

基于数据仓库的行政事业单位采购管理内部控制评价

——以重庆海事局为例

程平^{1,2}(博士生导师), 杜姗^{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)》等相关文件,设计基于Hadoop平台的采购管理内部控制评价数据仓库体系结构,分析数据的采集、数据清洗的过程与方法,通过确定主题域、分析粒度以及概念模型、逻辑模型、物理模型的设计详细阐述了采购管理内部控制评价数据仓库的构建。最后利用HQL语言在Hive中完成对海量数据的查询分析,展示了采购管理中异常点、风险点的发现方法,以期提升行政事业单位采购管理内部控制评价的效率和质量。

【关键词】行政事业单位; 采购管理; 数据仓库; 内部控制评价

【中图分类号】F233;C931 **【文献标识码】**A **【文章编号】**1004-0994(2019)17-0053-5

近年来,随着云会计^[1]、物联网、大数据等技术的快速发展,以“数据驱动”进行各类决策的趋势逐渐得到认可与推广,其中,围绕特定业务主题的数据仓库与联机分析处理(OLAP)由于直接对接具体应用业务与数据源,得到了广泛应用。鉴于此,本文以重庆海事局为例,设计面向采购管理内部控制评价的数据仓库体系结构,探讨行政事业单位采购管理内部控制评价数据仓库的构建与应用,以期为相关部门开展更为全面的内部控制评价信息化工作提供支持。

一、采购管理内部控制评价数据仓库体系的架构设计

(一)设计思路

重庆海事局是交通运输部设置在长江干线负责重庆段水上安全监督管理、防止船舶污染水域和水上人命救助的行政执法机关。采购管理作为行政事业单位的重要经济业务活动,经过多年的探索与实践,重庆海事局单位层面上的采购管理内部控制建

设取得了良好的效果。采购业务流程主要包括采购预算编审、采购计划申报与审批、采购合同审核、采购预算与计划的执行、资金结算以及供应商评价等环节。根据《行政事业单位内部控制规范(试行)》(以下简称《内控规范》)的要求,重庆海事局按照政府集中采购和协议采购流程,实现了“先预算、后计划、再采购”,对采购业务各环节的关键点进行重点控制。采用表单数据采集方式,管理和使用部门可以随时从数据库中提取供应商、采购内容、合同金额、合同时间等采购过程数据,实现对采购业务的全程透明控制。

对于整体采购业务流程而言,目前重庆海事局运用大数据、云会计等技术将采购业务及其控制流程成功嵌入到财务云平台中,基本完成了采购业务内部控制信息化的事前与事中审核,但是很难实现对采购供应商的商业信用评估、采购业务整体情况分析等事后评价体系的建立。对于分析型的数据处理而言,重庆海事局现有的各数据源之间没有很好的信息交互与共享功能,导致海量数据与“孤岛信

【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)

息”的情况并存,而且数据质量不高,仅仅停留在业务层面,因此为了便于做出管理决策,有必要构建面向分析层面的数据仓库。

(二)构建过程

基于财务云平台的采购管理内部控制数据仓库以《内控规范》为法规依据,结合重庆海事局采购业务流程,本文将数据仓库体系结构分为采购管理数据源、ETL(Extract, Transform, Load)过程、采购管理内部控制数据仓库、采购管理内部控制数据分析评价、采购管理内部控制评价结果输出五个部分。具体采购管理内部控制评价数据仓库体系结构设计过程图参考《财会月刊》2019年第13期中的文章《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》相关内容^[2]。

首先,采购管理内部控制评价数据是从海量的相关数据源中,经过数据采集和数据的抽取、清洗、转换,按照不同的主题加载到采购管理内部控制评价主题数据库中。采购管理主题数据库中按照不同的主题,如采购预算编制科学性、采购方式合理性、质疑投诉答复率等对数据进行分类存放形成数据集市。其次,将采购管理主题数据库与采购管理内部控制评价分析系统直接相连,实现采购管理内部控制评价数据的挖掘与分析。最后,将分析结果以可视化的方式呈现给决策用户端,实现用户需求的响应与反馈。

二、基于采购管理内部控制评价数据仓库的数据采集与清洗

数据采集即ETL过程,是数据仓库的核心,是将各个存储着财务信息的业务板块数据库中的数据进行抽取、清洗、转换、加载的过程。对与采购管理内部控制评价相关联的多种数据信息进行抓取,通过ETL过程将重复冗余的信息进行清洗,对缺失数据进行处理,进而加载到采购管理内部控制评价主题数据仓库中。

目前,重庆海事局的数据来源主要有财务云平台、金蝶K3系统、各类文件扫描件等多种采购业务相关数据,涵盖了海量的结构化、非结构化以及半结构化数据。其中,结构化数据是指由二维表结构进行逻辑表达和实现的数据,严格遵循数据格式与长度规范,通过关系型数据库进行存储和管理。采购管理内部控制涉及的结构化数据主要来源于财务云平台业务数据库中。非结构化数据是指数据结构不规则

或不完整,没有预定义的数据模型,不方便用数据库二维表结构来表现的数据。采购管理内部控制涉及的非结构化数据主要包含采购合同文本扫描件、采购物资的相关票据、《内控规范》等相关文本文件。半结构化数据是指非关系模型的、有基本固定结构模式的数据。采购管理内部控制涉及的半结构化数据主要是指中国政府采购网等以html文件格式存储的采购相关信息。

各种数据需要通过进一步清洗、转换才能将有助于数据分析的信息抽取到数据仓库中进行存储。数据清洗的任务是过滤掉不符合要求的数据,主要包括不完整数据(如供应商名称缺失、采购商品种类缺失等)、错误数据(如采购商品日期格式错误、采购金额等数值数据被输成全角数字字符等)、重复数据(如不同维度表中包含的相同字段等)。清洗过程就是将不完整数据补充完善后重新录入,将错误数据在业务系统中更正后录入并删除重复数据的过程。数据转换则主要是进行不一致的数据转换、数据粒度的转换。不一致数据转换是指将不同业务系统的相同类型的数据统一,如同一个供应商在财务云平台中的编码为XX001,在金蝶K3系统中的编码为YY001,需在数据抽取后将其转换为同一编码。数据粒度的转换主要是将不同业务系统中的数据按照采购管理内部控制评价数据仓库的粒度进行聚合。

三、基于采购管理内部控制评价数据仓库的设计

(一)确定主题域

数据仓库的主题是在较高层次上将业务模型和面向事务的数据进行分析、归类和综合的一个过程,每个主题对应一个分析领域。结合《内控规范》及相关规定的要求,采购管理内部控制评价应满足采购预算编制、采购招投标管理、采购质疑投诉等方面的需求,实现对重庆海事局采购管理不同模块数据的分析及评价。因此,本文将重庆海事局采购管理模块数据仓库的主题域分为:采购预算编制科学性、采购计划执行有效性、采购方式合理性、招投标控制合规性、质疑投诉答复率、采购档案完整性、采购信息分析统计合规性等七个主题,具体见表1。

在表1中,以采购预算编制科学性主题为例,可通过考察预算编制表“项目名称”“申报金额”“核定金额”和采购管理办法中“产品配置数量”“产品使用年限”是否符合实际需求和相关标准编制政府采购预算,从而评价采购预算的编制是否科学。

表 1

采购管理内部控制评价数据仓库主题域

主题域名称	主题描述	设计依据
采购预算编制科学性	评价该单位采购资金安排和使用的合理性	《内控规范》第三十四条规定:单位应当加强对政府采购业务的预算管理,根据本单位实际需求和相关标准编制政府采购预算
采购计划执行有效性	检查该单位是否根据已批复的采购预算安排采购活动	《内控规范》第三十四条规定:根据本单位实际需求和相关标准编制政府采购预算,按照已批复的预算安排政府采购计划
采购方式合理性	检查该单位是否按照规定采用了合适的采购方式	《内控规范》第三十五条规定:单位应当加强对政府采购申请的内部审核,按照规定选择政府采购方式
招投标控制合规性	检查该单位招标、投标过程中是否严格执行了招投标规定,确保采购招标项目合规	《内控规范》第十一条规定:单位进行经济活动业务层面的风险评估时,应当重点关注是否建立了有效的招投标控制机制
质疑投诉答复率	检查该单位处理质疑投诉的答复情况	《内控规范》第三十七条规定:单位应当加强对政府采购业务质疑投诉答复的管理
采购档案完整性	反映该单位保管采购预算与计划、各类批复文件、合同文本、验收证明等采购业务相关资料的情况	《内控规范》第三十八条规定:单位应当加强对政府采购业务的记录控制。妥善保管政府采购预算与计划、各类批复文件、合同文本、验收证明等政府采购业务相关资料
采购信息统计分析合规性	反映该单位定期对采购业务信息进行分类统计的情况	《内控规范》第三十八条规定:单位应当定期对政府采购业务信息进行统计,并在内部进行通报

(二)确定分析粒度

数据仓库的粒度划分层次是数据仓库构建设计中要解决的一个重要问题,粒度层次划分适当与否直接影响数据仓库中的数据量和所适合的查询类型。一般数据的粒度分为单一粒度和双重粒度:单一粒度是指将数据的最细粒度数据增量存储,并在最细粒度模型上聚合数据;双重粒度是指只将最近几个月的最新粒度数据保留在数据仓库中,当到达某个保留周期时,将保留周期之后的数据导出到磁盘上存储,从而为最新的数据腾出空间。单一粒度和双重粒度的区别在于:单一粒度将最细粒度的数据永久保存于数据仓库中,而双重粒度保留周期内最新的最细粒度数据。重庆海事局内有多套业务系统,历史数据量大,若采用单一粒度,那么对于存储的要求极高,因此,本文数据仓库将采用双重粒度。例如采购预算编制科学性主题可从时间维度进行分析,按照采购日期、预算编制日期以日为最小粒度进行粒度分割;采购方式合理性可从采购商品的地区维度进行分析,按照采购商品地点以区为最小粒度进行粒度分割。

(三)模型设计

数据仓库的模型设计是数据仓库建设的核心,也是决定数据仓库项目成功完成与否的关键因素。基于上文对采购管理内部控制评价数据仓库主题域的划分,将其划分为七个数据集市,并基于各主题特性,构建数据仓库模型,不同主题域的设计见表 2。本文以采购预算编制科学性主题为例,对其进行

具体的数据仓库模型设计。

1. 概念模型。数据仓库的概念模型设计可采用 E-R 模型以及面向对象分析两种方法,本文采用 E-R 模型进行采购管理内部控制评价数据仓库的概念模型设计,如图 1 所示。

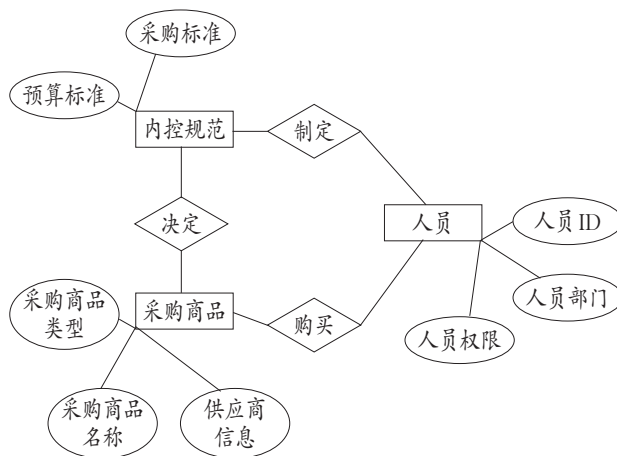


图 1 采购管理内部控制评价数据仓库概念模型

2. 逻辑模型。数据仓库的逻辑模型则是概念模型的具体化。采购管理内部控制评价数据仓库的每个主题都是由多个表来实现的,这些表之间依靠主题的主键联系在一起,形成一个完整的主题。维度表是用户分析数据的窗口,决定了数据分析的广度。事实表则用来存放业务的数字度量值信息,这些度量值信息是从各维度的交点处得来的。在选取数据仓库维度时应该将实体作为一个对象,把与该对象相关的所有重要属性都提取出来作为独立维度。以采

表2 行政事业单位采购管理内部控制评价

主题域名称	数据仓库模型	事实表	维度表	模型描述
采购预算编制科学性	雪花模型	采购预算编制科学性事实表 NK_YW_CG_SS_YSBZKX	采购商品维度表 NK_YW_CG_WD_CGSP	重点关注采购是否符合实际需求和相关标准编制政府采购预算
			采购方式维度表 NK_YW_CG_WD_CGFS	
			资产维度表 NK_YW_CG_WD_Asset	
			人员维度表 NK_YW_CG_WD_People	
			预算项目维度表 NK_YW_CG_WD_YSXM	
			时间维度表 NK_YW_CG_WD_TIME	
采购计划执行有效性	雪花模型	采购计划执行有效性事实表 NK_YW_CG_SS_CGJHZXYX	采购商品维度表 NK_YW_CG_WD_CGSP	重点检查是否根据已批复的采购预算安排采购活动
			采购方式维度表 NK_YW_CG_WD_CGFS	
			预算项目维度表 NK_YW_CG_WD_YSXM	
采购方式合理性	雪花模型	采购方式合理性事实表 NK_YW_CG_SS_CGFSHLX	采购商品维度表 NK_YW_CG_WD_CGSP	重点对比不同金额的采购商品采用的采购方式是否合规
			采购方式维度表 NK_YW_CG_WD_CGFS	
			采购标准维度表 NK_YW_CG_WD_Standard	
			人员维度表 NK_YW_CG_WD_People	
招投标控制合规性	雪花模型	招投标控制合规性事实表 NK_YW_CG_SS_ZTBKZHG	采购商品维度表 NK_YW_CG_WD_CGSP	重点检查该单位是否和信用等级良好的单位合作
			采购方式维度表 NK_YW_CG_WD_CGFS	
			供应商维度表 NK_YW_CG_WD_SPY	
			招投标信息维度表 NK_YW_CG_WD_ZTBXX	
质疑投诉答复率	雪花模型	质疑投诉答复率事实表 NK_YW_CG_SS_ZYTSDF	采购商品维度表 NK_YW_CG_WD_CGSP	重点关注“答复满意度”是否低于考核指标值
			质疑投诉维度表 NK_YW_CG_WD_ZYTS	
			质疑投诉回复维度表 NK_YW_CG_WD_ZYTSHF	
			时间维度表 NK_YW_CG_WD_TIME	
采购档案完整性	雪花模型	采购档案完整性事实表 NK_YW_CG_SS_CGDAWZX	采购商品维度表 NK_YW_CG_WD_CGSP	重点检查该单位各项采购业务相关资料是否齐全
			采购档案维度表 NK_YW_CG_WD_CGDA	
采购信息分析统计合规性	雪花模型	采购信息分析统计合规性事实表 NK_YW_CG_SS_XXFXTJ	采购商品维度表 NK_YW_CG_WD_CGSP	重点关注该单位是否定期对政府采购业务信息进行统计,并在内部进行通报
			采购方式维度表 NK_YW_CG_WD_CGFS	
			人员维度表 NK_YW_CG_WD_People	
			时间维度表 NK_YW_CG_WD_TIME	

购预算编制科学性主题为例,其维度可分为采购商品维、资产维、采购方式维、预算编制维、预算标准维、人员维等,通过以上维度与事实表相连,具体分析采购预算编制是否符合相关规定以及采购预算是否符合实际需求。具体的逻辑模型设计如图2所示。

3. 物理模型。数据仓库的物理模型设计就是数据仓库逻辑模型在物理系统的实现,即基于逻辑模型,在数据仓库中建表、索引等。本文以采购预算编制科学性主题为例的物理模型设计如下:

(1) NK_YW_CG_WD_CGSP【采购商品维度表】包含字段:CG-SP_ID【采购商品ID】、CG-SP_NAME【采购商品名称】、CG-SP_NOTE【采购商品信息】。

(2) NK_YW_CG_WD_CGFS【采购方式维度表】包含字段:CGFS_ID【采购方式ID】、CGFS_NAME【采购方式名称】、CGFS_NOTE【采购说明】、CGFS_Depot【采购部门】。

(3) NK_YW_CG_WD_ZC【资产维度表】包含字段:ZC_ID【资产ID】、ZC_NAME【资产名称】、ZC_STYLE【资产类型】、ZC_De-
pot【资产使用部门】。

(4) NK_YW_CG_WD_People【人员维度表】包含字段:P_ID【人员ID】、P_NAME【人员名称】、P_AUTHORITY【人员权限】、P_Depot【人员部门】。

(5) NK_YW_CG_WD_YSXM【预算项目维度表】包含字段:B_ID【预算项目ID】、B_NAME【预算项目名称】、B_STYLE【预算项目类型】。

(6) NK_YW_CG_WD_TIME【时间维度表】包含字段:Year【年】、Month【月】、Day【日】。

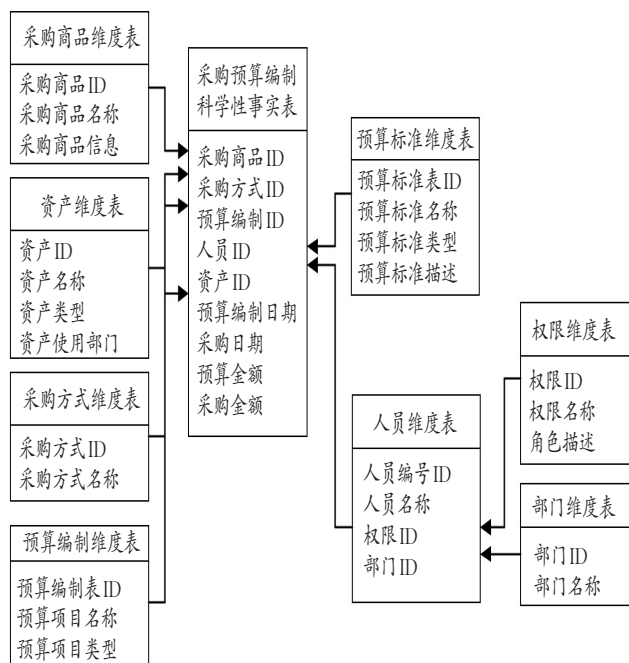


图 2 采购预算编制科学性主题数据仓库雪花模型

四、基于数据仓库的采购管理内部控制评价应用

(一) 采购管理内部控制数据分析与评价

基于采购管理内部控制评价数据仓库的决策分析方法主要有联机分析处理程序(OLAP)、运用HQL查询语句以及数据挖掘三种方法。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,主要优势在于能够直接仿照用户的多角度思考模式,预先为用户组建多维数据模型,通过快捷、一致、交互的访问使用户更好地理解数据。Hive是一个以Hadoop为基础的、建立在Hadoop生态系统之上的数据仓库,能够对大量的结构化数据进行存储与处理。它将数据存储在HDFS中,采用类SQL语言HQL语句对数据进行操作与管理。Hive中的解释器负责对HQL进行解析和转换,将其解释为Map/Reduce任务,并通过执行Map/Reduce任务返回对Hive中数据的查询或处理的结果。数据挖掘一般是指从大量的数据中通过算法搜索隐藏于其中信息的过程。本文重点介绍利用HQL查询语句进行分析查询的方法。

(二) 采购管理内部控制评价建模

根据上文确立的采购管理内部控制评价数据仓库主题域,可对七个分析主题建立评价模型,本文以基于重庆海事局财务云平台的采购预算编制科学性主题模型为例,利用HQL语言在Hive中完成对海

量数据的查询操作,实现对采购管理中异常点、风险点的捕捉。采购预算编制科学性主题旨在考察单位采购资金安排和使用的合理性,是否符合实际需求和相关标准编制政府采购预算。首先检查采购预算的编制是否符合《内控规范》等相关条例,如采购方式、采购金额等;其次检查采购商品的实际情况与采购预算编制是否相符,是否存在预算未执行或超预算购买的情况。

1. 步骤一。运用HQL语句从数据仓库中选取相关数据项,筛选出采购项目中不符合采购预算编制科学性的项目。通过“预算项目ID”连接预算编制表和预算标准表,将预算项目的“预算金额”与对应预算项目的“预算标准金额”相比对,若“预算金额”大于“预算标准金额”,则说明该项预算项目的编制存在问题。对应的分析评价核心HQL语句如下:

```

Select I_ProjectID
From ZC_budget_formation 预算编制表
Join ZC_budget_standard 预算标准表
On 预算编制表.Project_ID=预算标准表.Project_ID
Group by 预算编制表.subject_ID
Having Sum(预算编制表.M_budget)>=预算标准表.M_standard

```

2. 步骤二。通过“采购商品ID”连接采购明细表和预算明细表,将采购商品的“预算金额”与“采购金额”对比,若不相等,则说明实际采购与采购预算不符,检查出存在异常的项目。对应的分析评价核心HQL语句如下:

```

Select I_Project_ID
From ZC_purchase_detail 采购明细表
Join SYS_budget_detail 预算明细表
On 采购明细表. Project_ID=预算明细表. Project_ID
Group by 采购明细表.USE_ID
Having Sum(预算明细表. M_Budget)!=Sum(采购明细表.USE_Money)

```

主要参考文献:

[1] 程平,何雪峰.“云会计”在中小企业会计信息化中的应用[J].重庆理工大学学报(社会科学),2011(1):55~60.
[2] 程平,范洵.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[J].财会月刊,2019(13):71~76.
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