、数据呈现三个分系统的建设的介绍,阐述了中国银行省、市两级 FMIS 示范系统的整体构造和实际应用,提出了管理信息系统数据仓库的数据模型。

关键词:数据仓库 联机分析处理(OLAP) 联机交易处理(OLTP)

《中国银行省、市两级 FMIS 示范系统》是我们承接的国家“八五”重点科技攻关项目,该项目是运用了数据仓库技术和多维数据库软件工具,依靠广东省中国银行现有的网络和系统资源建设起来的,为中国银行省、市分行各级管理人员提供全面的管理信息服务。

一、FMIS 系统概述

中国银行省、市两级金融管理信息系统(FMIS)示范系统,是中国银行广东省分行承担开发的国家“八五”科技攻关项目。FMIS 是采用数据仓库技术,依靠广东省中国银行统一的基础系统网络,实现了 IBM ES9000、AS400、RS6000 和 PC 服务器等异种机的互联,衔接了现行的各业务系统,建立了集中式的数据仓库体系结构,初

步实现了管理信息系统的联机分析处理(OLAP)和动态报表的应用。该系统重点围绕中国银行资产负债管理的要求,建立覆盖全省 22 家分行的数据采集网络,初步实现了计算机业务系统数据和手工报表数据采集、存储的自动化,同时面向各级经营管理人员,开发出财务分析、业务管理、动态报表管理信息应用和办公自动化等功能。

1. 系统组成

FMIS 主要包括数据采集、数据整理、数据呈现三个分系统功能。其中数据采集分系统是采集现有联机交易系统的数据和补充的报表数据;数据整理分系统对采集来的数据进行分类、归纳、重组,下面称此为狭义的数据仓库功能;数据呈现分系统对数据进行联机分析处理,为用户提供了交互式的图形界面。FMIS 的功能组

成见图1。

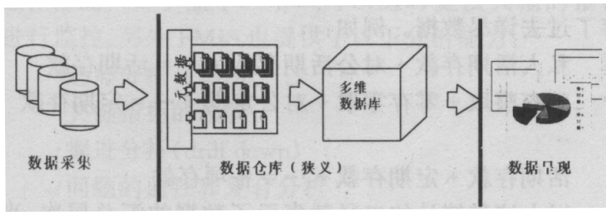


图1 管理信息系统的组成

2. 体系结构

FMIS以向中国银行省、市分行提供全面的管理信息服务为目标,因此FMIS系统要对广东省各分行的数据进行定期的采集和汇总,建立省行集中式的数据仓库。

图2为广东省中国银行FMIS系统结构示意图,FMIS是建立在一个由64KBPS DDN数据专线组成的全省通信网络基础上,通过支持SDLC、TCP/IP、NETBIOS多种通信协议下的局域网互联,组成全省的广域网,将省行及辖内21家地市分行的主机(ES9000、AS400等),微机系统互联,成为统一的客户/服务机系统环境,使面向用户的终端一机多能,可以同时运行多种应用,从而为数据的自动采集和数据仓库的共享奠定了基础。

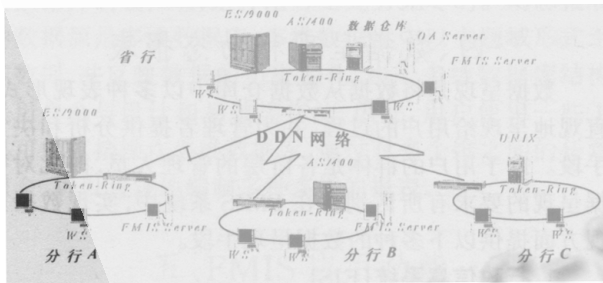


图2 广东省中国银行FMIS系统结构

二、数据采集

数据采集就是从分散的业务系统中将操作层面的数据加载到数据仓库里的过程。数据仓库中的数据来源于操作层面的各业务系统,例如:存款、信贷、会计、客户信息、信用卡等业务系统。数据采集分系统的主要任务是与各业务系统连接,对数据进行审核和转换,并将数据加载到狭义的数据仓库中。数据采集分系统,按照数据的来源及采集途径,制定了面向数据的采集和面向报表的采集两种方式。

1. 面向报表的采集

这种采集方式主要是解决一些非规范的业务系统(或业务系统未能提供)的数据的间接采集。在现阶段由于业务系统未能全部提供管理信息系统需要的数据,因此,本方法按照数据仓库的数据规范定义一系列的中间文件,由数据源的提供者将非规范的业务系统的数据按此格式进行转换或手工录入,通过采用工作组应用软件LOTUS NOTES,在全省网络的环境上将目前业务系统暂时未能提供,但又是管理信息所需要的数据以报表的形式报送到省行,如分行的员工人数、逾期贷款、中长期存款等数据,数据采集的周期根据是月。采集到的数据经过整理加载到省行的数据仓库中。

2. 面向数据的采集

这种采集方式是针对规范的业务系统,采用DRDA技术在各分行的主要业务系统进行数据的直接采集。面向数据的采集由财会报表应用系统(FRS)、MERVA应用软件实现。

广东省中国银行有5家业务系统在IBM ES9000上的分行,10多家在IBM AS400上的分行,FRS通过DRDA技术连接到业务系统上,将需要的数据从业务系统的数据库中下载到FRS系统中的DB2/2数据库中,并对数据进行合法性检查、平衡检查、报表衔接的检查等一系列的审核工作。基于全省的广域网络,省行电脑中心的FRS系统通过DDCS软件将分行的FRS的数据库中的数据自动地,定时采集到省行的数据仓库中,并进行数据的再审核及汇总等工作。现阶段的采集周期是月,逐步将采集周期缩短为旬,直至吸附在各业务系统上进行实时数据采集。

3. 数据的加载

数据经过转换和审核后,通过批量程序的定时将数据加载到狭义的数据仓库中。数据加载包括两个阶段,首先是数据加载到集中的关系数据库中的过程,其中包括当前详尽数据、过去详尽数据。其次是将需要汇总的数据从关系数据库中加载到多维数据库的过程。多维数据库中包含了轻汇总数据和高度汇总数据,多维数据库中的数据会被经常读取和进行历史分析与比较。如何提高优化程序的效率,缩短数据加载过程的时间是数据加载的主要目标,而检查加载后的数据的正确性是即关键又繁复的工作。

三、数据仓库

1. 数据仓库与OLAP

早期的数据库被理解为单一的、无所不包的中心数据集(群),数据库系统主要是服务于操作层面的应用,

通常是联机交易处理,人们关心的是响应时间、数据的安全性和完整性。信息分析层面所关心的是数据的历史分析、数据之间的因果关系和发展趋势等问题,因此信息分析层面的处理与操作层面的处理有本质的区别,主要区别表现为以下几点:

- 处理数据不同
- 支持技术不同
- 用户群体不同
- 处理特性不同

由于存在以上的区别,所以我们要采用数据仓库理论和 技术来进行管理信息系统的建设。

数据仓库中着重于解决在现代计算机系统环境下信息处理的体系结构的理论和技术,数据仓库是将管理和决策支持所需要的信息从日常交易产生的数据分离出来,将此分散的,难于访问得到交易数据转换为集中统一,随时可用的信息,并为此建立一种数据存储结构。数据仓库主要服务于信息分析层面的应用,通常是指联机分析处理,数据仓库具有以下四个特点:

- 面向主题(subject-oriented)
- 集约化(integrated)
- 非挥发性(nonvolatile)
- 时间序列(time-variant)

2. 数据仓库的结构

数据仓库中存放了经过数据采集来的大量数据,数据仓库在物理上是由数据仓库的关系数据表构成。FMIS的数据仓库(狭义)采用了IBM DB2关系数据库和COMSHARE ESSBASE多维数据库技术,这些DB2数据表结构是通过高级模型(又称ERD,实体关系层)、中级模型(又称DIS,数据项目集)、低级模型(又称物理模型)三个数据模型分析设计出来的。经过数据采集到的数据按照数据主题分门别类地存放在一百多个数据表结构中,这些表中包含了数据的生效日期和失效日期的属性。

数据仓库中的数据是按照处理的程度划分为不同的粒度层次,粒度水平的高低反映了数据汇总程度的高低。数据仓库中的数据包括有当前详尽数据、过去详尽数据、轻度汇总数据、高度汇总数据。FMIS中的数据有不同的数据采集周期,在每一个数据采集周期,随着新数据的加入进来,当前详尽数据就转化为过去详尽数据,并同时当前详尽数据按照时间、机构等主题域进行数据汇总,形成轻度及高度汇总数据。数据仓库中的数据采集了全辖22分行的当前详尽数据,这些数据加载到数据仓库时,会按分行、时间、数值、科目等路径进行汇总,在分行层次汇总得到每月、季、年的汇总数据;

在省行的层次按分行、货币、科目等路径汇总,成为全辖的按各统计口径的汇总数据,并进行了历史的积累,保存了过去详尽数据。例如:

私人活期存款 + 对公活期存款 + ... = 活期存款
 整存整取 + 零存零取 + 对公定期 + ... = 定期存款

...

活期存款 + 定期存款 + ... = 各项存款

以上这种统计的口径就表示了数据的汇总层次,当数据加载到数据仓库中的时候就按这个路径进行数据的汇总。

按照数据仓库的原理和OLAP处理准则,根据广东中国银行的计算机应用的系统环境,运用原形法构造了广东省中国银行FMIS的数据仓库的结构,见图3。

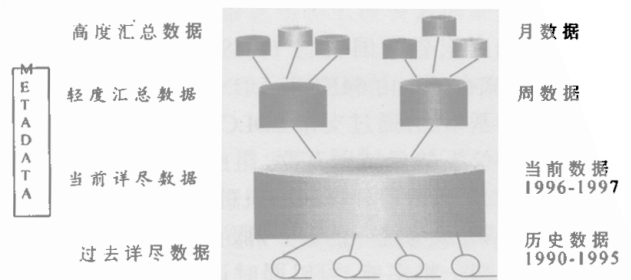


图3 FMIS数据仓库结构

四、数据呈现和动态报表

数据呈现是将数据从数据仓库中以多种表现形式,直观地呈现给用户的过程,并为管理者提供分析和决策手段。由于用户的群体是各阶层的管理人员,因此对数据呈现的要求有所区别。在FMIS系统中,实现数据呈现方面提供以下多种的数据呈现手段。

1. 行政信息系统(EIS)

EIS主要为行长及各级管理人员提供简洁、直观、易操作的图形用户界面,将管理者常用的功能,如资产负债管理、业务管理、汇率信息等功能配合以各种图形、报表和地图等形式表现出来。FMIS对一些数据进行了历史分析、分行比较分析和趋势分析,并对一些指标进行了定量分析,设置了安全警戒线,以不同的颜色表示指标的范围。例如:资产负债指标是衡量一个商业化银行运作好坏的一面镜子,我们开发了资产负债的25项指标的监控应用,对每一项指标设定了一个标准值。如果某个分行的某个指标的实际值超过其标准值,那么在地图上的这个分行区域就会显示为警告色(通常设置成红色),没有超标的分行区域就显示为安全色(通常设置成

绿色),这样管理人员就对全省各分行在某个指标下的运作情况一目了然了。用户可以随时更换其他的指标进行监控,另外 FMIS 也提供了以下分析能力:

- 趋势分析与诊断
- 关键指标的预警
- 掘进分析(drill down)
- 问题的监控和复合分析

2. 方法库

FMIS 采用了比较分析法、对比分析法、结构分析法、异常分析法、指标管理、趋势分析法、历史分析法、因素分析法等,并提供了方差分析、聚类分析、平滑指数法、回归分析、多元分析等十多种预测数学模型为管理人员的决策工具。

3. 动态报表

报表对管理者来说是不可缺少的,管理人员需要的报表变化多样,稳定性差,如果通过程序来编制报表,不但工作量相当大,而且需要经常修改程序,即使如此也难以满足管理人员的需求。

动态报表是通过一些软件工具与数据仓库的结合,通过用户对数据仓库中的数据进行随意的、任意的组合来生成报表的技术集成。FMIS 采用了个人制表软件(如:MS EXCEL 或 LOTUS 1-2-3)连接到多维数据库中,用户可以针对多维数据中维的任意组合得到所需要的数据并制成报表,这种组合是无穷无尽的。动态报表的数据源是多维数据库,多维数据库中以主题域形式组织数据,并反映数据的历史变化,由于多维数据库结构是用户参与设计和易被操作的,因此,用户使用一些工具可以直接操作多维数据库,通过对多个维之间的任意组合,可以直观地编制多种多样的报表。

五、FMIS 应用功能

FMIS 以银行的资产负债管理应用为突破口,开发了业务管理、行政管理、办公自动化、财经信息、决策支持和公共信息等应用和系统集成,见图 4。

存款业务管理包括存款期限分析、存款结构的变化及原因分析、存款利率分析、存款平均利率分析、存款大户情况分析。

放款业务管理包括放款期限分析、放款种类分析、押品分析、放款用途(行业)分析、行业分析、收息率分析、放款额度使用情况。

信用卡管理包括商户分析、客户分析、发卡量分析、交易量分析等。

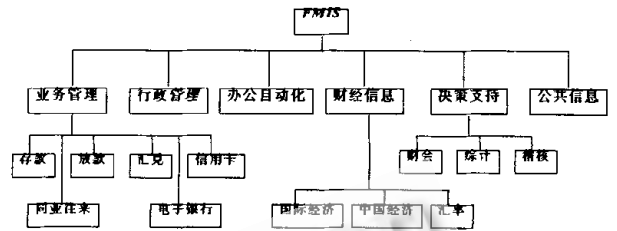


图 4 广东省中国银行 FMIS 应用功能

电子银行管理包括提供全国 18 家分行(含中银集团)的跨行 ATM 交易金额、交易笔数分析。对全国 18 家分行(含中银集团)跨行 ATM 交易不成功率及其原因的统计、分析和比较。

决策支持包括财会指标、综合指标资产负债表、财务收支情况和损益分析,提供了 25 项资产负债指标的管理和监控。

对以上各种管理应用,FMIS 提供分行比较、历史比较、掘进分析等功能。用户对选定的某些感兴趣的数据进行分行比较时,FMIS 按照分行数据值的大小进行排序,并以直方图形式显示出来;如果用户需要了解某个分行该数据在一段时间的变化情况时,FMIS 会对该行的该数据以报表及曲线形式表现出来;FMIS 也提供了对汇总的数据逐层掘进的分析功能,以发现问题的根源。

办公自动化系统采用了 LOTUS NOTES 工作组软件,提供試算表 LOTUS 1-2-3 或 MICROSOFT EXCEL 软件工具与 ESSBASE 多维数据库的结合,编制多维动态报表。在全辖各分行开通了电子邮箱应用,为数据采集提供了良好途径,与美联社系统相联提供公共信息、经济信息等服务。提供文字处理、电子简报、个人数据库、效率手册、电子邮箱等功能。

参考文献

- [1] W. H. Immon 《Building the Data Warchous》 John Wiley & Sons, INC.
- [2] W. H. Inmon & R. D. Hackathorn 《Using the Data Warehouse》 John Wiley & SOns, INC.
- [3] 《Data Warehousing-A Blueprint for Success》 SAS Institute Inc.

(来稿时间:1998 年 2 月)