

文章编号: 1005-8451 (2011) 11-0014-05

## 铁路信息综合库的研究与应用

苑晓明<sup>1</sup>, 陈光伟<sup>2</sup>

(1.北京铁路局 信息处, 北京 100860; 2.铁道部信息技术中心, 北京 100844)

**摘要:** 指出综合多变的需求与相对独立的信息应用之间的矛盾是当前铁路信息系统面临的主要矛盾之一, 提出构建基于信息资源规划和数据仓库的铁路综合信息库 (RIIS) 是解决之道, 结合信息资源规划和数据仓库的研究, 给出 RIIS 的具体应用。

**关键词:** 信息资源规划; 数据仓库; 铁路信息; 信息整合

**中图分类号:** U29 : TP39 **文献标识码:** A

### Study and application on Railway Integration Information Server

YUAN Xiao-ming<sup>1</sup>, CHEN Guang-wei<sup>2</sup>

(1. Information Technology Department, Beijing Railway Administration, Beijing 100860, China;

2. Information Technology Center, Ministry of Railways, Beijing 100844, China)

**Abstract:** This paper pointed out that the main issue of currently Railway Information System was the contradiction between the composite and changeably requirements and the relative independent information applications, brought forward that to construct Railway Integration Information Server(RIIS) based on Information Resource Planning(IRP) and Data Warehouse (DW) was a way to solve this problem, combined the study on IRP and DW, showed the application of RIIS.

**Key words:** information resource planning(IRP); data warehouse(DW); railway information; information integration

铁路运输管理信息系统 (TMIS) 等一批铁路信息系统的建成投产, 铁路主要业务基本实现了信息化, 但作为铁路运输主要组织者的铁路局 (集团公司), 其信息系统仍然存在很多问题, 不能满足业务部门日益高涨的需求。经过仔细研究和项目实践, 本文提出在铁路局构建基于信息资源规划和数据仓库的铁路综合信息库 (RIIS) 是解决此问题的一条有效途径。

### 1 当前信息系统面临的主要矛盾与原因分析

(1) 激烈的市场竞争越来越要求铁路局以全方位、综合准确的信息来分析自身生产、经营的情况, 对自身的运营状况做出全面的预测和分析, 发货企业也需要了解从提报计划到货物运到的全面及时的信息。相对综合多变的需求, 现有信息系统相对独立、综合功能不强的矛盾越来越突出。

(2) 铁路各级信息部门负责制订执行信息规划, 但由于信息规划偏宏观、缺乏具体指导手段等原因, 规划对信息系统研发的具体指导性不强。新

的信息系统多数仍然自成体系, 虽然个别系统也尝试整合已有的相关系统, 但这些系统普遍存在用户面狭窄、涉及信息系统少、数据质量不高、数据权威性不足等问题。占用大量资源的重复开发、重复整合和重复维护, 造成很大浪费。

(3) 由于没有统一的公共平台提供数据共享, 各信息系统间通过接口相互联接, 经常是每个系统建立多条接口直接联接到目标数据库表上, 一个目标数据库表可能同时被多个系统关联, 接口复杂混乱。这样的接口大多是可读又可写, 带来极大的安全隐患。

(4) 铁路业务部门各自为政, 信息系统也相对独立; 信息系统采用项目化、分散化开发与维护, 各项目组之间缺乏有效的协调与组织, 更深层的原因是铁路信息资源的管理观念相对落后于需求, 铁路信息资源缺乏有效的、操作性强的信息规划所致。

### 2 信息资源规划

信息资源规划 (IRP), 是把信息作为企业 (组织) 重要资源之一进行战略规划设计的方法。信

收稿日期: 2010-01-22

作者简介: 苑晓明, 高级工程师; 陈光伟, 研究员。

息资源规划的实质是在做好信息资源规划的基础上,建立主题数据库。主题数据库最大的特点是面向主题而不是面向某个程序,是“共建共用”而不是“自建自用”,从而实现信息资源的整合与共享,实现信息系统集成。铁路信息系统突出问题的解决应该借鉴 IRP 思想,对铁路信息资源有效规划,构建统一的数据库平台,实现资源整合、信息共享。进行信息资源规划的主要步骤如下:

(1) 进行需求分析,建立业务模型。需求分析阶段是要对业务流程进行梳理,熟悉明确现有业务流程并进行适当优化,按业务性质把主要业务划分为不同的业务域-业务过程-业务活动。

(2) 对应业务模型,建立功能模型。分别对应业务模型建立功能模型:子系统-功能模块-程序模块。

(3) 进行数据分析,建立数据模型。按照业务模型和功能模型对整个信息资源数据进行分析、规整,定义主题数据库模型。

(4) 建立信息规范和主题数据库。在数据模型的基础上,进行规范化处理,形成信息规范和主题数据库。

### 3 数据仓库

数据仓库(DW)的定义是:面向主题、集成、随时间变化的非易失性数据的集合,用于支持管理层的决策过程。数据仓库也对不同主题的历史信息集成,目的是分析统计和决策支持等。

构建数据仓库的主要步骤是:

(1) 进行需求分析,划分主题域,设计概念模型。对业务流程进行分析,划分为不同的主题域。通常的数据仓库概念模型分为星型、雪花型两种。

(2) 设计数据仓库逻辑模型。逻辑模型是对概念模型的细分,通过逻辑模型的设计,可以进一步得到事实表、维表等模型。同时,还可以把关于聚集的设计和建立统一信息规范等也涵盖其中。

(3) 设计数据仓库物理模型,构建数据仓库。物理模型用来确定表的存储方式以及表间的索引模式,对数据仓库的性能进行优化等。

(4) 建立数据仓库管理模型。包括数据抽取

(ETL)、转换、清洗、装载,对数据仓库进行安全性管理和用户授权管理;对数据仓库数据加载和数据清洗的管理;元数据管理等。

(5) 建立在线分析处理(OLAP)多维分析模型。

### 4 信息资源规划和数据仓库的结合

信息资源规划的目的是把信息看做重要资源,整合信息资源,提高资源使用效率;数据仓库的作用是整合历史信息,提供辅助决策等。只有抓住主要技术实质,才能使两者无缝结合。

结合信息资源规划和数据仓库,通过分析 C-U 矩阵,根据 C 的创建信息,决定数据抽取(ETL)提取的数据源和提取方式;根据 U 的使用信息,形成面向不同用户的数据集市。

图1描述综合资源库的建立过程,数据存储模型分为3部分:

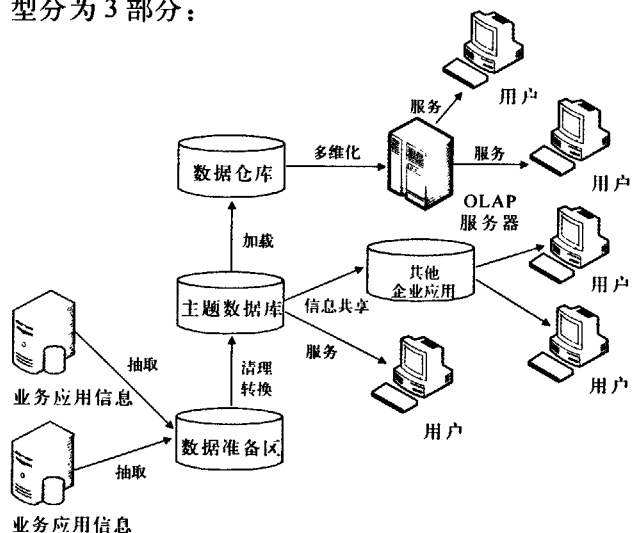


图1 数据存储模型

(1) 数据准备区,这部分视实际情况可以灵活配置。

(2) 数据集成区即主题数据库。

(3) 数据仓库。

在 ETL 和统一信息规范的基础上建立的主题库主要功能是:

(1) 提供数据整合与共享功能,可以直接为其他系统或用户直接提供服务。

(2) 作为数据仓库的一部分,为数据仓库提供数据基础。

(3) 用数据仓库的统一信息规范处理主题数据库数据,这样更有利于主题库数据的规范化,有

利于主题库与数据仓库的结合。

不论是信息资源规划还是数据仓库,都要进行需求分析,并在需求分析的基础上,划分业务域或者主题域。无论业务域还是主题域,都是以业务需求为导向划分主题。划分业务域的结果是建立主题数据库,划分主题域的结果是建立不同主题的数据仓库。业务域与主题域不同之处在于,面向业务域建立的主题数据库的数据是一次一处输入系统,以保证其准确性和完整性,而面向主题域的数据仓库的数据则可能是重复使用的。

在分析信息资源规划和数据仓库的形成机制基础上,实现二者的结合应遵循如下步骤:

#### (1) 需求分析,划分业务域和主题域

不论是数据仓库还是信息资源规划,都是在明确用户需求的基础上,划分不同主题。

#### (2) 设计业务域业务模型和功能模型

按照划分的业务域,细化为不同的业务过程,对应设计功能模型:子系统,功能模块-程序模块。

#### (3) 设计数据仓库概念模型

按照划分的主题域,设计对应的数据仓库概念模型。针对特定主题域,可以用实体-关系(E-R)图进一步分析实体内在关系。星型数据仓库模型查询性能较高,雪花型模型灵活性较好。

#### (4) 设计主题库数据模型与数据仓库逻辑模型

建立主题库数据模型,不但要在对业务模型和功能模型分析的基础上,还要在ETL和统一信息规范的基础上进行,这样的主题库不但是信息整合和共享的中心,还是数据仓库的一部分和数据基础。在主题库数据模型的基础上,建立数据仓库逻辑模型,其中包括事实表、维表、聚集模型、统一信息规范等。

#### (5) 设计数据仓库物理模型

物理模型用来确定表的存储方式以及表间的索引模式,对数据仓库的性能进行优化等。

#### (6) 建立主题数据库,构建数据仓库

在主题库数据模型和数据仓库物理模型的基础上,建立主题数据库,构建数据仓库。

#### (7) 建立数据仓库管理模型和多维分析模型

管理模型包括ETL设计,对数据仓库进行安全性、用户授权管理;对数据仓库数据清洗、加载策略管理;元数据管理等。基于多维分析的OLAP技术可以提供多维分析展示功能。

## 5 铁路综合信息库的功能

(1) 整合铁路局各主要信息资源并经过处理、加工,能快速、灵活、全面提供各种查询、报表和分析功能,以及辅助决策、综合效益评价、商业智能(BI)等功能,满足用户各种需求。

(2) 成为一个权威性数据支撑平台,其他项目可以在该平台上二次开发,避免重复整合开发。

(3) 具有高度可扩展性,对符合标准的新数据源可以很方便地引入,并且不会明显降低系统性能。

(4) 统一管理安全和用户授权,主要信息系统形成一个出口并严格管理控制,从而消除安全隐患。

## 6 铁路综合信息库的构建

结合铁路信息应用实际,我们建立的RIIS采用Oracle数据库,对来自不同业务应用的数据经过抽取、清洗转换和加载,加工形成主题数据库、数据仓库,并进行OLAP分析。

### 6.1 需求分析,划分业务域和主题域

进行信息资源规划,以业务需求为导向,整理业务部门关心的内容,设计形成业务主题,同时,要熟悉不同数据仓库主题的构建方法,筛选不同业务主题的信息源。

综合分析铁路主要业务活动,结合多年铁路信息技术建设的经验,我们把主要铁路业务分为货物、客流、机车车辆、调度、财务收入、计划、货运营销和客运营销等8个业务域;把主要铁路业务分为装车情况分析、货运客户营销、货运效益分析、运输能力分析、运输效益分析等主题,数据仓库开发一般采用螺旋式周期性开发方法,各主题可以单独设计,逐步构建。篇幅限制,文内主要对货物业务域和装车情况分析主题域为例进行说明。

### 6.2 货物业务域业务模型和功能模型的设计

货物业务域业务模型和功能模型如图2以及图3。

### 6.3 设计数据仓库概念模型

考虑到装车主题域信息数据数量大、业务逻辑相对集中的特点,建立多维(日期维、发收货单

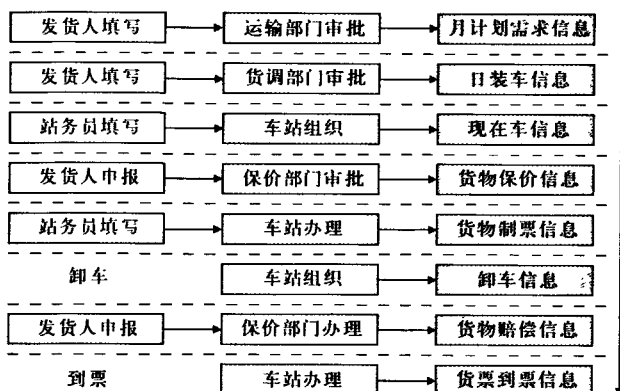


图2 货物业务域业务模型

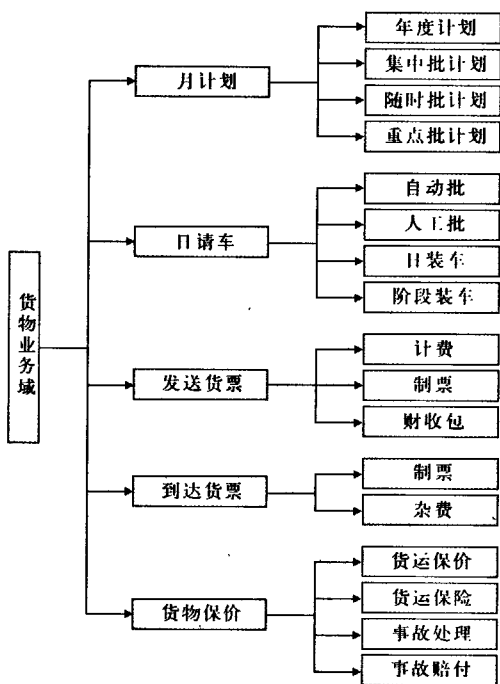


图3 货物业务域功能模型

位维、装卸单位维、品类维等)的星形模型是理想的选择。

在装车主题域中,需要在货运月计划、货运调度、货票、保价等信息基础上,以“计划号”关键字关联建立计划-实际事实表,以日期维、品类维表、发收货人维表、装卸车单位维表作为维表,构建星型数据仓库,如图4。

#### 6.4 设计主题库数据模型与数据仓库逻辑模型

按照业务特点和源信息结构,货物业务域建立4个主题数据库:月计划库、日请车库、发送货票和到达货票。

装车情况数据仓库的事实表模型:日期关键

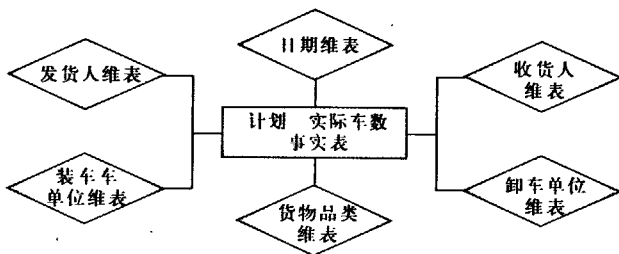


图4 星型数据仓库

字、发货人关键字、装车单位关键字、货物品类关键字、卸车单位关键字、收货人关键字、原提车数、月计划车数、日请求车数、日计划车数、实际装车数。

装车情况数据仓库的维表包括:日期维表、发货人维表、装车单位维表、货物品类维表、卸车单位维表、收货人维表等。每个维表分别有关键字字段,分别与事实表中的相应关键字对应。

#### 6.5 设计数据仓库物理模型

物理模型包括设计物理存储结构、建立联合索引、物理模型优化等。考虑到 Oracle 实体化视图可以加快对聚集的查询速度,我们把对数据仓库聚集等预处理数据存储在实体化视图中。

#### 6.6 建立主题数据库,构建数据仓库

在主题库数据模型和数据仓库物理模型的基础上,建立主题数据库,构建数据仓库。

#### 6.7 建立数据仓库管理模型

数据仓库管理模型包括 ETL 设计;对数据仓库进行安全性、用户授权管理;对数据仓库数据加载策略、清理管理;元数据管理等。基于多维分析的 OLAP 技术可以提供多维分析展示功能。

ETL的质量问题具体表现为正确性、完整性、一致性、完备性、有效性、时效性和可获取性等几个特性。在构建 RIIS 时,按统一标准规范清洗数据、进行空值处理、规范化数据格式、验证数据正确性、数据替换、对非法数据替换或导出等方法保证 ETL 质量。

为了加强安全管理、构建统一开发平台,RIIS 统一对系统和用户授权,既可以满足对数据查询需求,又可以达到安全共享目的。

根据业务特点,采取不同的信息加载策略。可以在晚上服务器空闲时期,如凌晨 1:00 开始,每隔 0.5 h 依次对不同信息资源执行加载、预处理等

(下转 P20)

其中,车站概况、铁路大事记、列车查询、车站公告、票务信息、铁路出行常识、换乘信息、旅游出行以及天气预报等功能模块,是在自助查询机上呈现给旅客查询所用;各车站可根据车站情况不同和地域不同,灵活增减内容。其中,列车查询提供对本站始发、途经、终到各车次的实时信息查询,票务信息提供本站始发和途经各车次的实时票务查询。

后台维护是在维护终端上由操作人员进行监视和控制的。接口数据管理是查看集成管理系统发送给查询系统的列车、余票等相关信息;设备状态监控可查看查询机设备的实时状态,如开关机状态、硬盘、CPU和内存等重要部件的运转情况;公告信息管理,由车站管理人员增加或删除本车站日常及临时公告信息,旅客可通过前台系统进行查看;广告维护,若系统需要引入商业广告进行展示,可在后台对广告内容、时长、上传和播放列表等功能进行维护。

通过对车站自助查询系统功能上的优化,更大幅度地发挥了系统的使用价值,有助于提高系统的使用率,以求得最佳的社会效益与经济效益。

## 5 结束语

旅客服务系统集中管理的模式更好地统筹了车站设备、应用服务等资源,使任务更集中、数据更统一,并且实效性更高。车站自助查询系统作为旅客服务系统一部分,实现了在架构上的优化和功能上的扩展,内容也更加充实,应用效果更加明显。

该系统已在高速铁路系统试验国家工程实验室的客运服务实验室试用,在实验室模拟运行环境下取得了良好效果。

### 参考文献:

- [1] 刘育欣,张彦,陈靖,朱韦桥.铁路旅客服务集成管理系统设计[J].铁路计算机应用,2010,19(5):17-20.
- [2] Yusuke MIENO.'On-train Regional Information Offering System'[J]. Japanese Railway Engineering, 2001, 6(147).
- [3] 史天运.铁路客运专线智能信息系统总体方案研究[C].2005年中国智能自动化会议论文集,2005.
- [4] 张霞,赵瑜.关于铁路旅客综合服务信息系统的研究[J].交通运输系统工程与信息,2004(2):64-67.

责任编辑 杨利明

(上接 P17)

任务,并输出相关监控信息到日志表。

元数据是指关于数据的数据,可以对数据进行详细描述和说明,对数据仓库的使用有非常重要的作用。元数据分为静态和动态两种,静态主要与数据结构有关,动态主要与数据的状态与使用方法有关。

超立方多维分析是OLAP技术的核心所在,多维分析可以采用切片、切块、旋转、上卷、下钻、钻透等操作,可以为用户展现多角度、多侧面的视图。采用BO服务器作为OLAP服务器,可以满足各类决策支持和统计分析查询的需求。

## 7 结束语

构建基于信息资源规划和数据仓库的铁路综合信息库(RIIS)并不能一蹴而就,需要长期的过程去逐步完善。我们在北京铁路局设计并构建了RIIS,通过这个平台,基本实现了主要货运信息

规划整合,实现了跨信息系统的货运客户关系管理、各种台账及货运指标综合查询统计分析、到达预测、后期运输指标辅助决策等功能,效果良好。我们还计划以此平台为基础,进行铁路局客货运输效益评价等系统的研发。

构建(RIIS)可以加强铁路局信息规划的指导性,打破既有信息系统相对独立体系,减少信息系统安全隐患,为解决铁路局信息系统面临的诸多矛盾提供了一条解决之道。

### 参考文献:

- [1] 陈京民.数据仓库原理、设计与应用[M].北京:中国水利水电出版社,2004.
- [2] 高复先.信息资源规划-信息化建设基础工程[M].北京:清华大学出版社,2002.
- [3] 张维曙.铁路局货运工作辅助决策系统技术报告[D].北京:北京铁路局,北京交通大学,2009.

责任编辑 杨利明