

基于行政服务的政务智能系统研究与设计^①

郑雄锋 侯济恭 (华侨大学 计算机科学与技术学院 福建 泉州 362021)

摘要: 针对行政服务中心中不断增加的数据量, 如何处理这些海量的数据已成为一个关键问题。由此, 引入了商业智能及其相关技术, 并以行政服务中心中的实际数据为背景, 提出基于行政服务的政务智能系统体系结构、系统框架、功能模块, 并做了较详细的说明。最后分析了数据仓库的创建过程。

关键词: 商业智能; 行政服务; 政务智能系统; 数据仓库

Research and Design of E-Government Intelligent System Based on Administrative Service

ZHENG Xiong-Feng, HOU Ji-Gong

(College of Computer Science and Technology, Huaqiao University, Quanzhou 362021, China)

Abstract: With the increasing amount of data in Administrative Service Center, how to handle these vast data has become a key issue. Thus, this paper introduces the Business Intelligence and its related technologies, taking the actual data in Administrative Service Center as background. It proposes the E-Government Intelligence System architecture, system framework, functional module, and explains in detail. Finally, it analyzes the process of building data warehouse.

Keywords: business intelligence; administrative services; e-government intelligence system; data warehouse

市行政服务中心是市为深化行政审批制度改革、推进网上审批而创立的。市行政服务中心通过几年的运行, 积累了庞大数据, 如何才能把这些数据财产转换为可靠的、有用的知识来辅助政府提高决策效率与准确性? 综合了数据仓库、OLAP、数据挖掘的商业智能的引入成为必然。

商业智能(Business Intelligence, 简称 BI)又称商务智能, 其核心技术包括数据仓库、联机分析处理、数据挖掘。数据仓库能够有效地存储和管理各类数据信息, 以提供一个集成的、可靠的和一致的统一视图^[1]。联机分析处理是一种分析技术, 具有汇总、合并和聚集功能, 以及从不同的角度观察信息的能力。数据挖掘是从大量的、无规律的、有噪声的数据集中提取或“挖掘”出有效的、可理解的、潜在有用的知识^[2]。其数据流程如下图 1 所示。

1 基于行政服务的政务智能系统体系结构

基于行政服务的政务智能系统就是利用商业智能

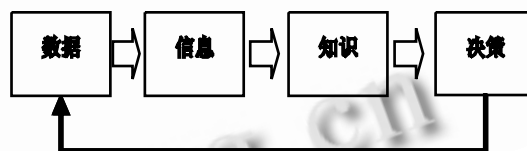


图 1 数据流程图

的一系列方法、工具与技术来“加工、提炼”行政服务中心的数据的系统。该系统具有以下几个特点:①所处理数据的海量、多样性与分散性。②较多人员的参与。该系统需要相关部门专业人员与技术人员进行协同合作。③服务对象的多样化与多层次: 服务对象包括政府领导、普通政府工作人员、分析人员以及群众。

基于行政服务的政务智能系统体系结构如图 2, 包含数据源层、数据存储层、数据分析层、表示层。数据源层中的数据主要来自: ①行政服务中心的一系列数据, 如网上行政审批一系列数据、监察数据等等

① 基金项目: 福建省科技重大专项资助 (2006HZ0001)

收稿时间: 2009-10-20;收到修改稿时间:2009-12-02

②外部数据,包括各种文档(如 Word, Excel 等)、书面材料等。数据存储层中存储的是经过抽取、转换装载后的数据。可以根据具体需要,细化各个主题,形成多个主题表,据此从数据仓库中选出多个数据子集,形成数据集市^[3]。数据集市的是面向特定部门(如工商局、建设局等)的特定功能需求而建立的,只面向某个特定的主题,如工商局关注的是各类企业投资情况等主题,而建设局关注的则是土地开发使用等一些主题。数据分析层是面向决策应用的,是在数据仓库基础上进行即时查询、报表生成、OLAP、数据挖掘等一系列操作。表现层是将数据分析层的结果通过政府信息门户网站等方式向公众、分析人员、政府人员展现。

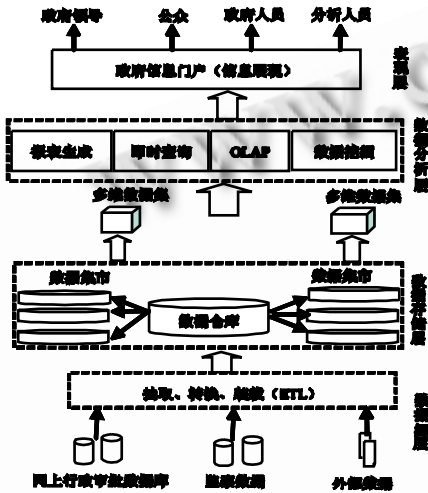


图2 基于行政服务的政务智能系统体系结构

2 行政服务的政务智能系统框架与功能模块

政务智能系统框架如图3所示。系统框架主要分为信息发布管理、数据仓库管理、数据分析管理、数据挖掘管理、系统管理。

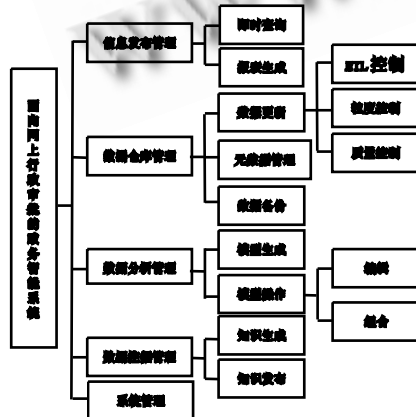


图3 基于行政服务的政务智能系统框架

基于行政服务的政务智能系统功能模块如图4,包括数据仓库管理模块、数据分析管理模块、数据挖掘管理模块、信息发布模块、系统管理模块。

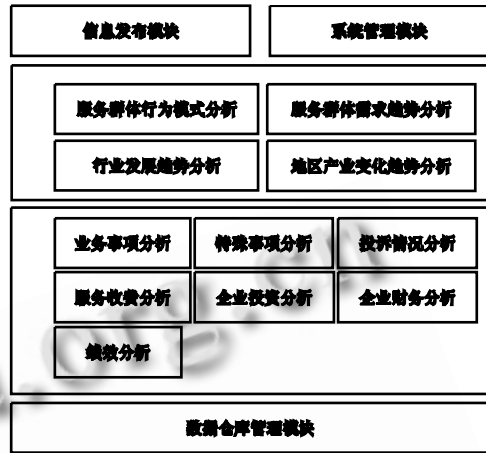


图4 基于行政服务的政务智能系统功能模块图

(1) 数据仓库管理模块。该模块是系统的核心模块,包括数据的ETL(抽取、转换、装载)、数据质量控制、数据粒度控制、元数据管理等。主要实现将大量异构数据源通过有效可靠的抽取、转换、集成、装载等操作形成统一的信息存储在数据仓库中。数据仓库中可以使用表分区(如时间分区、列表分区、散列分区)、位图索引等技术来存储这些信息,以提高其访问查询效率。

(2) 数据分析模块。由于行政服务中心拥有现成大量的历史业务数据,因此该模块就以行政服务中心的数据平台为基础,结合各种分析领域的实际需求抽取出包含有各自独立的逻辑内涵的主题,并且确定每一个主题所应该包含的数据内容,将其对应为一个分析对象。比如,业务事项分析、投诉情况分析、企业投资规模分析等。然后对每个主题利用OLAP技术生成不同模型,并可以根据需要,对不同模型进行组合从多角度、多层次深入分析生成另外的分析模型。

(3) 数据挖掘模块。该模块通过利用数据挖掘技术对不同的主题多方面进行深入分析,比如,可以利用Oracle Data Mining(ODM)工具中的关联规则、决策树、聚类算法等数据挖掘方法,实现对如行业的发展变化趋势、各地区的产业变化趋势及分布情况、服务群体行为模式分析、服务群体需求趋势分析等。从而能够辅助政府从多方位了解某个地区的经济发展状况、大众的需求趋势等。

(4) 信息发布模块。该模块将一些基本统计信息、

OLAP 分析结果信息、数据挖掘结果信息以多种表现方式向公众、分析人员、政府人员展现，并提供预定义的查询功能和报表生成功能。比如通过将建立好的元数据导入到 Oracle Business Intelligence enterprise edition(OBIEE)中，利用 OBIEE 中的 BI Answers 来实现^[4]：完全的即席查询分析；以多种方式展现——图形，报表，交叉表等；交互式生成报表；可保存和共享个人报表。然后利用 Interactive Dashboards(简称仪表盘)设计合理简洁的页面来显示各种各样的信息，比如表格、图片、图表、文字、文件等等。用户可通过访问 Web 站点，经过身份安全、权限认证后访问该数据和进行各种分析、查询操作，同时预留与应用系统的接口，实现各种功能的无缝连接，并提供集成化的认证、信息发布和管理环境，是系统使用人员无需关心具体的技术实现途径，即可实现对系统数据的访问和分析^[5]。

(5) 系统管理模块。主要包括对用户一些信息的管理、系统的一些设置、日志管理等。

3 基于行政服务的数据仓库的设计

数据仓库已成为数据分析和联机数据分析处理日趋重要的平台，并将为数据挖掘提供有效地平台。因此，创建高效合理的数据仓库显得至关重要。

3.1 数据仓库逻辑模型设计

(1) 确定主题。事实表用于存放大量的事实数据，通常都很大，而且非规范化程度很高；维度表用于存放描述性数据，它是围绕事实表建立的较小的表。事实表中的数据是不允许修改的，新数据只是简单地增加进去。维度表的数据可以改变，每一个维度表通过一个关键字直接与事实表关联^[6]。根据行政服务中心拥有现成大量的历史业务数据，结合各种分析领域的实际需求抽取包含有各自独立的逻辑内涵的主题。比如，业务事项分析主题即对行政服务中心各部门的行政审批事项的各种情况，包括受理数、受理成功率、正常办结数、超期办结数、正常办结率等情况做出分析，并与上年该时间段作对比分析^[7]。业务事项(WorkEvent)事实表如图 5 所示。

基于行政服务的主题共细化为 7 个，包括业务事项主题、特殊事项主题、投诉情况主题、企业投资信息主题、企业财务信息主题、服务收费情况主题、绩效分析主题。其星型模型汇总如图 6。

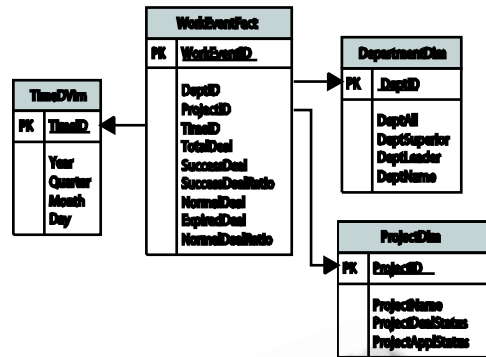


图 5 业务事项事实表

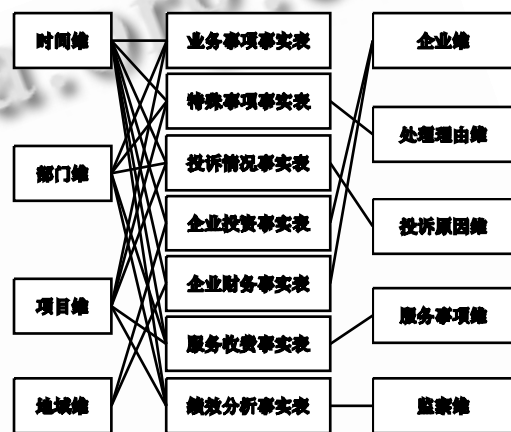


图 6 基于行政服务的星型模型汇总

(2) 粒度选择。粒度是数据仓库中数据单元的细节程度或综合程度的级别。粒度问题是设计数据仓库的最重要的方面，对数据仓库环境所处的整个体系结构都有影响^[8]。而基于数据仓库的多维分析一个显著优点是可以对维度进行任意细化(下钻)和汇总(上钻)操作。因此，维度的粒度直接影响多维数据集的浏览层次。但是，实际应用中粒度也不能无意识的细化，而要根据实际需要进行处理。行政服务中心中的数据仓库包括两种类型的数据：当前细节数据和轻度综合数据。当前细节数据是来自操作系统的细节数据，行政服务中心的每一条审批记录、每一条投诉信息都被放入数据仓库中，其数据量相对轻度综合数据大的多。轻度综合数据是经过一定程度综合的细节数据，比如对审批数据可以按周进行汇总审批量，对投诉信息、企业注册信息按月进行汇总，而对企业财务信息的时间维粒度则只能确定到“年”，因为企业的一些有关财务信息是表示一个会计年度的信息。根据主题应用之分析目的，以上划分按概化到细化进行，

可按需进行适当调整。因此对数据仓库采用双重粒度级别进行设计。

(3) 数据分割策略。选择正确的数据仓库分割策略可以极大的提高数据仓库的访问查询效率。由于行政服务中心的数据量较大,可以使用表分区技术(如范围分区、列表分区、散列分区等)对数据进行分割存储,并可结合使用合适的索引技术(如 B 树索引、位图索引等),同时将表基础数据与索引数据分别存放到不同的表空间上。

3.2 数据仓库物理模型设计

3.2.1 ETL 处理

ETL(Extract、Transform、Load)主要负责对行政服务中心中大量分布广泛的、异构的数据进行抽取,对数据按照统一制定的转换规则进行转换、集成,并按照一定的规则进行数据清洗,最后把经过转换、清洗的高质量数据按计划增量或全量装载到数据仓库。

ETL 步骤如下:①明确数据源中数据的结构和含义。首先需要了解行政服务中心中数据源的详细情况,包括所在计算机平台、拥有者、数据的结构、语义、内容和异常,然后评估数据的质量以及数据需要清理和转换到的程度。②数据抽取。根据已确定好的主题,从异构数据源中抽取相关数据到数据仓库中。③数据转换。主要包括数据的转化(数据的合并、汇总、过滤等)、数据的重新格式化和计算、关键数据的重新构建和数据总结、数据定位^[9]。④数据清洗。行政服务中心中的原始数据可能存在错误类型,这些类型包括单个数据源中的错误和不一致,以及多个数据源中的模式冲突,语义定义不一致等,所有这些数据都必须处理,若系统不能自动处理,还必须借助手工操作来完成^[10]。⑤数据装载。数据经过清洗、转换、一致性等操作后,就可以按预先定义好的维度粒度装入到数据仓库中。

3.2.2 数据仓库的创建

主要包括确定数据的存储结构、确定索引策略、确定数据存放位置和确定存储分配。如以创建业务事项(WorkEvent)事实表及其维表为例。可以采用本地管理的表空间(自动段空间管理),索引与表放在各自不同的磁盘上。对于业务事项事实表上的外键列的索引使用位图索引,因为他们的基数较小,其他表上的主键列则使用 B*索引。严格按照如下步骤建表:①创建事实表及其维表,表中主键、外键等一些约束先不创

建。②创建事实表及其维表各主键索引。③创建事实表上的外键位图索引。④为事实表和维表添加主键约束。⑤为事实表添加外键约束。

4 结语

基于行政服务的政务智能系统具备了强大的数据管理、处理、分析能力,并为政府决策者提供了科学可靠的决策支持。虽然基于行政服务的政务智能系统的创建是一项长期的过程,尤其是数据仓库,需要以反复迭代的方式进行建造,根据实际的应用和实践经验不断进行调整,以最大限度满足需求。但随着政府在电子政务智能化方面资金投入的加大,相关技术人才培养与引进,在不久的将来,基于政服务的政务智能系统将更加成熟与完善,并极大地影响着政府的办事效能。

参考文献

- 1 Inmon W H.王志海等译.数据仓库(原书第四版).北京:机械工业出版社,2007.
- 2 Han JW, Kamber M.范明,孟小峰译.数据挖掘概念与技术(原书第二版).北京:机械工业出版社,2007.
- 3 郑洪源,周良.商业智能解决方案的研究与应用.计算机应用研究,2005(9):92-93.
- 4 Stackowiak R, Rayman J, Greenwald R. Oracle Data Warehousing and Business Intelligence Solutions. Indianapolis: Wiley Publishing Inc., 2007.
- 5 侯济恭.政府智能系统体系结构.计算机工程与设计,2007,(18):4494-4497.
- 6 周亮.电子政务决策支持系统中数据仓库的研究与设计.武汉理工大学学报(信息与管理工程版),2005,(1):31-34.
- 7 侯济恭.政府智能系统的核心技术.武汉理工大学学报(信息与管理工程版),2009,(1):58-62.
- 8 谢燕,郑有才,张立勇等.数据仓库技术在呼叫中心中的应用.计算机工程与设计,2006,27(21):4150-4152.
- 9 宋鹏,廉继红.ETL 技术在复杂数据迁移项目中的应用.西安工程大学学报,2008,22(4):493-494.
- 10 王克龙,王玲,王平等.数据仓库中 ETL 技术的探讨与实践.计算机应用与软件,2005,22(11):30-31.