

基于数据仓库的行政事业单位建设项目内部控制评价

——以重庆海事局为例

程平^{1,2}(博士生导师), 彭兰雅^{1,2}, 郭奕君^{1,2}

【摘要】以重庆海事局为例,根据财政部《行政事业单位内部控制规范(试行)》和重庆海事局《长江海事局基本建设管理办法》等文件,设计建设项目内部控制评价的数据仓库体系结构,分析相关的数据源及数据采集方法,详细阐述Hive数据仓库的设计,包括建设项目内部控制的主题域确立、粒度设计,以及概念模型、逻辑模型和物理模型的设计等,并基于数据仓库,运用HiveQL语言对建设项目内部控制进行查询分析,展示基于数据仓库的建设项目内部控制评价应用实施方法。

【关键词】行政事业单位; 建设项目; 数据仓库; 内部控制评价

【中图分类号】F232; TP315 **【文献标识码】**A **【文章编号】**1004-0994(2019)15-0059-6

目前行政事业单位建设项目内部控制评价的相关研究主要集中在内部控制评价指标构建以及内部控制评价方法方面,而关于行政事业单位内部控制评价信息化的研究较少,基本不涉及基于数据仓库的建设项目内部控制评价研究。鉴于此,本文以重庆海事局为例,根据财政部《行政事业单位内部控制规范(试行)》(简称《内控规范》)和重庆海事局《长江海事局基本建设管理办法》(简称《管理办法》)的要求,展现基于数据仓库的数据分析方法在建设项目内部控制评价中的具体运用。

一、建设项目内部控制评价数据仓库的体系结构设计

(一)设计思路

重庆海事局从2016年开始实施《管理办法》,规范了基本建设(含船舶建造)工作程序和管理行为,以及建设项目的工程立项、初步设计、招投标、工程施工、竣工验收、档案管理等,形成了建设项目管理的内部控制体系。目前重庆海事局已经建立了以财

务集中核算、实时动态监管、数据分析预测等为核心功能的财务云平台,同时根据《管理办法》等的要求开发了建设项目管理模块,从事前、事中和事后全方位控制建设项目的风险^[1]。

依据建设项目管理相关规范进行内部控制评价,以评价主题为导向,结合数据仓库技术来建立行政事业单位建设项目内部控制评价机制。首先要确认建设项目内部控制评价的主题,然后从多个数据源提取相关数据,对数据进行清洗、转换和加载(ETL),再根据数据仓库模型建立建设项目内部控制数据仓库,最后利用HiveQL语句查询出评价主题相关的数据,发现内部控制中的异常点与风险点,完成建设项目内部控制评价。

(二)构建过程

行政事业单位建设项目内部控制评价数据仓库体系结构主要由四个部分组成:建设项目管理相关数据源的分析与处理、内部控制评价数据仓库构建、内部控制数据分析与评价以及内部控制评价结果输出。以重庆海事局为例,其建设项目内部控制评

【基金项目】国家社会科学基金项目(项目编号:17BGL194); 重庆海事局基于财务云平台的内部控制信息化系统建设项目(项目编号:2018Q60)

价数据仓库体系结构的设计过程参考《财会月刊》2019年第13期文章《基于数据仓库的行政事业单位层面内部控制评价——以重庆海事局为例》相关内容^[2]。

重庆海事局建设项目内部控制评价数据仓库的数据源主要包括《内控规范》、《管理办法》、海事局财务云平台的业务数据,以及金蝶K3系统中建设项目信息数据、建设项目资金数据、建设项目档案数据、竣工结算数据和建设项目其他相关数据。该数据仓库的构建过程包含以下五个步骤:第一,根据建设项目内部控制评价需求,采集所需数据;第二,将从不同数据源采集来的数据运用ETL技术进行清洗、转换,然后加载到数据仓库中;第三,围绕不同的主题,搭建各个主题的数据集市,构成以Hadoop为基础架构的建设项目内部控制评价大数据平台,形成包含资金管理、档案管理和项目管理三个方面共九个主题数据集市的建设管理内部控制评价数据仓库(Hive);第四,针对特定的目标,利用HiveQL语言、OLAP工具、数据挖掘算法(Mahout)等技术对内部控制评价模型进行多维度智能分析;第五,利用可视化技术图文并茂地展示评价结果,并形成建设项目内部控制评价报告。

二、建设项目内部控制评价数据采集与ETL

(一)数据来源

重庆海事局建设项目内部控制评价数据仓库构建的数据包括结构化数据(如预算数据、财务数据)、半结构化数据(如HTML网页文件)和非结构化数据(如招标报告、施工报告)^[3]。在进行数据源分析时,既要考虑系统的结构化数据,又要考虑业务层面的非结构化数据;既要考虑财务云平台的数据,又要根据需求采集其他数据源中的相关数据。建设项目内部控制评价数据仓库的具体数据来源如下:

1. 重庆海事局财务云平台。该平台中有系统数据库和业务数据库两类数据库,系统数据库主要记录重庆海事局人员、资产、部门等方面的基础信息,业务数据库主要包括资产、合同、预算、收支、建设项目等业务数据。建设项目的会议签到表、人员表、项目预算表、项目表等结构化数据以及规范文件汇编等非结构化数据,可以从财务云平台中获取。

2. 金蝶K3账务处理软件。主要记录重庆海事局相关的财务数据,建设项目财务管理模块中项目资金明细表等结构化数据可从中提取,例如项目转

账明细表、项目预算明细表、项目拨款明细表。

3. 外部数据。包括会议纪要、项目建议书、可行性研究报告、竣工决算报告等非结构化数据以及施工进度表、档案信息表等结构化数据。

(二)数据采集与ETL

数据的采集要保证数据的完整性、及时性,为建设项目内部控制评价分析提供准确、可靠的数据支持。获取数据的方式有多种途径,在财务云平台以及金蝶K3账务处理软件上的数据,可以将与主题相关的数据直接提取到目标数据仓库;外部数据可直接导入数据仓库,也可先运用Python网络爬虫从网页中获取再导入数据仓库。数据来源途径多样,导致存在大量的数据冗余、格式不统一、语义矛盾等现象,可以运用ETL工具来处理数据。例如,从外部导入的施工进度表中的“施工进度”的数据类型为数值型,而工程发承包合同中要求“进度”的数据类型为字符型,应统一数据类型,将工程发承包合同的“进度”改为数值型。

三、建设项目内部控制评价数据仓库设计

(一)确定主题域

根据《内控规范》和《管理办法》等文件,基于重庆海事局建设项目内部控制的现状,确定建设项目内部控制评价主题。建设项目内部控制评价应包括项目资金管理、议事决策、竣工决算等方面,实现对重庆海事局建设项目不同模块数据的分析与评价。

参考程平、张敏济^[4]的研究,本文将重庆海事局建设项目内部控制评价划分为九个主题,主题描述与依据见表1。

(二)粒度设计

粒度是指数据仓库数据单位中保存数据的细化或综合程度的级别,粒度会影响数据仓库的数据量 and 数据仓库的分析能力。细化程度越高,粒度越小,业务分析内容则越详细,但数据量过大会造成数据分析效率较低;细化程度越低,粒度越大,则数据的聚集程度越高,可分析的问题越少^[5]。因此,粒度设计应考虑数据仓库中数据的存储量大小及数据是否满足需求,恰当的粒度大小有助于提高业务分析能力,保持较高的数据分析效率。例如:专款专用合规性主题可将拨款用途作为粒度,分析各个项目每种用途的拨款总额;档案归档及时性主题则应从时间角度进行分析,该主题划分的最小粒度为日,分析项目的归档事件是否超过规定时间。

表 1

主题确定及其描述

主题名称	主题描述	依据
集体决策合规性	该主题检查单位集体决策过程中是否存在单独决策或擅自修改决策结果的行为	《内控规范》第四十七条
项目建议书审核合规性	该主题检查项目建议书是否按照规定编制并审核	《内控规范》第四十八条 《管理办法》第二十五条
可行性研究报告审核合规性	该主题考察可行性研究报告是否按照规定编制并审核	
概预算审核合规性	该主题考察概预算是否按照预算编制要求编制并审核	
竣工决算报告审核合规性	该主题考察竣工决算报告是否与竣工决算报告编制要求相符,检查是否满足相关规定	
专款专用合规性	该主题考察单位是否存在未按照审批单位下达的投资计划和预算对建设项目资金实行专款专用。	《内控规范》第五十条 《管理办法》第二十条、 第二十七条
款项支付进度合规性	该主题考察项目的支付进度是否符合规定。	
档案归档及时性	该主题考察单位是否规定时限内办理档案移交归档	《内控规范》第五十一条 《管理办法》第三十三条
竣工决算及时性	该主题考察建设项目竣工后,是否在规定时间内办理竣工决算	《内控规范》第五十三条 《管理办法》第二十一条、 第二十八条

(三)设计模型

根据建设项目内部控制评价内容确定的主题域,将建设项目内部控制评价数据仓库划分为九个数据集市,并基于每个主题的特性,构建数据仓库模型,模型及数据组织见表2。

1. 构建概念模型。概念模型可以将实体间的关系抽象化,明确各个建设项目主题活动之间的关系以及表之间的联系,运用E-R模型来设计。以重庆海事局为例,建设项目内部控制评价数据仓库概念模型如图1所示。

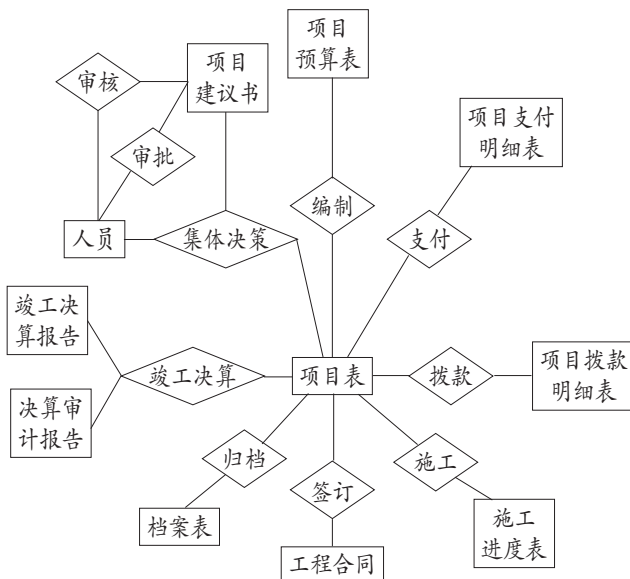


图 1 内部控制评价数据仓库概念模型

2. 构建逻辑模型。逻辑模型是概念模型的具体化,用来描述各个主题之间的关系。以重庆海事局建设项目专款专用合规性为例,其逻辑模型设计包括

三个事实表(项目拨款明细表、项目预算明细表、项目支付明细表)和四个维度表(项目基本信息表、拨款用途表、科目表、项目状态表),针对不同项目,利用唯一项目ID标识,通过项目ID将三个事实表连接,拨款用途与项目用途相匹配,确定项目具体的预算金额,再匹配预算科目和付款科目,确定实际支付金额。专款专用合规性的逻辑模型如图2所示。

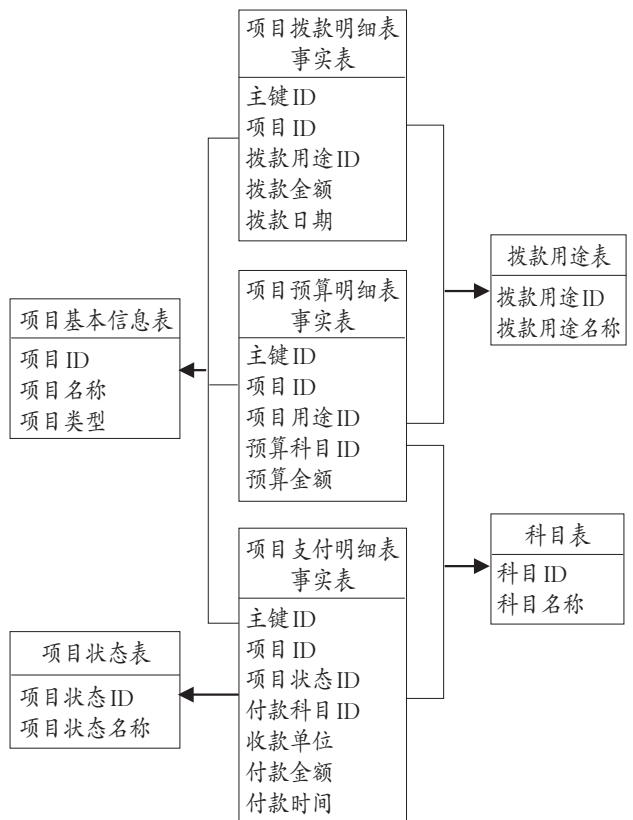


图 2 专款专用合规性逻辑模型

表 2

行政事业单位建设项目内部控制评价数据仓库模型及其数据组织

主题域名称	模型	事实表	维度表	模型描述
集体决策合规性	星座模型	会议签到表 NK_YW_JJ_SS_HYQDB	项目基本信息表 NK_YW_JJ_WD_XMJBXX	该模型通过检查会议签到表、会议纪要表中参会人员名称、人数、部门、职务等字段,与议事决策标准对比
		会议纪要表 NK_YW_JJ_SS_HYJYB	部门表 NK_YW_JJ_WD_BM	
		项目意见书 NK_YW_JJ_SS_XMYJS	人员表 NK_YW_JJ_WD_RY	
		议事决策标准 NK_YW_JJ_SS_YSJCBZ	项目类型表 NK_YW_JJ_WD_XMLX	
项目建议书审核合规性	星座模型	项目建议书 NK_YW_JJ_SS_XMJYS	人员表 NK_YW_JJ_WD_RY	该模型提取项目建议书中审核人员等字段与人员表中的姓名、职务、部门等字段对比,检查项目建议书审核是否符合要求
			项目基本信息表 NK_YW_JJ_WD_XMJBXX	
可行性研究报告审核合规性	星座模型	可行性研究报告 NK_YW_JJ_SS_KXXYJBG	人员表 NK_YW_JJ_WD_RY	该模型提取可行性研究报告中编制人员名称、审核人员名称等字段与人员表中姓名、职务、部门等字段对比,检查可行性研究报告是否符合要求
			项目基本信息表 NK_YW_JJ_WD_XMJBXX	
概预算审核合规性	星座模型	概预算文件 NK_YW_JJ_SS_GYS	人员表 NK_YW_JJ_WD_RY	该模型通过提取概算编制人员名称、职务等字段与法规文件汇编中的预算编制要求进行比对,检查项目概预算是否按照要求编制
		预算编制要求 NK_YW_JJ_SS_YSBZYQ		
竣工决算报告审核合规性	星座模型	竣工决算报告 NK_YW_JJ_SS_JGJSBG	项目基本信息表 NK_YW_JJ_WD_XMJBXX	该模型通过将竣工决算报告中编制人员名称、职务、审核人员名称、职务等与竣工决算报告编制要求进行核对,检查是否满足相关规定
		决算报告编制要求 NK_YW_JJ_SS_JSBGBZYQ	人员表 NK_YW_JJ_WD_RY	
专款专用合规性	星座模型	项目拨款明细表 NK_YW_JJ_SS_XMBK	项目基本信息表 NK_YW_JJ_WD_XMJBXX	该模型通过项目拨款明细表中拨款金额和项目转账明细表中付款金额、方式、备注等字段进行比对,检查单位对下达的投资和预算资金是否实行专款专用
		项目预算明细表 NK_YW_JJ_SS_XMYS	项目状态表 NK_YW_JJ_WD_XMZT	
		项目支付明细表 NK_YW_JJ_SS_XMZF	拨款用途表 NK_YW_JJ_WD_BKYT	
			科目表 NK_YW_JJ_WD_KM	
款项支付进度合规性	星座模型	项目支付明细表 NK_YW_JJ_SS_XMZF	项目基本信息表 NK_YW_JJ_WD_XMJBXX	该模型通过将项目支付金额汇总,除以预算金额,计算结果与施工进度表中对应的施工进度对比,判断是否按工程进度进行款项支付
		项目预算明细表 NK_YW_JJ_SS_XMYS		
		施工进度表 NK_YW_JJ_SS_SGJD		
		工程发承包合同 NK_YW_JJ_SS_GCFCBHT		
档案归档及时性	星座模型	档案归档记录表 NK_YW_JJ_SS_DAGD	项目基本信息表 NK_YW_JJ_WD_XMJBXX	该模型通过将项目信息表中项目状态“完工”时的录入日期与档案信息表中的归档日期进行比对,检查单位是否在规定时间内办理档案移交归档手续
竣工决算及时性	星座模型	竣工决算报告 NK_YW_JJ_SS_JGJS	项目基本信息表 NK_YW_JJ_WD_XMJBXX	该模型检查是否在规定的时间内办理竣工决算以及组织竣工决算审计,并且办理建设项目档案和资产移交

3. 构建物理模型。物理模型设计是指根据逻辑模型来确定实体间的关系和属性,通过具体的物理介质实现,是逻辑模型中各种实体表的具体化,包括数据的表结构、数据的存储方式、数据的获取方式的设计,可分为维度表和事实表。其中,维度表用于从不同角度分析主题,事实表用于存储每个主题的主要内容。以专款专用合规性主题为例,其事实表和维度表的结构、字段设计如下:

(1)NK_YW_JJ_SS_XMBK【项目拨款明细表】。表包含字段:I_ProjectID【项目ID】、M_Money【拨款金额】、DT_Account【拨款日期】、I_UseID【拨款用途ID】。

(2)NK_YW_JJ_SS_XMZF【项目支付明细表】。表包含字段:I_ProjectID【项目ID】、VC_receiveunit【收款单位】、I_State【项目状态】、M_paymoney【付款金额】、I_subjectID【付款科目ID】、DT_paytime【付款时间】、I_UseID【拨款用途ID】。

(3)NK_YW_JJ_SS_XMYS【项目预算明细表】。表包含字段:I_ProjectID【项目ID】、M_Budget【预算金额】、I_subjectID【预算科目ID】、I_UseID【预算用途ID】。

(4)NK_YW_JJ_WD_XMJBXX【项目基本信息表】。表包含字段:I_ProjectID【项目ID】、I_Style【项目类型】、VC_Name【项目名称】。

(5)NK_YW_JJ_WD_XMZT【项目状态表】。表包含字段:I_StateID【项目状态ID】、VC_StateName【项目状态名称】。

(6)NK_YW_JJ_WD_BKYT【拨款用途表】。表包含字段:I_UseID【项目用途ID】、VC_UseName【项目用途名称】。

(7)NK_YW_JJ_WD_KM【科目表】。表包含字段:I_subjectID【科目ID】、VC_SubjectName【科目名称】。

四、基于数据仓库的建设项目内部控制评价应用

(一)建设项目内部控制评价方法

基于数据仓库的决策分析方法主要有HiveQL语言查询分析、OLAP分析及数据挖掘,本文根据建设项目内部控制的分析主题来建立内部控制评价模型。

1. HiveQL语言查询分析。数据仓库构建完成后,用户可以根据建设项目的各个主题设计主题分析点,编写HiveQL语句对其进行访问并对其中的

数据进行多维度查询分析,如数据是否符合要求、是否存在风险或异常等。

2. OLAP分析。联机分析处理程序(OLAP)是一种归纳型的决策过程,基本操作包括切片、切块、旋转、上卷、下钻^[6]。对于建设项目内部控制评价,结合上文设计的事实表和维度表,在每次查询时只需确定OLAP需要展示的维度、事实字段和操作类型。例如:专款专用合规性可以对项目基本信息、拨款用途、项目状态、科目维度进行分析。若要查询每个项目的拨款及使用情况,则进行切块操作;若要查询某个项目的拨款用途,则进行切片操作。

3. 数据挖掘。数据挖掘是指从大量的数据中获取隐含在其中的有用信息的过程,将数据挖掘算法应用到建设项目内部控制评价中,可以识别潜在风险,实时跟踪资金流动情况。鉴于篇幅有限,本文主要阐述HiveQL语言查询分析在建设项目内部控制评价中的具体应用。

(二)基于HiveQL语言的内部控制情况查询分析

现以专款专用合规性、款项支付进度合规性和档案归档及时性为例,展现数据仓库技术在行政事业单位建设项目内部控制评价中的应用。

1. 专款专用合规性。数据仓库在专款专用合规性方面的应用是检查专项资金是否专项使用,首先根据项目拨款的用途检查预算编制的科目是否符合专项资金的用途,其次检查项目支付情况是否符合预算要求,根据项目状态是在建还是完工和付款时间、时间上是否合规以及付款科目及金额,检查是否按照预算科目进行付款,是否有超预算或者挪用等情况。

分析点1:运用HiveQL语句从数据仓库中选取相关数据项,筛选出不符合专款专用合规性的建设项目。首先,通过“项目ID”连接拨款明细表和预算明细表,按照“拨款用途”分类,筛选出各用途的“预算金额”,将各个用途的“预算金额”汇总与“拨款金额”进行对比,若不相等,则说明未按拨款做预算,该建设项目的专款专用存在问题。对应的分析评价核心HiveQL语句如下:

```
Hive>
Select I_ProjectID
From NK_YW_JJ_SS_fund_detail 拨款明细表
Join NK_YW_JJ_SS_budget_detail 预算明细表
On 预算明细表. Project_ID= 预算明细表. Project_ID
Group by 拨款明细表. USE_ID
Having Sum( 预算明细表. M_Budget) != Sum( 拨款明
```

细表.Use_Money)

分析点2:通过“项目ID”连接预算明细表与项目支付明细表,按照“科目ID”分类,汇总每个科目的“支付金额”,与相应科目的“预算金额”进行对比,若“支付金额”大于“预算金额”,则说明出现了超预算使用资金的情况,该建设项目专款专用存在问题。对应的分析评价核心HiveQL语句如下:

```
Hive>
Select I_ProjectID
From NK_YW_JJ_SS_payment_detail 项目支付明细表
Join NK_YW_JJ_SS_budget_detail 预算明细表
On 项目支付明细表.Project_ID=预算明细表.Project_ID
Group by 预算明细表.subject_ID
Having Sum(项目支付明细表.M_paymoney)>=预算明细表.M_Budget
```

2. 款项支付进度合规性。数据仓库在款项支付进度合规性方面的应用是检查是否按工程进度进行款项支付,以及某一时点项目工程进度是否符合合同规定的计划进度。

分析点1:运用HiveQL语句从数据仓库中选取相关数据项,筛选出不符合款项支付进度合规性的建设项目。首先通过“项目ID”将支付明细表和预算表相连,汇总项目“支付金额”并除以项目“预算金额”,计算出支付的比例,然后与施工进度表的“施工进度”进行对比,若支付的比例超过施工项目进度,则出现异常。对应的分析评价核心HiveQL语句如下:

```
Hive>
Select I_ProjectID
From NK_YW_JJ_SS_payment_detail 项目支付明细表
Join NK_YW_JJ_SS_progress 施工进度表
On 项目支付明细表. Project_ID=施工进度表. Project_ID
Join NK_YW_JJ_SS_budget_detail 预算明细表
On 项目支付明细表. Project_ID=预算明细表. Project_ID
Where Sum(项目支付明细表. M_paymoney)/预算明细表. M_Budget!>施工进度表.Project_progress
```

分析点2:连接该项目的合同,在相同的时间节点,将“施工进度”与“合同计划进度”进行对比,若实际施工进度小于计划进度,则出现异常,将此项目查找出来,进一步分析原因。对应的分析评价核心HiveQL语句如下:

```
Hive>
Select I_ProjectID
From NK_YW_JJ_SS_progress 施工进度表
Join NK_YW_JJ_SS_contact_project 工程发承包合同
On 施工进度表.Project_ID=工程发承包合同.Project_ID
Where 施工进度表. Project_progress !=工程发承包合同. Schedule_progress
```

3. 档案归档及时性。根据《管理办法》的要求,项目通过竣工验收后三个月内,建设单位应办理档案移交归档手续。通过HiveQL语句,查询出建设项目的完工时间和归档时间,计算时间差,若时间差超过3个月,则出现异常,需查询出来做进一步分析。对应的分析评价核心HiveQL语句如下:

```
Hive>
Select *
From NK_YW_JJ_SS_file 档案表
Join NK_YW_JJ_WD_project 项目表
On 档案表. Project_ID=项目表.ID
Where 档案表. Date_finish_file_date>90 and 项目表.I_State=="完工"
```

主要参考文献:

- [1] 程平,张敏济. 数据视角下基于财务云平台的建设项目内部控制审计——以重庆海事局为例[J]. 财会月刊,2017(32):93~101.
- [2] 程平,范洵. 基于数据仓库的行政事业单位单位层面内部控制评价——以重庆海事局为例[J]. 财会月刊,2019(13):71~76.
- [3] 程平,张敏济,刘艳. 基于财务云平台的重庆海事局建设项目控制探析[J]. 交通财会,2017(6):39~42.
- [4] 程平,张敏济. 数据视角下基于财务云平台的建设项目内部控制审计——以重庆海事局为例[J]. 财会月刊,2017(32):93~101.
- [5] 吕海燕,车晓伟. 数据仓库中数据粒度的划分[J]. 计算机工程与设计,2009(9):2323~2325.
- [6] 宋旭东,胡墨谦,刘晓冰. 基于模型驱动体系架构的OLAP开发方法[J]. 计算机集成制造系统,2010(2):423~430.

作者单位:1.重庆理工大学会计学院,重庆400054;
2.重庆海事局内部控制信息化课题组,重庆400054