

基于数据仓库的行政事业单位预算管理内部控制评价

——以重庆海事局为例

程平^{1,2}(博士生导师), 文少波^{1,2}

【摘要】随着“互联网+”的快速发展,行政事业单位利用信息技术改善其预算管理内部控制评价的需求愈加强烈。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)》等相关文件和财务云平台预算管理业务应用,设计了基于Hadoop平台的预算管理内部控制评价数据仓库体系结构,对数据仓库的数据采集、清洗、转换过程进行了详细阐述,并从主题确定、声明粒度、确认维度、确认事实以及模型设计五个步骤具体分析了预算管理内部控制评价数据仓库的构建过程,最后利用基于Hive的HQL语言进行数据分析,展示了行政事业单位预算管理内部控制评价的具体应用方法。

【关键词】数据仓库;行政事业单位;预算管理;内部控制;Hadoop平台

【中图分类号】F232;TP315 **【文献标识码】**A **【文章编号】**1004-0994(2019)18-0059-5

目前,利用云会计和数据仓库等信息技术对行政事业单位预算管理内部控制进行评价的研究十分有限。鉴于此,本文以重庆海事局为例,将《内控规范》和《重庆海事局预算管理办法》作为指导,结合现有的财务云平台,在Hadoop平台上设计并构建预算管理内部控制评价的数据仓库,以实现预算管理内部控制相关数据的整合和利用。

一、预算管理内部控制评价数据仓库基本架构设计

1. 设计思路。预算管理内部控制评价数据仓库的构建,首先需要根据内部控制评价指标初步确定数据仓库评价主题,再对数据进行采集、清洗、转换、加载,以保证获取高质量的数据,为数据集市的设计、多维分析及数据挖掘提供数据基础。

2. 构建过程。重庆海事局通过建立数据仓库,对预算管理内部控制进行评价,其预算管理内部控制评价数据仓库体系结构设计图详见《财会月刊》2019年第13期文章《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^[1]。预算管理内部控制评价数据仓库体系结构主要由内控相关数据源、内控评价数据仓库构建、内控数据分析与评价、内控评价结果输出四部分组成。具体设计过程是以《内控规范》和预算管理业务流程作为指导基础,将预算管理内部控制相关数据从金蝶K3账务处理软件、财务云平台、互联网等利用Sqoop工具抽取到Hive的RDS库中(RDS:原始数据存储,充当操作性系统与数据仓库之间的过渡区),再对数据进行清洗转换,接着加载到预算管理内部控制评价数据仓库中。加载到数据仓库后,根据对应的评价主题形成预算管理内部控制评价数据集,再利用Hive的HQL语言对预算管理内部控制进行即席查询,同时通过OLAP对预算管理内部控制进行不同角度的切片切块分析和不同层级的上卷下钻处理,并结合每个主题的特定风险设置分析点,对预算管理内部控制评价进行建模,最后将评价结果通过可视化或报告的形式输出展示。

【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)

二、数据仓库的数据来源分析及 ETL 过程

1. 数据来源。重庆海事局利用大数据、云计算等信息技术,有效地整合了各方面内部数据、外部数据。具体数据来源如下:①重庆海事局财务云平台。该平台中有大量数据库,基本可分为系统数据库和业务数据库两类。②金蝶 K3 账务处理软件。该账务处理软件记录了重庆海事局的财务数据,预算管理内部控制评价数据仓库设计主要是基于此系统的财务管理模块记录的项目收入预算建议数、支出预算建议数、实际收入、实际支出等数据。③外部数据。预算管理内部控制评价数据仓库设计还需从外部获取数据,如预算编制依据、预算审核依据等自描述的半结构化数据多来源于纸质文件或互联网,需要额外的技术录入(如影像扫描技术或网络爬虫技术)。

综上,重庆海事局预算管理内部控制评价数据仓库主要是将财务云平台及金蝶 K3 账务处理软件中的部门人员信息、预算管理数据等内部数据和互联网信息等外部数据进行全方位采集,按照数据源的不同进行分类处理。业务系统中的数据形式多种多样,可能是 MySQL、SQL Server 等关系数据库中的结构化数据,也可能是 HTML、XML 等自描述的半结构化数据,还可能是文本、CSV 等平面文件或图像、影像文件等非结构化数据。对这些预算业务相关数据进行 ETL 处理,建立整合的预算管理内部控制评价的数据仓库。

2. ETL 过程。ETL 即数据抽取、转换、加载的过程:①数据抽取是从外部异构数据源抽取数据,抽取技术分为完全抽取和变化数据捕获。例如,重庆海事局部门人员信息这类数据,由于数据量不大且易处理,就采取完全数据抽取方式;而各种预算操作日志记录因其数据量大,就只能抽取变化的源数据,即最后一次抽取之后发生了变化的数据。②数据转换的重点是数据清洗,数据清洗是对数据进行重新审查和校验的过程。对于重复记录,可通过在查询语句中使用 distinct 关键字去重。而对于数据的结构要求,就得根据具体业务分析需求做出调整。③数据加载是指将清洗和转换后的数据加载到 DW 中,可通过直接追加、全面覆盖等策略进行加载。

三、预算管理内部控制评价数据仓库设计

1. 确定主题域。数据仓库是面向主题的,所以主题域的确定是数据仓库设计的关键。参考程平、李雪燃^[2]提出的重庆海事局预算管理内部控制审计指

标体系,以《内控规范》为指导,结合《重庆海事局预算管理暂行办法》,整理出预算组织机构建设完整性、预算岗位人员管理合理性、预算编制流程合规性、预算编制方法科学性、预算编制内容完整性、预算编制及时性、预算批复方法科学性、预算分解方法科学性、预算追加调整流程合规性、预算执行有效性、预算预警机制有效性、决算管理合规性和绩效目标差异分析十三个主题,详见表 1。

2. 声明粒度。粒度可简单理解为数据的粗细,用于确定事实中表示的是什么,同时数据粒度也是数据分析时进行上卷、下钻操作的基础。在预算管理内部控制评价数据仓库中,以预算编制流程合规性评价主题为例,需具体到每个部门、每个操作员、每天关于预算的操作行为,因此数据仓库粒度层是日→月→季度→年,如××部门××操作员关于预算编制的××操作行为。

3. 确认维度。维度表是事实表的基础,通常是一组层次关系或描述信息,用来定义事实,说明事实表的数据是从哪里采集的,即维度是观察数据的角度。在预算管理内部控制评价数据仓库中,以预算编制流程合规性评价主题为例,维度可分为编制依据维、审核依据维、人员基本信息维、部门维、角色维、时间维。①编制依据维/审核依据维,主要是从依据的角度分析预算编制、审批时是否有理有据,而非凭空想象。依据维主要包括有关法律法规、财政财务规章制度,以及财政部门、上级单位对年度预算编制工作方针、政策和工作布置。②人员基本信息维/角色维,描述预算相关人员的信息,用它来分析预算编制各个流程操作的操作员是否符合规定,即编制人员、审批人员是否有权限、是否实现三权分立等。③部门维,因为预算编制是按部门进行的,所以将部门作为一个维度,按部门角度来分析预算编制流程是否合规。④时间维,数据仓库是随时间变化的数据集合,需要记录数据的历史,因此每个数据仓库都需要一个时间维度表。

4. 确认事实。事实表用于存放预算业务的数字度量值信息,这些度量值信息是将各个维度作为条件,联合查询得到的。在预算管理内部控制评价数据仓库中,以预算编制流程合规性评价为例,其事实就是从时间、部门、操作员、编制依据等维度的交点处得到的预算金额。

5. 模型设计。根据预算管理内部控制评价确定的主题域,将预算管理内部控制评价数据仓库划分

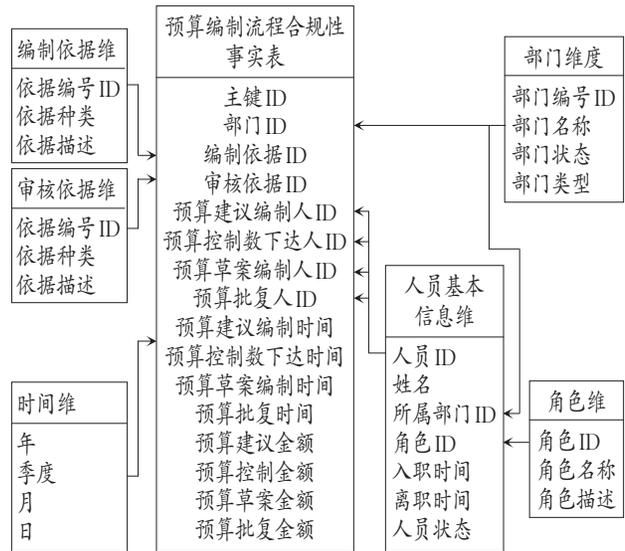
表 1 重庆海事局预算管理内部控制评价主题域

序号	控制流程	主题名称	主题描述	依据
1	组织管理体系控制	预算组织机构建设完整性	评价单位是否设置了预算管理委员会,是否有预算牵头部门,是否有确定的预算责任主体,是否建立预算编制、审批、执行、决算与评价等预算内部管理制度	《内控规范》第十九条
		预算岗位人员管理合理性	评价预算编制、审批、执行、考核等不相容岗位是否做到职责分离,是否建立了预算工作岗位责任制等	《内控规范》第十九条
2	预算编报控制	预算编制流程合规性	评价编制流程是否依循了“二上二下”的要求	《内控规范》第二十条,《重庆海事局预算管理办法》第三十条、第三十一条、第三十二条、第三十三条
		预算编制方法科学性	评价预算编制方法的选择是否匹配了环境的变化、是否科学	
		预算编制内容完整性	评价预算编制内容是否做到精细化,是否有编制依据、审核依据、部门、预算项目、预算科目、操作人、操作时间等数据	
		预算编制及时性	评价预算是否严格按照时间进度进行编制	
3	预算批复控制	预算批复方法科学性	评价预算审批方法的选择是否以以前年度的业务支出金额和本年度的业务工作计划为依据进行批复	《内控规范》第二十一条,《重庆海事局预算管理办法》第三十四条
4	预算下达控制	预算分解方法科学性	评价预算指标分解方法是否按照法定程序批复的预算在单位内部进行层层分解,是否落实到单位各个内设机构、各项经济活动和各个岗位	《内控规范》第二十一条,《重庆海事局预算管理办法》第三十四条
5	预算追加调整控制	预算追加调整流程合规性	评价预算追加调整是否符合法律法规,是否存在未经批准的预算调整	《内控规范》第二十一条,《重庆海事局预算管理办法》第六十一条
6	预算执行控制	预算执行有效性	评价执行项目/科目是否有相应的预算项目/科目,且执行金额与预算金额是否有较大出入	《内控规范》第二十二条,《重庆海事局预算管理办法》第六十三条
		预算预警机制有效性	评价单位是否建立预算执行跟踪及预警机制,当预算执行出现异常时是否能及时发现并采取措施	《内控规范》第二十二条,《重庆海事局预算管理办法》第六十六条
7	决算控制	决算管理合规性	评价决算的数据是否真实、完整地反映了海事局当年各项活动执行的情况,以及决算数据有没有及时披露	《内控规范》第二十三条
8	考评控制	绩效目标差异分析	评价是否对预算管理活动的各项实际绩效与目标绩效进行了差异分析	《内控规范》第二十四条

为十三个数据集市,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)概念模型设计。概念模型是确定如何组建数据及数据之间的相互关系,以满足业务应用的需要。在预算管理内部控制评价数据仓库中,其概念模型设计是预算内部控制评价所涉及的实体及实体之间的关系。以预算编制流程合规性评价主题为例,其主要是明确部门、编制流程、操作人、操作时间、编制依据、审核依据和预算金额之间的关系。

(2)逻辑模型设计。逻辑模型指数据的逻辑结构,推动概念模型向物理模型转换。逻辑模型由事实表和维度表组成,每个事实表引用任意数量的维度表,在预算管理内部控制评价数据仓库中,以预算编制流程合规性评价为例,其逻辑模型如图所示。



预算编制流程合规性评价数据仓库逻辑模型图

(3)物理模型设计。物理模型主要解决如何组织和存储数据的问题,以满足系统处理的要求,在预算管理内部控制评价数据仓库中,以预算编制流程合规性评价为例,其事实表、维度表及表包包含字段如下:①预算编制流程合规性事实表(NK_YW_YS_SS_Flow_organize)。表包含字段: I_ID【主键ID】、I_Depot_ID【部门ID】、I_BZYL_ID【编制依据ID】、I_SHYJ_ID【审核依据ID】、I_YSJY_ID【预算建议编制人ID】、I_YSKZ_ID【预算控制数下达人ID】、I_YSCA_ID【预算草案编制人ID】、I_YSPF_ID【预算批复人ID】、D_YSJY【预算建议编制时间】、D_YSKZ【预算控制数下达时间】、D_YSCA【预算草案编制时间】、D_YSPF【预算批复时间】、M_YSJY【预算建议金额】、M_YSKZ【预算控制金额】、M_

表 2 行政事业单位预算管理内部控制评价数据仓库模型及数据组织

序号	主题名称	模型	事实表	维度表	模型描述
1	预算组织机构建设完整性	星型模型	预算组织机构建设完整性事实表	用户信息表、员工信息表、角色表、权限表、部门表	该模型查询单位是否设置了预算管理委员会,是否有预算牵头部门,是否有确定的预算责任主体
2	预算岗位人员管理合理性	星型模型	预算岗位人员管理合理性事实表	用户信息表、员工信息表、角色表、权限表、部门表	该模型评价预算各个不相容岗位是否做到职责分离
3	预算编制流程合规性	雪花模型	预算编制流程合规性事实表	编制依据表、审核依据表、人员基本信息表、部门表、角色表	该模型评价编制流程是否依循了“二上二下”的要求
4	预算编制方法科学性	星型模型	预算编制方法科学性事实表	预算明细表、编制依据表、部门表、人员基本信息表	该模型评价预算编制方法的选择是否匹配了环境的变化、是否科学
5	预算编制内容完整性	雪花模型	预算编制内容完整性事实表	科目表、项目表、预算明细表、收入明细表、支出明细表、部门表、人员基本信息表	该模型评价预算编制内容是否做到精细化,是否有编制依据、审核依据、部门、预算项目、预算科目、操作人、操作时间等数据
6	预算编制及时性	星型模型	预算编制及时性事实表	科目表、项目表、预算明细表	该模型评价预算是否严格按照时间进度进行编制
7	预算批复方法科学性	星座模型	预算批复方法科学性事实表	预算明细表、审核依据表、部门表、人员基本信息表	该模型评价预算审批方法的选择是否以以前年度的业务支出金额和本年度的业务工作计划为依据进行批复
8	预算分解方法科学性	星座模型	预算分解方法科学性事实表	预算明细表、科目表、项目表、部门表、人员信息表	该模型评价预算指标分解方法是否按照法定程序批复的预算在单位内部进行层层分解,是否落实到单位各个内设机构、各项经济活动和各个岗位
9	预算追加调整流程合规性	星座模型	预算追加调整流程合规性事实表	预算调整表、预算编制表、部门表、人员信息表、编制依据表、审核依据表	该模型评价预算追加调整是否符合法律法规,是否存在未经批准的预算调整
10	预算执行有效性	星型模型	预算执行有效性事实表	预算明细表、预算执行表	该模型评价执行项目/科目是否有相应的预算项目/科目,且执行金额与预算金额是否有较大出入
11	预算预警机制有效性	星型模型	预算预警机制有效性事实表	预算明细表、预算执行表、施工进度表	该模型评价单位是否建立预算执行跟踪及预警机制,当预算执行出现异常时是否能及时发现并采取措施
12	决算管理合规性	星型模型	决算管理合规性事实表	预算明细表、决算明细表	该模型评价决算的数据是否真实、完整地反映了海事局当年各项活动执行的情况,以及决算数据有没有及时披露
13	绩效目标差异分析	星型模型	绩效目标差异分析事实表	绩效表、预算执行表	该模型评价是否对预算管理活动的各项实际绩效与目标绩效进行了差异分析

YSCA【预算草案金额】、M_YSPF【预算批复金额】。
 ②编制依据维度表(NK_YW_YS_WD_Gist_organize)。表包含字段:ID【依据编号ID】、I_Style【依据种类】、VC_Gistdesc【依据描述】。
 ③审核依据维度表(NK_YW_YS_WD_Gist_check)。表包含字段ID【依据编号ID】、I_Style【依据种类】、VC_Gistdesc【依据描述】。
 ④人员基本信息维度表(NK_YW_YS_WD_SYS_PeopleBaseSet)。表包含字段:ID【人员ID】、VC_Name【姓名】、I_Depot_ID【所属部门ID】、I_Role_ID【角色ID】、D_Rztime【入职时间】、D_Lztime【离职时间】、I_State【人员状态】。
 ⑤部门维度表(NK_YW_YS_WD_SYS_Unit)。表包含字段:I_Depot_ID【部门编号ID】、VC_Name【部门名称】、

I_State【部门状态】、I_Style【部门类型】。
 ⑥角色维度表(NK_YW_YS_WD_SYS_Roleinfo)。表包含字段:ID【角色ID】、RoleName【角色名称】、VC_Roledesc【角色描述】。
 ⑦时间维度表:(NK_YW_YS_WD_Time)。表包含字段Year【年】、Quarter【季度】、Month【月】、Day【日】。

四、基于数据分析的预算管理内部控制评价应用

1. 预算管理内控即席查询分析。预算管理内部控制评价数据仓库的建立,实现了即席查询功能,即根据预算管理内部控制评价需求,灵活地选择查询条件,进而生成相应的统计报表。基于Hive的HQL语句按预算部门、预算流程、预算相关人员、预算金额、预算科目、预算项目、预算编制依据和预算审核

依据等条件进行查询分析,以预算编制流程合规性评价为例,选择预算编制部门、预算编制流程中所处的具体编制流程环节(如“一上”环节),查询到编制时间、编制人员、编制依据和预算编制“一上”环节的编制金额,根据预算建议数上报金额,分析编制人员权限是否符合要求、编制是否有理有据、编制金额是否存在异常。

2. 预算管理内控 OLAP 分析与数据挖掘。预算管理内部控制联机分析处理(OLAP)集中于数据的分析,从预算管理内部控制评价数据仓库的集成数据出发,构建面向内控评价分析的多维数据模型,再使用多维分析方法从不同角度对多维数据集合进行分析比较。OLAP 由上卷、下钻和切片三个基本的数据操作构成。上卷是数据的聚合,即数据可以在一个或多个维度上进行累积和计算,如将部门预算的每月预算金额进行叠加,可得出部门年度预算金额。下钻是一种从汇总数据开始向下浏览细节数据的技术,如将年度预算金额按预算科目进行下钻,可得出各个科目具体的预算金额。切片则是获取 OLAP 立方体中的特定数据集合,从不同的视角观察这些数据,如年度预算可从时间维度分析每月具体预算情况,也可从预算科目维度分析各个科目的具体预算情况。预算管理内部控制的数据挖掘,可利用决策树、支持向量机、随机森林等算法进行研究。

3. 预算管理内控评价建模。预算管理内部控制评价主要对预算组织管理体系、预算编制、预算审批、预算下达、预算执行、预算追加调整、决算和考评八个环节的控制进行评价分析,而每个环节又细化为多个控制评价分析点,以预算编制流程合规性为例,对其内部控制评价进行建模。根据《内控规范》和《重庆海事局预算管理办法》,判断预算编制流程是否合规,具体需从两个分析点进行判断。

分析点 1:判断预算是否按照“二上二下”进行编制,其中“二上二下”主要包括预算建议数上报、预算控制数下达、预算草案数上报、预算批复数下达。而“二上二下”每个流程中是否都有必要的数据是判断预算是否按照“二上二下”进行的依据,这些必要数据包括预算部门、操作人、时间、金额、编制依据和审核依据。若数据完整即流程合规,若数据不完整即流程不合规,对应的分析评价核心 HQL 语句如下:

```
Hive>
Select
  a.I_Tap,a.I_Inputer,a.DT_Time,a.M_Money,b.VC_
Name
From
  SYS_Budget_PreInputer a join SYS_Unit b
on
  a.I_Depot=b.ID and a.I_Inputer is not null
  and a.DT_Time is not null and a.M_Money is not
null
order by a.ID
```

上述 HQL 语句中,SYS_Budget_PreInputer 为预算编制表,SYS_Unit 为部门信息表,I_Tap 为预算编制流程所处环节(存储数据为 1、2、3、4,依次代表预算建议数上报、预算控制数下达、预算草案数上报和预算批复数下达),I_Inputer 为操作员,DT_Time 为操作时间,M_Money 为预算金额,VC_Gist 为依据,VC_Name 为部门名称。

分析点 2:判断编制流程中的四个环节的时间是否为依次进行,若不为依次进行,则数据无效,对应的分析评价核心 HQL 语句如下:

```
Hive>
Select
  a.DT_Time,b.VC_Name
From
  SYS_Budget_PreInputer a join SYS_Unit b
on
  a.I_Depot=b.ID and a.I_Tap=1
```

上述 HQL 语句中,a.I_Tap=1 的意思为编制流程处于第一状态(预算建议数上报),查询此时的时间,再将 a.I_Tap 的值分别赋予 2、3、4,即编制流程状态依次为预算控制数下达、预算草案数上报和预算批复数下达,查询当时时间,最后判断 4 个时间是否遵循先后顺序。

主要参考文献:

- [1] 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.
- [2] 程平,李雪燃.数据视角下基于财务云平台的预算管理内部控制审计——以重庆海事局为例[J].财会月刊,2017(23):103~108.

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