

# 面向银行的一种中间件的前后处理过程<sup>①</sup>

陈继平 方旭升 (南京航空航天大学 经济与管理学院 江苏 南京 211100)

**摘要:** 银行风险数据处理中间件是一种主要采用数据集市技术来支持银行各种风险监管系统的高层应用中间件。本文首先探讨了提出银行风险数据处理中间件的客观原因和对其进行研究的意义,构建出了中间件系统的体系结构,详细描述了各主要功能模块,并着重设计了面向前端和后端的应用服务接口,最后给出了中间件系统的一个应用实例。中间件系统提炼和封装了共同的银行风险数据处理逻辑,为前端各种风险管理应用提供统一的服务接口和平台,大大缩小了银行风险管理系统的开发周期和开发成本。

**关键词:** 中间件; 数据处理; 数据集市; 银行风险; 接口

## Front-End and Back-End Data Processing of a Middleware for Bank

CHEN Ji-Ping, FANG Xu-Sheng (School of Economic and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China)

**Abstract:** Data processing middleware for banking risk is a high-level application middleware mainly using data mart technology to support the supervision system of banking risk. This paper firstly discusses the objective reasons for proposing the DPMFBR and the significance of studying it, then builds out the architecture of middleware system, gives a detailed description of the major functional modules, and focuses on designing the front-end and back-end application service interface. At last, an example of application of middleware system is given. Middleware system refines and packages the same logic of data processing for banking risk and provides a unified service interface and platform for various front-end risk management applications. It can greatly shorten the development cycle of banking risk management system and reduce the development costs.

**Keywords:** middleware; data processing; data mart; banking risk; interface

## 1 引言

随着加入 WTO 以及之后五年过渡期的结束,我国金融业已进入全面开放的新时期。银行业作为我国金融业的核⼼,面临着比以往更多的风险和挑⼚。加强银行风险监管,提高风险识别的能力和水平,变得愈来愈迫切。当前,每家银行均需按照银监会(银行业监督管理委员会)的报表模版要求,利用数据报送系统生成每期的报表文件,在规定的⼊间要求内打包上报银监会。除此之外,四大国有银行以及一些资金雄厚的股份制商业银行还部署了非现场监管分析系统,用来时刻掌握自身的风险状况。中国银行使用的早期风

险预警系统,更是进一步提升了银行风险的早期识别能力。然而不管怎样,这三类系统的运行都离不开银行的业务系统,最终展示出的供支持决策的风险分析结果都是用银行业务系统中的数据经过不同程度的加工处理形成的(如图 1 所示)。

往往一家银行所需的风险管理系统涵盖上述三类中的多个,如果每个系统都从原始业务数据库中提取数据进行加工,不仅会加重业务系统数据库服务器的负荷,干扰业务系统的正常运行,还会造成软件资源的极大浪费。针对这种情况,本文提出了银行风险数据处理中间件(Data Processing Middleware For

<sup>①</sup> 收稿时间:2010-04-12;收到修改稿时间:2010-05-22

Banking Risk, 简称 DPMFBR)的概念,研究和设计了该中间件系统的体系架构和工作原理,目的是为了能给银行的各种风险管理应用搭建一个系统平台(如图 2),减少风险管理系统的开发成本和复杂度,使银行的风险管理信息化能迅速地展开。

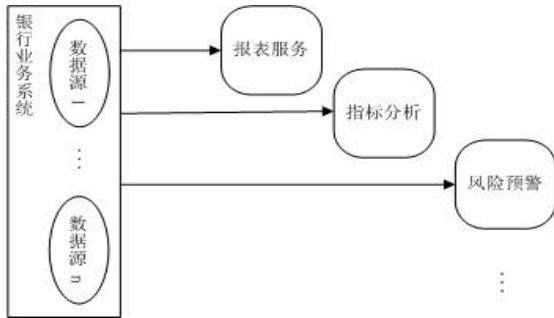


图 1 风险管理系统与银行业务系统的关系

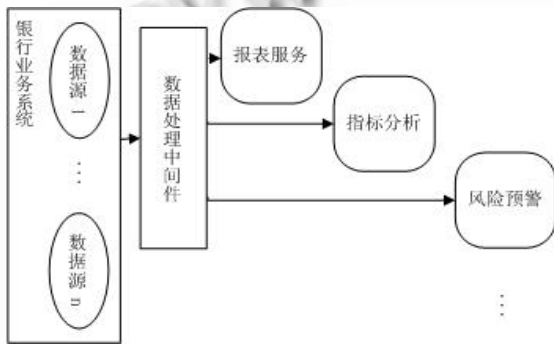


图 2 数据处理中间件的作用

## 2 相关研究工作

DPMFBR 主要是封装了各种风险应用的公共数据处理部分并对外提供统一的调用接口。这公共处理部分需要将下属各营业部门的交易数据、客户数据、经营数据等重要数据汇总到总部的主服务器上进行处理<sup>[1]</sup>。数据仓库、数据集市技术正好符合这种需求,它既能做到数据的集中处理,又可进一步提供多种灵活的数据分析方法<sup>[1]</sup>。选用专业的数据集成工具 DataStage,抽取、清洗、转换、加载(Extract-Transform-Load,简称 ETL)数据到数据集市,难点在于中间业务模型(数据集市存储模型)的设计上,要保证能提供前端风险分析应用所需的所有数据。本文的中间业务模型没有按照面向前端的各种风险分析主题来设计,而是按照面向后端的业务主题来设计

组织数据,这样中间业务模型能与银行的业务系统一一对应,很容易从中提取、加载数据,同时不会遗漏重要业务数据,从而保证中间业务模型的完整性。

另外,由于银行风险应用环境的复杂性,这包括它们使用不同的操作系统平台和实现技术,DPMFBR 采用 web 服务技术实现前端接口,从而能方便前端各种风险应用调用。

## 3 数据处理中间件体系结构和主要功能模块

### 3.1 体系结构

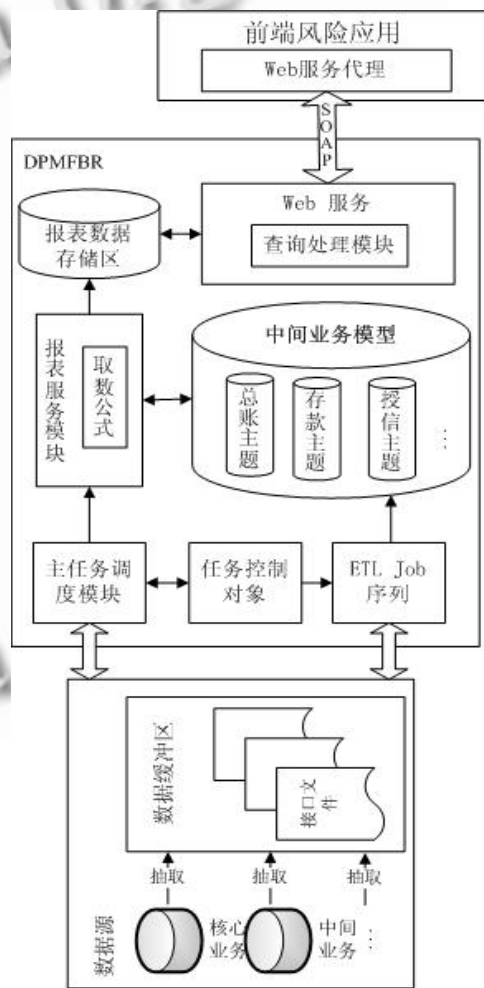


图 3 DPMFBR 体系结构

DPMFBR 可以部署在一台专门的中间件服务器上,监听前端应用发送来的请求,提供相应的数据服务。银行数据报送系统通过向 DPMFBR 不断发送获取某张报表的单元格值的请求,能快速自动生成报表,

大大减轻银行风险部门人员的负担。非现场监管分析系统发送指标的计算公式,从而能获得风险指标值。早期风险预警系统依靠中间件,可以快速展示银行的业务经营风险趋势。中间件体系结构如图3所示。中间件系统的功能可以分为两个部分<sup>[2]</sup>:后端数据处理部分和前端请求服务部分。

### 3.1.1 后端数据处理部分

为了避免中间件工作时对业务系统的影响,由各业务系统数据库服务器自身空闲时,将自己业务数据按照后端接口(后面详述)规定的接口文件格式,复制到数据缓冲区指定的目录中。

图中主任务调度模块是整个后台数据处理的入口主程序,负责启动整个后台应用。它按照一定的时间间隔监视数据缓冲区中的固定目录,如果发现有新的一批文件导入且已导入完毕,则新建一个任务控制对象。任务控制对象会启动相应的 ETL Job 序列,从接口文件中抽取、清洗、转换、加载数据到中间业务模型。ETL Job 序列执行完毕,任务控制对象会扫描 Job 运行过程中产生的日志信息(如图4所示),主要是通过日志记录的 Type 字段,检查 Job 运行有无出错,如果有一个 Job 发生错误,则回滚整个任务。任务控制对象返回任务执行结果。主任务调度模块获得执行结果后,销毁任务控制对象。

>Occurred	>On date	Type	Event
7:31:26 AM	8/27/2002	Control	Starting Job JobNames
7:31:26 AM	8/27/2002	Info	Environment variable st
7:31:26 AM	8/27/2002	Info	JobNames.GetNames:
7:31:26 AM	8/27/2002	Info	JobNames.GetNames:
7:31:26 AM	8/27/2002	Control	Finished Job JobNames

图4 ETL Job 运行日志

中间件系统的后台数据处理部分主要采用的是数据集市技术。中间业务模型是数据集市的核心,是从银行内部的不同系统收集数据,经过抽取、整合和统计,形成的一个中心数据集,是数据集市的主要数据源<sup>[3]</sup>,可以提供全面的银行信息。它设计的好坏影响到数据集市的最终物理实现。中间业务模型按照银行的业务主题进行设计和组织数据存储,分别为总帐主题、存款主题、授信主题、债券发行主题、同业往来主题、银行卡主题、衍生业务主题、债券头寸主题。

报表服务模块用于提高中间件系统的请求响应速

度。新一期的数据从接口文件中成功装载到中间业务模型后,主任务调度模块会启用报表服务模块。报表服务模块按照当期涉及到的报表的每个单元格的取数公式,从中间业务模型中取出数据,存储到如下关系的报表数据存储区中:

R(报表编号,单元格标识,机构,统计口径,年份,月份,数据值)。

### 3.1.2 前端请求服务部分

前端风险应用通过 web 服务代理,向 web 服务器发送 SOAP 信息查询请求。Web 服务器端程序获取请求后,调用查询处理模块中的语法分析器来解析数据请求表达式<sup>[4]</sup>,如果发现表达式不符合前端接口规范,则给予出错提示,否则从报表数据存储区中取出表达式中涉及到的报表的单元格数据,经过相应算术运算,将计算结果返回给请求客户。前台请求服务部分所需要做的处理工作非常简单,响应前端请求所需的时间就非常短,并减轻了中间件服务器的负荷,这都是因为大部分的数据处理操作都在后端已完成。

表1 接口文件区目录组织结构

目录结构	存放内容
{ETL_HOME}	接口文件缓冲区
-- app <sub>1</sub>	业务系统 1 文件接收目录
--daily	日变化数据文件接收目录
--monthly	月变化数据文件接收目录
⋮	
-- app <sub>n</sub>	业务系统 n 文件接收目录
--monthly	月变化数据文件接收目录
--daily	日变化数据抽取工作目录

## 3.2 中间件系统的接口设计

### 3.2.1 前端接口设计

银行风险报表的属性包括报表编号、报表版本号、统计口径。所以某张报表中的单元格,可以用报表的这些属性加单元格的行列信息来唯一指定,具体请求表达式设计成:报表编号\_报表版本号\_单元格行列\_统计口径,如 G2200\_0690\_E15\_1,这里把它称之为原子式<sup>[5]</sup>。

前端风险应用向中间件系统发送查询表达式来获取结果,把这查询表达式叫做组合式,它是由原子式

与运算符组合而成，具体定义如下：

(1) 运算符

- ① 常数项： $a, b, c, \dots$ ；
- ② 普通四则运算符，包括 $+, -, *, /, ()$ ；
- ③ 独特运算符两个， $[ ]$ 和 $*N$ 。“ $[ ]$ ”含义是求中括号中的组合式的本期值和去年同期值的算术平均数。“ $*N$ ”叫折年系数，含义是将折年系数之前的组合式的本期值  $v$  折算为一年的值，即  $v/\text{本期月份数} * 12$ 。

(2) 组合式的递归定义

原子式是组合式；

若  $A$  是组合式，则  $a+A, a-A, a*A, a/A$  是组合式；

若  $A, B$  是组合式，则  $A+B, A-B, A*B, A/B$  是组合式；

若  $A$  是组合式，则  $[A], A*N$  是组合式，其中  $*N$  必须出现在整个组合式的末尾。

前端风险应用通过调用 **web** 服务代理方法向中间件系统发送具体的组合式和期数，来获取服务器响应结果，如 **WebMethodProxy**(组合式,年份,月份)。

3.2.2 后端接口设计

后端接口的设计主要是对接口文件区目录组织结构的设计。合理地组织接口文件区的内容，可以方便 **ETL** 程序的处理。假设  $\{ETL\_Home\}$  代表接口文件区根目录，然后再按照数据来源进一步划分目录(如表 1)。其中日表化数据用来产生当日日报，主要是银行内部使用。月变化数据用来产生月报，用来上报给上级监管机构。所有接口文件区中的数据以 **csv** 文件格式存储，字段内容以逗号(,)分割。因为任何关系数据库表中的数据和银行经常使用的报表文件(如 **Excel**)都能很容易转换为 **csv** 文件。关系数据库中的表转换后以“数据表名.csv”命名传送至相应接受目录。如果该表无数据，则也同样传送该文件名称的空文本至文件交换区。同一批数据的 **csv** 文件到相应接口文件接收目录时，新建一个以日期命名的目录存放(yyyy-MM-dd)，并在传送完成后，在该目录下创建一个名为 **pass.end** 文件，文件内容为该批数据的产生日期，来告之中间件系统该批文件已经传完。

4 应用案例

本实验用某城市商业银行的 2009 年 12 月份的

2009年12月XXXX银行指标分析表				
指标名称	类内汇总数据			
	本期值	比上期	比上年初	比上年同期
资产类准备金充足率	141.94%	↓20.30%	↓20.34%	↓20.34%
贷款类准备金充足率	141.94%	未报送	↓29.62%	↓29.62%
不良资产率	0.21%	↓0.01%	↓0.38%	↓0.38%
不良贷款率	0.44%	未报送	↓0.32%	↓0.32%
流动性缺口率	-16.50%	↑23.53%	↓19.25%	↓19.25%
人民币超额备付金率	3.53%	未报送	↓3.97%	↓3.97%
流动性比例_本外币合计	70.20%	未报送	↑5.37%	↑5.37%
流动性比例_外币	97.96%	未报送	↓150.36%	↓150.36%
流动性比例_人民币	70.12%	未报送	↑5.72%	↑5.72%
资产类准备金充足率	5.90%	↓0.30%	↓1.18%	↓1.18%
贷款类准备金充足率	9.72%	↓0.02%	↓1.95%	↓1.95%
净稳定资金比例	10.29%	↓17.04%	未报送	未报送
流动性覆盖率	0.39%	↓0.69%	未报送	未报送
净息差	0.03%	↓1.32%	未报送	未报送
成本收入比率	51.02%	↑10.15%	↑10.56%	↑10.56%
总行资产利润率	0.73%	↓1.16%	未报送	未报送
总行利润率	10.29%	↓17.04%	未报送	未报送
资产利润率	0.39%	↓0.69%	未报送	未报送
存款集中度	64.57%	未报送	↑16.05%	↑16.05%

图5 指标分析结果

(下转第163页)

(上接第 108 页)

业务数据,按照 DPMFBR 的后端接口规定的目录结构组织方式,将其放置到接口文件缓冲区月变化数据文件接收目录中。DPMFBR 监测到缓冲区中有新数据导入,会启动 ETL 程序将导入的业务数据经清洗、转换处理后加载到中间业务模型,然后再经过报表服务模块的进一步处理。该城市商业银行设计的需观察的公共指标数共 53 个,监测的风险涉及流动性风险、市场风险、信用风险、资本充足、盈利性等。根据中间件的前端接口规范,每个指标都有自己的组合式,通过向中间件发送请求,可获得如下指标分析结果(如图 6 所示)。每个指标名称右侧的连续四列,分别是指标的本期值、比上期变化值、比年初变化值、比上年同期变化值。从图中可以看出指标的值基本上都能从 DPMFBR 中得到,从而证明本文研究的 DPMFBR 是切实有用的。

## 5 结束语

本文针对银行业的风险管理信息化情况,提出了银行风险数据处理中间件的概念,设计了中间件系统

的体系结构、工作原理和对外接口,最后用某城市商业银行的前端风险指标分析应用实例证明了中间件系统的可用性,望能给银行风险管理信息化的开展带来一定的参考价值。由于各金融机构的业务数据源的不一致性,导致中间件系统的通用性受到局部限制,后端接口的设计以及 ETL 程序需要针对各银行定制,这方面的研究有待进一步深化。

### 参考文献

- 1 朱玲凡.基于数据仓库的金融统计信息平台及其应用研究[硕士学位论文].南京:南京航空航天大学,2009.
- 2 王念滨,宋益波,姚念民,刘大昕.一种基于群集的并行数据处理中间件.计算机研究与发展,2007,44(10):1702-1708.
- 3 苏新宁.数据仓库和数据挖掘.北京:清华大学出版社,2006.4.
- 4 安然,陈驰,徐震.数据库加密中间件的设计与实现.计算机工程与设计,2009,30(14):3261-3265.
- 5 耿素云,屈婉玲,张立昂.离散数学.北京:清华大学出版社,2008.