

文章编号: 2095-2163(2020)01-0211-04

中图分类号: TP311.13

文献标志码: A

ETL 技术在银行成本分摊系统数据处理中的应用

杨政, 康磊

(西安石油大学 计算机学院, 西安 710065)

摘要: 银行成本分摊系统中数据处理尤为重要, 本文针对银行成本分摊系统中数据处理, 采用 ETL 数据处理技术从多数据源或子系统中获取数据, 对其进行抽取、转换、装载, 规范数据格式, 满足数据处理的需求。

关键词: 银行; 成本分摊; ETL

Application of ETL technology in data processing of bank cost sharing system

YANG Zheng, KANG Lei

(School of Computer Science, Xi'an Shiyou University, Xi'an 710065, China)

【Abstract】 Data processing in the bank cost sharing system is particularly important. In this paper, the data processing in the bank cost sharing system uses ETL data processing technology to obtain data from multiple data sources or subsystems, extract, convert, load, and standardize the data format, which could meet the requirements for data processing.

【Key words】 bank; cost sharing; ETL

0 引言

成本分摊是指将某成本中心在特定会计期间内发生的各项成本, 以合理的原则向受益的成本中心、利润中心、产品等成本接收对象进行分摊, 以充分反映这些成本接收对象所应该承担的成本信息。

银行成本分摊需分摊的成本包括人员费用、运营费用、业务费用、行政费用、税费及其他费用等。各项费用根据其业务的关联程度可划分为直接费用和间接费用。直接费用是指与产品销售或业务活动直接关联的费用支出; 间接费用是指那些与产品销售或业务活动无直接关联, 但是间接支持了产品或业务活动正常运转的费用支出。

1 银行成本分摊系统数据部分组成

银行成本分摊系统组成如图 1 所示。该系统由 5 大模块构成, 对此可做阐释分述如下。

(1) 公共管理模块: 完成对系统所涉及的基本档案、影响因素、映射规则的统一管理;

(2) 责任会计模块: 完成系统所需要的责任凭证管理, 实现内部结算;

(3) 成本分摊计算模块: 完成成本分摊的方案设置, 实现日常计算执行;

(4) 盈利性核算模块: 完成账户级 EVA (EVA 是英文 Economic Value Added 的缩写, 一般译为附加经济价值, 这是全面衡量企业生产经营真正盈利

或创造价值的一个指标或一种方法) 指标所涉及的各项收入、支出的计算, 实现各个维度的 EVA 汇总计算;

(5) 盈利性分析报表模块: 提供满足日常经营管理应用需要的成本分摊及盈利性分析报表, 提供给全行不同层次的报表使用者。

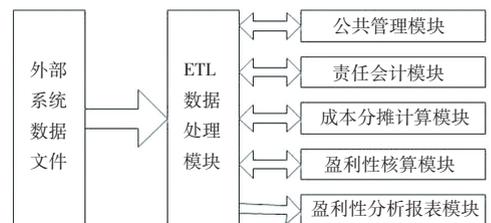


图 1 成本分摊系统组成

Fig. 1 Cost sharing system

除上述模块以外, 银行成本分摊系统离不开数据, 与数据相关的模块有 2 个。这里推得设计表述如下。

(1) 外部系统数据文件: 根据成本分摊外部数据接口规范的约定, 由源系统或子系统通过源数据向成本分摊系统提供数据, 以文件的形式; 外部系统即源系统或子系统, 如人力资源系统、内部资金转移定价系统等。

(2) ETL 数据处理: 成本分摊系统接受加载外部数据后, 根据自身需要进行 ETL 数据加工处理,

作者简介: 杨政 (1994-), 男, 硕士研究生, 主要研究方向: 计算机系统结构、数据库数据处理; 康磊 (1968-), 女, 副教授, 主要研究方向: 计算机体系结构。

收稿日期: 2019-09-28

加工过程中将利用设置的相关映射规则,为成本分摊计算模块、盈利性报表模块等分别分步提供数据。

2 ETL 模块处理

各个子系统将数据推送到成本分摊系统中,得到源数据,成本分摊系统对数据进行 ETL 模块处理,也就是将数据进行抽取、转化、加载到成本分摊系统中,对成本分摊的后续计算模块和报表模块提供数据支持。

2.1 数据的来源

成本分摊系统的数据来源分布在各个子系统中,需要人力资源系统、内部资金转移定价(FTP,下文 FTP 均指资金转移定价)系统、增值税系统、财务系统等系统提供数据。采用 ETL 技术将所需数据从各系统中抽取、转换、加载到成本分摊系统中,具体如图 2 所示。

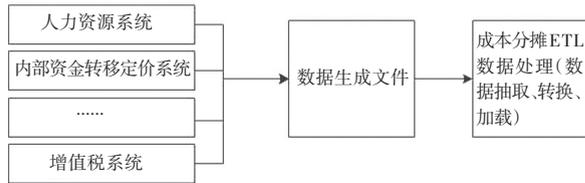


图 2 子系统与成本分摊系统数据处理关系

Fig. 2 Data processing relationship between subsystem and cost sharing system

由图 2 可知,人力资源系统、内部资金转移定价系统、增值税系统等子系统将数据以文件格式推送到成本分摊系统中,成本分摊系统中进行 ETL 数据抽取、转换、加载,将数据格式转换成本系统需求的数据格式。人力资源系统为成本分摊系统提供机构信息、部门信息、人员信息、薪酬信息、人员费用、运营费用、业务费用、行政费用等数据,FTP 系统为成本分摊系统提供定价收入、定价支出等数据,增值税系统为成本分摊系统提供增值税数据等税费。

因为现有业务数据源或子系统多,如何保证数据的一致性,真正理解数据的业务含义,跨越多系统整合数据,最大可能提高数据质量,迎合业务需求不断变化的特性,是 ETL 技术处理的关键。

2.2 数据的抽取

数据的抽取是数据进入数据库的入口。由于数据库是一个独立的数据环境,需要通过抽取过程将数据从业务系统、外部数据源、脱机的数据存储介质中导入到成本分摊的数据库中。

数据抽取方式是:各子系统生成数据文件放在指定服务器路径下,然后通过 XManager 软件登录到对应服务器路径下,通过文件传输功能,获取到成本

分摊服务器路径下。

数据从其他外围系统或子系统中获取,外围系统列表举例见表 1。

表 1 外围系统列表举例

数据来源	表名	表中中文名
中间业务系统	ZZWJ	总账文件表
内部资金转移定价系统	FTP_RE_RST_FTP	FTP 结果集表
增值税系统	PT_INOUTFLOW	价税分离交易明细数据
人力资源系统	TBL_EMPLOYEE	员工表
...

由表 1 举例可知,资金转移定价系统中结果集表可以获得余额年累计数、当前余额等数据,人力资源系统中员工表可获取员工名称、薪资等数据。外围系统或子系统将运行所需表数据生成多个文件,推送到某一服务器路径下,成本分摊系统通过 XManager 软件登录到对应服务器路径下,通过文件传输功能,将每个文件获取到成本分摊服务器路径下。然后编写脚本,此脚本功能为将从外围系统或子系统抽取过来的数据文件加载到相对应的事实表中。外围系统或子系统提供的数据文件数据都是以‘|’分隔,所以在此时需要将文件中数据加载到表中才能使用,因此需编写相对应程序脚本才能完成这一步,到此时数据抽取才能完成。

2.3 数据的转换

数据转换是真正将数据源变为目标数据的关键环节,只有将原始数据转换成成本分摊系统定义的标准才能参与成本分摊计算,如果数据格式不准确,将会影响计算结果。根据数据调研结果,确认内容见表 2。

表 2 数据来源及字段

系统名称	数据颗粒度	字段
核心系统	账户级明细数据	存贷款时点余额、日均余额、利息收入、利息支出
FTP 系统	账户级明细数据	FTP 定价收入、FTP 定价支出(定价结果集)
增值税系统	账户级明细数据	增值税(含税金额、不含税金额)
中收业务(总账)	科目级明细数据	手续费及佣金收入和支出
财管系统	科目级明细数据	费用(凭证表 gl_detail, gl_voucher)
人力 HR 系统	账户级明细数据	机构、部门、人员、薪酬信息
总账系统	科目级明细数据	业务状况表(用于数据核对)

由表 2 可知,第一列是获取数据的源系统;账户级明细数据是指每个表的每条数据都应该是一条条账户信息,比如一个人的一个账户信息,科目级数据是指确认到每个科目的粒度信息;字段是从这个系统中需要获取到的哪些字段。

在数据抽取结束后,获取到最基本的源数据。如图 3 所示。

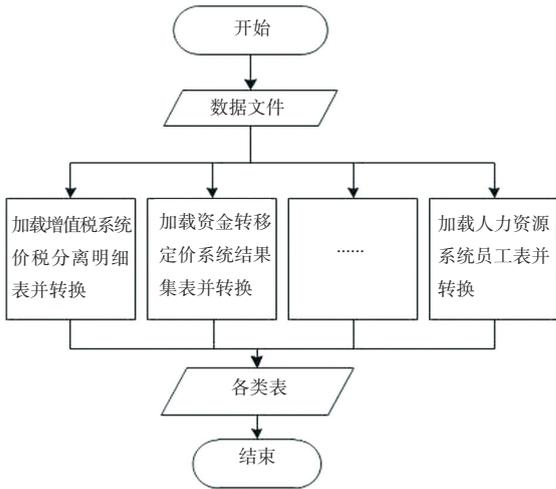


图 3 数据加载转换处理

Fig. 3 Data loading conversion processing

由图 3 可知,在 Oracle 环境下编写存储过程。存储过程将获取到的源数据通过定义的规则和要求转换成成本分摊系统的数据格式。加载增值税系统价税分离交易明细数据,然后获取增值税系统价税分离交易明细数据中的增值税字段,为后续报表计算提供数据支持;加载人力资源系统员工表,对员工表数据转换,比如部门 ID 字段, ID 字段会被加密成一段无意义字符,这时就需要转换成定义好的规则。部分字段转换规则见表 3。

由表 3 可知,将数据转换成某种占比后,将参与到成本分摊的计算中,会根据不同的占比进行费用分摊,这些占比数额将决定最终分摊数额的数额。

2.4 数据的装载

成本分摊系统的数据处理中的关键一步就是数据装载。在进行数据转换后,通过存储过程已经将

表 3 部分字段转换规则

Tab. 3 Partial field conversion rules

营业毛利占比	利息收入-利息支出+FTP 转移收入-FTP 转移支出
账户数占比	所有帐户默认为 1
产品规模占比	包括存款、贷款、资金、理财、票据等业务的月均余额
存贷款规模占比	包括存款,贷款业务的月均余额
贷款产品规模占比	贷款产品月均规模
风险加权资产占比	根据不同类别资产的风险性质确定不同的风险系数,以这种风险系数为权重求得的资产
战略系数占比	30% 营业毛利+70% 产品规模占比
不良贷款占比	不良贷款月日均余额,按照贷款五级分类中的次级、可疑、损失为不良

对应数据装载到表中,如图 3 所示。数据的加载就是将各个子系统的数据文件通过脚本文件加载到对应表里面,或者将数据转换后的结果插入到表中。

对于数据量小且数据变化频率较低的数据文件采用软件工具经过必要的转换后装载到相应的事实表中;对于数据量大的数据先用软件工具生成每日的增量数据,再将数据插入/更新到相应事实表中。

数据转换结束,成本分摊计算完成,生成结果集,将结果集汇总,最终将数据加载到成本分摊报表中。

2.5 数据的校验

在数据加载完成后,需要对加载后的数据进行校验。数据加载后的校验是对加载质量的检查,同时数据校验的结果也是判断新系统能否正式启用的重要依据。

加载后的数据校验的指标主要包括五方面:完整性检查,引用的外键是否存在;一致性检查,相同含义的数据在不同位置的值是否一致;记录条数检查,检查新旧数据库对应的记录条数是否一致;特殊样本数据的检查,检查同一样本在新旧数据库中是否一致。

在成本分摊报表部分,会检查数据展示信息,同时对数据的格式以及数额进行检查和校验,详见表 4。

如果查出有什么问题及时在存储过程中进行改正。

表 4 成本分摊报表部分数据

Tab. 4 Part of the cost allocation report data

资产余额	资产日均	贷款 FTP 净利润		负债余额	负债日均	存款 FTP 净利润	
		利息收入	FTP 转移支出			FTP 转移收入	利息支出
该期间最后一天时点余额	期间每日累计数/期间天数	本支行发生的贷款业务利息收入汇总	本支行发生的贷款业务 FTP 转移支出汇总	该期间最后一天时点余额	期间每日累计数/期间天数	本支行发生的存款业务 FTP 收入汇总	本支行发生的存款业务利息支出汇总