

**【编者按】**财政部对行政事业单位内部控制建设提出明确要求,即以量化评价为导向,开展单位内部控制基础性评价工作,规范行政事业单位内部控制报告的编制,促进行政事业单位内部控制信息公开。而在大数据、云会计、人工智能背景下,数据仓库是一个面向主题的、集成的、相对稳定的、反映历史变化的、主要用于组织决策的数据集合,它可以很好地满足行政事业单位内部控制评价中的数据分析需求,促使内部控制评价实现信息化、科学化和智能化。基于此,本刊推出以重庆海事局为研究对象的行政事业单位内部控制评价系列文章,包括单位层面和业务层面共7篇,探讨大数据下行政事业单位内部控制评价数据仓库体系结构的设计、数据仓库的构建、基于数据仓库的内部控制评价方法等问题。

# 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

程平<sup>1,2</sup>(博士生导师), 范洵<sup>1,2</sup>

**【摘要】**单位层面内部控制作为行政事业单位内部控制的基础,是内部控制正常运行的基本保障。数据仓库技术能够为内部控制评价提供各种类型的海量数据支持,以实现单位层面内部控制科学、智能的评价。以重庆海事局为例,依据《行政事业单位内部控制规范(试行)》构建单位层面内部控制评价体系,设计单位层面内部控制评价数据仓库体系结构,分析数据源和数据的抽取、转换与加载过程,依次从确定主题域、划分分析粒度、确定维度表、设计模型四个阶段详细阐述单位层面内部控制评价数据仓库的构建,并基于Hadoop Hive数据仓库的HQL语言结合案例分析单位层面内部控制评价的具体实施。

**【关键词】**数据仓库; 行政事业单位; 单位层面; 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)13-0071-6

## 一、引言

随着大数据、云会计<sup>[1]</sup>、人工智能等现代信息技术的不断发展及其在行政事业单位的运用,数据量呈几何式增长,仅仅依靠传统的数据统计分析技术已无法满足当前内部控制(简称“内控”)评价的需求。而数据仓库作为一个面向主题的、集成的、相对稳定的、反映历史变化的主要用于组织决策的数据集合,可以很好地满足单位层面内控评价中的数据分析需求,促使行政事业单位单位层面内控评价实

现信息化、科学化和智能化<sup>[1]</sup>。现有文献大多停留在对内控评价指标的构建、评价方法、评价实施等问题的宏观探讨上,很少涉及具体应用场景的运用分析以及结合大数据技术对单位层面内控进行评价的探究。鉴于此,本文以重庆海事局为例,拟设计行政事业单位单位层面内控评价数据仓库体系结构,分析数据源和数据的抽取、转换和加载过程,详细阐述单位层面内控数据仓库的构建过程,并探讨HiveQL语言数据查询分析和OLAP多维分析在单位层面内

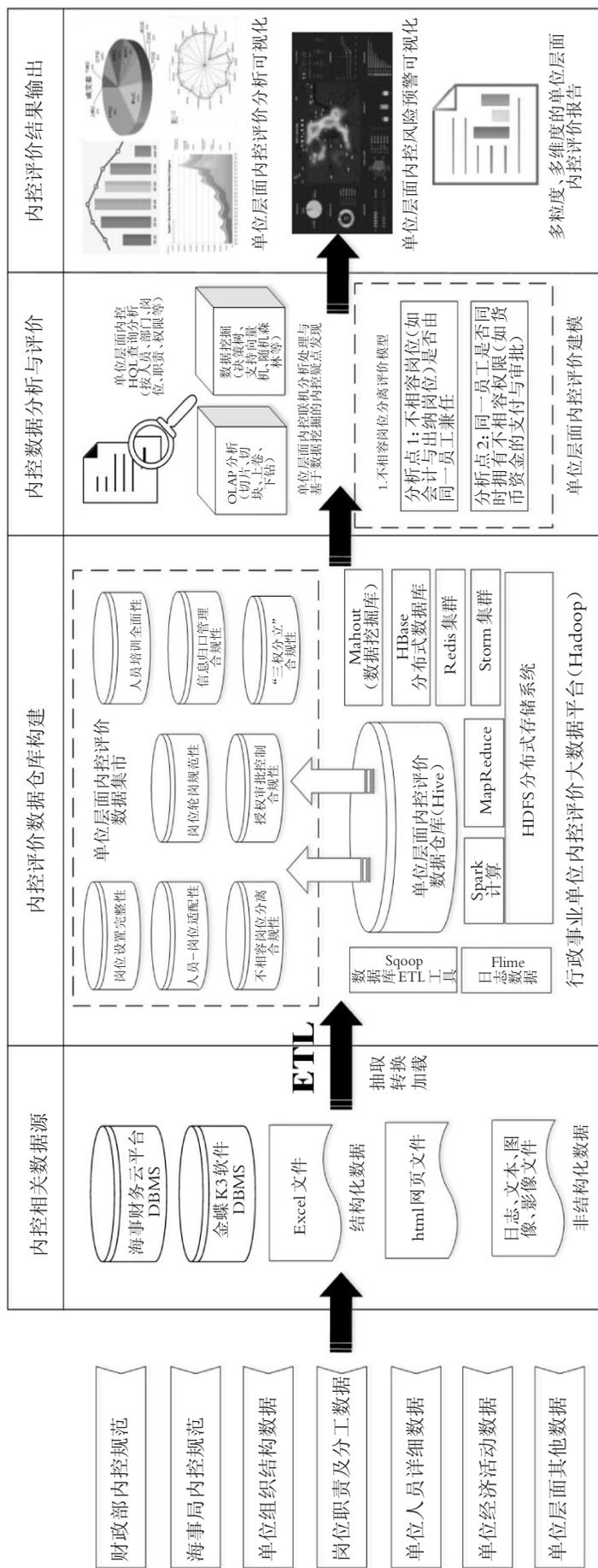
**【基金项目】** 国家社会科学基金项目(项目编号:17BGL194); 重庆海事局基于财务云平台的内部控制信息化系统建设项目(项目编号:2018Q60)

控评价中的具体应用。

## 二、单位层面内部控制评价数据仓库体系结构设计

**1. 设计思路。**目前，行政事业单位已逐步通过信息系统来记录并处理经济业务活动，虽积累了大量的历史数据，但由于系统之间分散运行，信息无法共享，导致无法全面采集数据。行政事业单位各个下属的管理系统记录和采集的数据在最终汇总时很容易出现数据的失真、缺失等现象，这些质量偏低的数据将影响管理者的分析和评价。在大数据技术背景下，数据仓库作为一个用于支持决策管理的、与时间相关的、稳定的数据集，可以很好地解决以上问题，将大量分散、结构不一致的数据转化为集中的、结构一致的、可利用的信息，从而实现对单位层面内控精准、智能的分析评价。以重庆海事局为例，其主要业务活动包括负责长江一带水上交通安全、航海保障、船舶和水上设施检验、环境保护、海洋管理等。通过实地调研发现其业务有以下特点：海事局资金全部来源于财政拨款，随着规费征收量和资金总量的不断增长，财务风险较大，特别是“十二五”期间是海事跨越式发展阶段，基本建设项目投资规模大，对资金和资产的监管任务重，而其单位财务人员相对较少，会计核算工作量大。因此，重庆海事局需要将数据仓库技术运用到内控建设中，并不断完善和优化，最终形成一个高度统一、资源整合、信息共享的行政事业单位内控信息化评价系统，以帮助管理者实时了解单位的内控风险点，进行精准有效的分析及智能的风险预警。

依据《行政事业单位内部控制规范(试行)》(简称“内控规范”)的单位层面相关规范进行内控评价，其根本是面向单位层面内控评价进行数据仓库体系结构设计的过程，其设计思路是对于单位的部门、员工权限等相关数据先通过数据抽取、转换和加载到数据仓库中进行集中存储和管理，再按照星型模型或雪



重庆海事局单位层面内部控制评价数据仓库体系结构图

花模型组织数据来建立若干数据集以构建单位层面内控数据仓库,最后利用HiveQL查询语言或者OLAP工具从数据仓库中读取数据并进行内控智能分析和评价,从而形成单位层面内控评价的数据仓库体系结构。

**2. 设计过程。**构建数据仓库是为了形成一个存储和使用数据的集成环境,将分散的结构化、半结构化及非结构化的数据转换为集中、一致的信息,以此辅助管理者利用这些业务数据对单位层面内控进行有效评价。重庆海事局单位层面内控评价数据仓库体系结构的构建,是将历史数据经过ETL(数据抽取、转换、加载的过程)输出标准的结构化数据加载至数据仓库,并通过Hadoop分布式文件系统进行存储,从而进行智能分析与评价。数据仓库体系结构见图。

图中,单位层面内控评价数据源主要来自重庆海事局单位内控制度文件、海事局财务云平台员工岗位、权限的基本信息和人员轮岗交接数据、金蝶K3系统的收入支出以及资产等业务中审核、审批等流程节点的数据。源数据处理完成后,基于Hadoop架构的单位层面内控评价大数据平台,构建形成包括不相容岗位分离合规性、人员培训全面性、岗位轮岗规范性等共八个主题数据集的单位层面内控评价数据仓库(Hive)。内控数据分析与评价则是利用HiveQL语言,根据评价指标的分析点建立单位层面内控评价模型,运用OLAP工具实现对数据仓库中的数据上卷、下钻、切片、切块的多维分析与智能评价。数据仓库还提供如饼状图、直方图、折线图等统计图以进行单位层面内控评价结果的实时可视化展示和风险预警,并能自定义输出各种多粒度、多维度的单位层面内控评价报告。

### 三、单位层面内部控制评价的数据采集及清洗

数据源是数据仓库的基础,分为内部数据和外部数据。当前行政事业单位经营业务数据大多分散于各个业务部门,数据割裂和私有化现象突出,由于业务标准不统一,数据质量普遍偏低,致使数据集中难、应用难,直接影响了其内控评价准确度。因此,为对行政事业单位内控的精准评价,数据采集的全面性和清洗质量至关重要。长期以来,重庆海事局投入使用的有办公管理系统、运行管理系统等多种信息系统,覆盖多个业务,涉及的数据源较广泛,主要涉及财务云平台、金蝶K3账务处理软件等内部平台数据源以及html网页文件、档案信息等外部数据源,主要涵盖单位组织架构、关键职能部门控制、岗

位分工等数据。因此,为实现对单位层面内控的有效评价,需集合各个内部业务系统和外部来源数据,对宏观层面的单位、员工相关数据进行全面采集。

人员贯穿了整个单位层面的内控,是单位层面内控的重要一环。重庆海事局财务云平台记录了单位员工的历史数据,包括员工的基本信息、家庭住址、岗位分配、轮岗记录等数据,涉及的数据表有:员工信息表、员工基本信息表、员工家庭信息表、员工附件信息表、员工地区信息表、工资查询人员表、操作员信息表、人员组织情况表、用户分组信息表、人员岗位交接表。单位层面内控要求对内控关键岗位进行控制,在权限分配时涉及单位的部门岗位。财务云平台已对重庆海事局的部门岗位信息进行记录,数据库表主要有:部门表、用户角色表、单位类型表、单位所在地区表、部门数据详情表、页面权限角色表等,以及平台外部数据如单位内控相关部署文件等。单位层面内控还要求对信息系统实行归口管理,按国家赋予的权利和承担的责任各司其职,防止重复管理、多头管理。涉及财务云平台的数据表主要有签报流程表、审批流程表等。为满足单位对员工培训的要求,需相应数据对员工培训的完成度和规范性进行评价,涉及的数据表有:培训参与人员表、培训计划表、培训附件表及培训通知和相关培训资料等。

数据采集完成后,上述数据表是用于单位日常经济业务,没有统一的录入标准,数据质量不高。所以,在把这些有用的数据加载进数据仓库之前需要对这些原始数据进行ETL标准化处理,将其整理成适用于数据仓库分析的数据,即数据清洗。

### 四、单位层面内部控制评价数据仓库的构建

**1. 确定主题域。**数据仓库设计首先是确定数据仓库的主题及相互关系,确定内控管理设计的边界。内控规范为建立和实施单位层面内控提供了标准依据并提出强制性要求。本文以内控规范为根本,融合单位的具体需求背景对程平、孙瑜<sup>[2]</sup>中的重庆海事局内控指标体系进行改造和优化,提出单位层面内控的评价主题域,见表1。

依据以上建立的内控评价指标体系,在重庆海事局内控数据仓库设计中,每个指标单独形成一个数据仓库主题,确定了如表1所示的八大主题。

**2. 确定分析粒度。**数据粒度是指数据仓库中保存数据的细化或综合程度,数据仓库中海量的数据库表以何种粒度存储,直接影响数据仓库的存储、查询质量以及是否能够满足用户的分析需求。依据重

表1 重庆海事局单位层面内控评价主题域确定及描述

主题域名称	主题域描述	内控规范依据
岗位设置完整性	单位岗位设置是否完整全面,例如是否设置预算业务管理、收支业务管理、政府采购业务管理、资产管理、建设项目管理、合同管理以及内部监督等经济活动的关键岗位	第十三条 单位应当单独设置内部控制职能部门或者确定内部控制牵头部门,负责组织协调内部控制工作。同时,应当充分发挥财会、内部审计、纪检监察、政府采购、基建、资产管理等部门或岗位在内部控制中的作用
授权审批控制合规性	明确业务经办人员职责范围与权限以及在办理相关业务时是否得到相应的授权	第十三条 应当充分发挥财会、内部审计、纪检监察、政府采购、基建、资产管理等部门或岗位在内部控制中的作用。同时,依据业务层面内部控制中预算(第十九条)、收支(第二十五条)、政府采购(第三十三条)、资产(第四十条)、建设项目(第四十六条)、合同(第五十四条)业务控制的“明确相关岗位的职责权限”规定
议事决策机制完整性	单位是否对重大经济事项建立健全集体研究、专家论证和技术咨询相结合的议事决策机制	第十四条 单位应当建立健全集体研究、专家论证和技术咨询相结合的议事决策机制
“三权”分立合规性	经济活动决策、执行、监督三个权限相互分离的控制机制是否嵌入人员权限分配和实际业务流程	第十四条 单位经济活动的决策、执行和监督应当相互分离
不相容岗位分离合规性	单位不相容岗位是否实现职责权限分离	第十五条 单位应当建立健全内部控制关键岗位责任制,明确岗位职责及分工,确保不相容岗位相互分离、相互制约和相互监督
关键岗位轮岗规范性	是否遵循单位内部轮岗制度	第十五条 单位应当实行内部控制关键岗位工作人员的轮岗制度,明确轮岗周期。不具备轮岗条件的单位应当采取专项审计等控制措施
人员培训全面性	根据相关政策及制度,结合本单位内部控制拟定的培训目标,各部门人员是否进行相关专题培训,如内控实施过程中的责任等内容	第十六条 单位应当加强内部控制关键岗位工作人员业务培训和职业道德教育,不断提升其业务水平和综合素质
人员一岗位匹配性	对应岗位员工的学历、资格证书是否符合岗位要求	第十七条 内部控制关键岗位工作人员应当具备与其工作岗位相适应的资格和能力

庆海事局的实际业务,为了提高存储和访问数据的效率,采用双重粒度设计<sup>[3]</sup>,将数据处理为真实完整的细节数据和轻度综合数据,分权限层级查询,权限层级越高可查询粒度越细,以“人员培训全面性”为例,对培训时间进行粒度划分,包括轻度综合(如2018年1月的培训人员ID、培训项目ID)、真实档案(如2018年1月1日的培训人员ID、培训项目ID)。

**3. 维度表设计。**维度是统计和分析数据的角度,在选取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。以人员培养全面性为例,可通过培训项目信息维度、人员组织情况维度、培训标准维度、培训类型维度、部门维度以及时间维度,来具体分析单位员工业务培训和职业道德教育等是否按照规定进行定期培训以及培训是否达到预期效果。

**4. 数据仓库模型设计。**单位层面内控评价各个指标具有多维特性,而且整个数据仓库中包含多个事实表,事实表之间共享多个维度表,不同的分析主题中存在着紧密的相关性和共享性。由于以上特性,

基于目前常见的星型模型、雪花模型和事实星座模型三种多维数据模型,结合具体分析应用需求对单位层面内部控制指标进行逻辑建模,具体模型类型、事实表、维度表及描述见表2所示。

选取“人员培训全面性”主题为例,通过培训开始时间和结束时间更新人员培训记录,使不同时间段的培训记录分布在一条时间轴上,从而可得到任一时间点的培训记录,事实表结构设计见表3。

### 五、基于数据分析的单位层面内控评价应用

在重庆海事局单位层面内控评价数据仓库中,利用HQL语言在Hive中创建用于分析的多维数据分析表,能够从不同维度进行切片、切块以及挖掘和钻取,满足管理者对单位层面内部控制评价的决策需求。下面通过单位层面内部控制的员工岗位设置方面的三个指标进行详细描述和相关数据的查询分析。

**1. 不相容岗位分离合规性。**不相容岗位分离控制对于发挥控制作用和实现内控目标起决定性作用,其目的是牵制和制衡,是内控核心的体现。在进行内控穿行测试时,不仅可以判断岗位是否为不相

表 2

行政事业单位单位层面内控评价数据仓库模型及其数据组织

主题域名称	数据仓库模型	事实表	维度表	模型描述
不相容岗位分离合规性	雪花	NK_DW_SS_RI GHT	NK_DW_WD_RoleNorm(不相容岗位标准维度表)	查询对比如编制岗与审批岗、业务经办岗与会计核算岗等不相容岗位是否由同一人担任
			.....	
人员—岗位匹配性	雪花	NK_DW_SS_Job	NK_DW_WD_Education(学历信息维度表)	比对员工是否与对应岗位所限制的有学历、资格证书等要求匹配
			NK_DW_WD_License(资格证书信息维度表)	
			.....	
岗位轮岗规范性	星型	NK_DW_SS_Us erRe	NK_DW_WD_UserReplace(人员岗位交接维度表)	查询员工职位变动表,内控制度规定轮岗的岗位是否按期、合规进行轮岗
人员培养全面性	雪花	NK_DW_SS_Tr aining	NK_DW_WD_TrainPlan(培训计划维度表)	查询员工培养计划表与实际培训记录,是否符合内部控制规定的员工培训类型与周期
			NK_DW_WD_TrainType(培训类型维度表)	
			NK_DW_WD_Unit(部门维度表)	
			.....	
“三权”分离合规性	星型	NK_DW_SS_Th ree	NK_DW_WD_ROLERIGHT(权限维度表)	查询单位经济活动的决策、执行和监督职能是否相互分离
			.....	
授权审批控制合规性	雪花	NK_DW_SS_SQ SP	NK_DW_WD_BusinessFlow(预算流程维度表)	查询预算、报销、合同签订收付、资金收入支出等业务流程中的经办人员是否拥有相应权限
			NK_DW_WD_BXFlow(报销流程维度表)	
			NK_DW_WD_Contract_Detail(合同流程维度表)	
			NK_DW_WD_PayInfoout(收入支出流程维度表)	
			.....	
议事决策机制完整性	星型	NK_DW_SS_YS	NK_DW_WD_Meeting(会议纪要维度表)	查询议事过程中的参与人员是否包括党委、行政、纪检等部门;对重大经济事项是否进行议事决策
			.....	
岗位设置完整性	星型	NK_DW_SS_Ro leFull	NK_DW_WD_UNIT(部门维度表)	查询单位所有岗位是否设置完整,重点观察内部控制关键岗位是否设置完整
			.....	

表 3

人员培训全面性评价物理表结构

表名	表包含字符
人员培训全面性事实表 NK_DW_SS_Training	I_User【人员ID】I_TrainProject【培训项目ID】I_TrainNorm【培训标准ID】BL_TrainState【培训状态(进行中1;结束0)】DT_TrainStart【培训开始时间】DT_TrainEnd【培训结束时间】DT_TrainTime【培训时长】
培训项目信息维度表 NK_DW_WD_TrainPlan	I_TrainProject【培训项目ID】VC_TrainName【培训项目名称】VC_TrainDec【培训内容】I_TrainType【培训类型】DT_TrainDay【培训天数】VC_TrainAddress【培训地点】I_Depot【培训部门ID】
人员组织信息维度表 NK_DW_WD_UserGroup	I_User【人员ID】I_gender【人员性别】VC_UserName【人员姓名】I_Depot【部门ID】I_Role【角色ID】I_Fuction【权限ID】DT_CreateTime【入职时间】DT_LeaveTime【离职时间】BL_UserState【人员状态(停职0;工作1)】VC_Note【备注】
培训标准维度表 NK_DW_WD_TrainNorm	ID【主键】I_TrainProject【培训项目ID】DT_TrianSC【规定时长】DT_TrianZQ【规定周期】VC_TrainState【预期效果】
培训类型维度表 NK_DW_WD_TrainType	I_TrainType【培训类型ID】VC_TrainType【培训类型名称】VC_TrainDec【培训类型描述】
部门维度表 NK_DW_WD_Unit	I_Depot【部门编号ID】VC_DeName【部门名称】VC_DeState【部门状态(停用0;启用1)】VC_DeArea【部门地区】VC_DeType【部门类型】VC_REMARK【部门备注】

容岗位,同时也可判断同一员工是否拥有两种或以上不相容权限,比如资金管理的支付与审批权限、审批与执行岗,若同时担任不相容职务或拥有不相容权限,极有可能发生舞弊、掩盖舞弊的行为。在对该指标进行评价中,可通过对系统中人员、岗位及权限等信息的采集,对不相容岗位分离有效性进行评价。连接操作员信息表(SYS\_USERINFO)、员工信息表(SYS\_EMPLOYEE)、用户分组信息表(SYS\_GROUPUSER)、角色信息表(SYS\_ROLEINFO)、角色权限表(SYS\_ROLERIGHT)、页面功能表(SYS\_FUNCTION)、部门表(SYS\_UNIT),可查询出单位每个员工的所任岗位及权限信息,与不相容岗位标准表(NK\_DW\_WD\_UNROLE)中的信息进行比对,检查是否出现同一个员工同时从事不相容岗位或拥有不相容权限。具体实现命令如下:

```
Hive>
Select d.ROLEDESC, b.ID,b.VC_NAME,b.VC_CARD,
g. VC_NAMESIMPLE ,
d.ROLENAME, f.displayname, f.nodeurl from SYS_USE
RINFO a
Join SYS_EMPLOYEE b on b.ID = a. I_EMPLOYEE
Join SYS_GROUPUSER c on c.I_USERID = a.userid
Join SYS_ROLEINFO d on d.ID = c. I_ROLEID
Join SYS_ROLERIGHT e on e.RoleID = d.ID
Join SYS_FUNCTION f on f.nodeID= e.nodeID
Join SYS_UNIT g on g.ID = d. I_DEPOT
Order by d. ROLEDESC
```

**2. 岗位轮岗规范性。**适时轮岗可有效防止内部腐败,特别在行政事业单位中,适时轮岗是内部控制建设的重要一环。行政事业单位需要制定详细的轮岗制度,明确轮岗周期以及轮岗过程中的规范工作进程、工作文件移交双方的责任与义务等。可通过连接员工信息表、人员交接表、角色信息表,查询出轮岗的员工是否按照规定进行定期轮岗,以及业务管理、收支业务管理、政府采购业务管理、资产管理、建设项目管理、合同管理以及内部监督等内部控制关键岗位是否进行定期、合规轮岗。具体实现命令如下:

```
Hive>
Select a.I_OldUser,b.VC_NAME, a.I_OldDepot,a.I_New
Depot,
c.ROLENAME,c.ROLEDESC from Dz_Sys_User Repl-
ace a
```

```
Join SYS_EMPLOYEE b on a.I_OldUser=b.VC_CODE
Join SYS_ROLEINFO c on a.I_OldDepot=c.ID
```

**3. 人员培训全面性。**人员培训是单位组织效益的重要途径,能够提升员工的综合素质,提高效率和单位形象,因此人员培训也不容忽视。以人员培训全面性指标为例,在Hadoop平台上利用OLAP,能够快速、准确地实现多层次、多角度、深层次的数据挖掘、统计和分析,并直观、多角度地反映数据分析结果。

利用HQL语言在Hive中创建用于分析“人员培训全面性”事实表,具体实现命令如下:

```
Hive>create table NK_DW_SS_Training,
>(People_ID STRING,
>Train_Project_ID STRING,
>Train_Standard_ID STRING,
>Train_State_ID STRING,
>Train_Start STRING,
>Train_End STRING);
OK
```

创建数据仓库和事实表后,用户可直接用HQL语言分析Hive数据仓库中的事实表NK\_DW\_SS\_Training。通过OLAP操作对培训时间按所需粒度进行切片查看,即可分析该时间线上整个单位员工的培训情况,例如对培训时间按照月粒度进行切片分析,实现命令如下:

```
Hive>SELECT*FORM NK_DW_SS_Training GROUP
BY DT_TrainStart;
```

通过以上语句可查询到重庆海事局每月的培训人数、培训项目是否符合单位内部的培训制度要求,并结合可视化工具以图表形式在前端展示给用户。

### 主要参考文献:

- [1] 程平. 云会计环境下、数据和系统对会计信息质量的影响[J]. 重庆理工大学学报(社会科学版), 2016(7):80~88.
- [2] 程平,孙瑜. 基于财务云平台的单位层面内部控制优化研究——以重庆海事局为例[J]. 会计之友,2016(4):150~154.
- [3] 丁祥武,黄忠睿. 使用微软商务智能平台构建医疗数据仓库[J]. 计算机应用与软件,2013(11): 238~242.

**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