

应用 ODS 技术解决电子政务系统数据一致性问题

张 宇¹, 薛永生¹, 张延松¹, 林坤辉², 韩绍兵¹

(1. 厦门大学计算机科学系, 2. 厦门大学软件学院, 福建 厦门 361005)

摘要: 讨论了 ODS 技术在电子政务系统中的应用. 将 ODS 引入到电子政务系统中形成更为合理的 DB-ODS-DW3 层结构, 并通过 ODS 记录系统和参考表的使用进行全局联机事务处理, 使各业务数据库内容可以实时更新, 保持数据的一致性. 从根本上解决密切相关的业务数据库数据不一致的问题.

关键词: 电子政务; ODS; DB; DW; 参考表

中图分类号: TP 311.12

文献标识码: A

随着现代信息技术特别是网络技术的飞速发展,“网上政府”、“电子政务”的建设方兴未艾,正成为信息化进程中的一个热点.我国政府对电子政务建设非常重视,已将其列为信息化建设的重点任务.

多年来,各地方、各部门、各级政府都建立了不少政务数据库,但由于政府各职能部门间相互独立、缺乏协调和统一,造成各业务数据库系统内大量数据重复,而本应一致的数据又不一致的现象严重.电子政务中经常遇到全局性的 OLTP(On-Line Transaction Processing)问题,即一个事务同时涉及到多个部门的数据.在事务要使用的数据应当是一致的.此时不一致的数据不仅使业务处理不便,而且也给基础信息资源库的建设造成麻烦.解决各部门数据不一致的问题不仅是为基础信息库的建设提供前提,更是完善日常业务处理的当务之急.

解决信息不一致与共享问题的主要方法有交换式共享和发布式共享两种方法.交换式共享是指在需要信息共享的各个政府部门之间通过统一的 XML 信息交换中心进行信息交换,通过权威信息源与各数据需求信息库之间的数据交换来统一数据,这是一种在各个部门之间通过数据交换方式实现的信息共享与信息统一;另一种方法是建立统一的共享信息库,将政府各部门业务处理所需的基础信息

统一发布到基础信息库中,需要权威基础信息的部门通过统一的数据交换平台从基础信息库中提取所需的信息,完成本部门数据与权威数据提供部门数据的一致,共享相关部门的数据信息.这是一种“推拉”式结构,权威信息提供部门将其他部门所需共享的信息按一定的数据更新策略与时间策略“推”到基础信息库中,需要共享其他部门信息的政府部门按相应的数据访问策略从基础信息库中将所需的信息“拉”到部门数据库中,实现信息的共享与一致性更新.在这两种信息共享方法中,发布式共享对于业务部门具有更好的可扩展性,而且通过集中的基础信息库可以进行面向整合数据的决策分析与趋势预测,通过不同部门的异常数据分析来预测突发事件.所以,发布式共享具有更好的实用性与发展前景.

1 电子政务中的信息资源共享与一致

“十五”期间,我国电子政务建设的工作重点是围绕“两网一站四库十二金”展开,整合政务信息资源,统一平台,统一标准,最终消除信息孤岛,实现信息资源的快速交换与全面共享.这一框架提出要加快建设两个统一的电子政务网络平台:一个是主要承担各级政府的办公业务和其他业务的“政务内网”,一个是主要处理企业、公众服务业和政府部门之间业务的“政务外网”;先行建设 4 个数据库(即人口基础信息库、法人单位基础信息库、空间地理和自然资源基础信息库、宏观经济数据库),为国家提供最基础的共享数据资源^[1].

数据资源的整合、共享是电子政务建设的重要

收稿日期:2004-05-09

作者简介:张宇(1977-),女,在职硕士研究生.工作单位:哈尔滨金融高等专科学校.

内容,是整个电子政务的基石,实现信息共享,对统一、一致的数据进行操作,各部门电子政务才能顺畅流动,才能充分发挥电子政务巨大的社会效益和经济效益。各部门建设、维护的信息资源是全市共有的重要资源,信息资源实行统一管理。政府信息以共享为原则,不共享为法定例外。基础信息库的建设是为了提供权威的数据供应用部门使用。

电子政务需要信息资源在各政府部门间的共享。长期以来,由于我国政府采取的垂直管理方式,各级政府各职能部门之间壁垒森严,各自为政,存在许多数据交叉采集问题,大量相同信息重复存储,从采集到存储的过程中由于各种原因造成本来应该相同的数据却不一致的现象。数据的不一致同时影响到政府部门间的业务整合,极大地影响着优化办公流程、实现“一站式”服务的实施。

在信息资源库的建设过程中,需要对各原有的业务数据库系统数据进行整合和集中,虽然在数据进入基础信息前经过了清洗、整理,达到最终入库的数据一致,但并不能从根本上解决业务系统数据不一致的问题。在一定时期内,各职能部门业务数据库中的数据仍然存在矛盾的现象,不一致的数据给公民、法人在政府各职能部门办理实际业务带来极大不便。

将 ODS(Operational Data Store) 技术引入电子政务中可以在一定程度上解决数据不一致问题。

2 基于 ODS 的电子政务系统

数据仓库技术的出现将数据划分为操作型数据和分析数据。数据仓库技术在电子政务信息资源建设中的应用提供了全局的、综合的、分析型的数据供决策部门使用。对于操作和分析应用的不同需求,电子政务也有基于 DB(Data Base) 的政务系统和基于 DW 的政务系统。一般的,电子政务的建设是将 DB 和 DW(Data Warehouse) 结合起来,形成 DB-DW 的双层体系结构。

基于 DB 的政务系统是面向应用和联机事务处理的,其处理的数据是当前各分散业务的细节数据,应用于日常业务的处理,其数据随业务处理而实时更新。此系统不能提供集成、统一的数据环境,数据共享程度低,不能将各业务数据之间的内在关联关系等信息挖掘出来,只能满足基本业务操作使用,不能提供决策支持环境。

基于 DW 的政务系统主要是面向分析、高层决策支持的,能够提供集成、统一、面向主题的数据环

境。将大量数据组织起来提供一致的共享,并通过对数据的挖掘分析为政府决策提供数据支持。

但基于数据仓库的政务系统需要一定量的数据源作为建立数据仓库的基础,数据仓库的投资在短期内很难见效;同时由于追求高层决策能力最终会导致系统的使用率降低。而目前政务信息系统在系统建设规模、即时 OLAP(On-Line Analytical Processing) 和数据要求等方面都没有完全具备采用完整数据仓库方案的条件。此外对于现有大量繁杂的业务系统,建设数据仓库之初,很难提出清晰完整的建设需要,并且对于建立全局级的大规模数据仓库,项目实施的周期和难度将很长,并且投资大、风险高,因此目前数据仓库系统并不一定是最佳方案。

操作数据存储 ODS 可以提供另外一种电子政务系统建设思路。

操作数据存储 ODS 是用于支持日常的全局应用的数据集合,其数据是面向主题组织的,是可变的,数据是当前或者近期的。ODS 中的数据在进入 ODS 之前经过清洗和整理,达到集成和一致的目的,在全局上保持一致,这使得 ODS 不同于传统分散的操作型系统,类似于数据仓库。但 ODS 中的数据是可以进行联机修改、存放的是当前或近期的数据,又有别于数据仓库^[2-4]。如:ODS 中保存现值和近期值,数据仓库中保存历史数据;ODS 中只保存细节数据,数据仓库中既有细节数据也有聚合数据;ODS 中的数据可以被更新,数据仓库中的数据不可更新等^[5]。

ODS 一方面满足全局应用,另一方面还为进一步建立企业数据仓库提供了一致的数据环境以供抽取,同时减轻了数据仓库的系统管理负荷。

ODS 按照数据更新的时间间隔分成 4 个等级,级别 I 是与数据源紧密同步等级,操作数据源的更新要立即反映到 ODS 中,数据更新以秒为单位;级别 II 提供以小时为单位的操作数据源与 ODS 的更新机制;级别 III 提供以天为单位的操作数据源与 ODS 的更新机制。级别 II 和级别 III 是常用的 ODS 更新机制,级别 I 由于技术上的代价过大,数据更新操作过于频繁而很少使用。级别 IV 是将预聚集、预分析的策略分析结果集存储于 ODS 中,提供在线的、实时的策略信息支持,这样,数据仓库就可以提供在线的、高性能的数据访问能力。

基于 ODS 的电子政务系统,ODS 是基于数据仓库所提出的面向主题的全局一致数据环境概念,为政务建设提供了多层次的信息处理环境,并建立

起 DB-ODS-DW3 层体系结构. 其中, ODS 作为一个中间层, 一方面包含全局一致的、细节的、当前或接近当前的数据, 可以进行全局联机操作型处理; 另一方面, 它是一种面向主题的、集成的数据环境, 且数据量小, 适用于辅助完成日常决策的数据分析处理. 可以先针对几个职能部门数据共享的需求建立局部一致的数据共享, 既能保证业务所需的一致数据, 又能提供综合数据以供分析. 随着政务系统的逐步发展和完善再推进数据仓库层的建设^[6]. 目前各级政府大都已经建立了较为完善的数据库应用系统, 从这些成功的应用系统中抽取数据来建立 ODS, 并最终形成一个完善的应用体系结构的技术路线更为可行. 因此, 当前政务信息系统的最佳选择是 ODS 解决方案.

3 基于 ODS 的电子政务系统对数据一致性的处理

ODS 在电子政务系统中的应用, 使原来 DB-DW 的两层结构转变为 DB-ODS-DW 的 3 层体系结构. 一个基于 DB-ODS-DW3 层体系的电子政务系统(图 1).

在 3 层体系结构中, 将原先在业务数据库系统和 DW 之间的保证数据一致性的工作从 DB-DW 接口转移到 DB-ODS 接口上, 数据在进入 ODS 之前先要经过清洗、转换, 保证进入 ODS 的数据是一致的. 数据仓库只需要定期从 ODS 中抽取数据作 OLAP 分析之用, 不必再做数据一致性工作. 当没有 ODS 时, 数据仓库的记录系统分散于各非集成的应用之中, 数据仓库的转换接口将很复杂, 还要处理好各应用在时间上的同步问题. ODS 的建立可以弥补了 DB-DW 两层体系结构中所存在的不足, 可以满足数据处理的多层次要求.

ODS 可以提供面向全局的联机事务处理, 相对

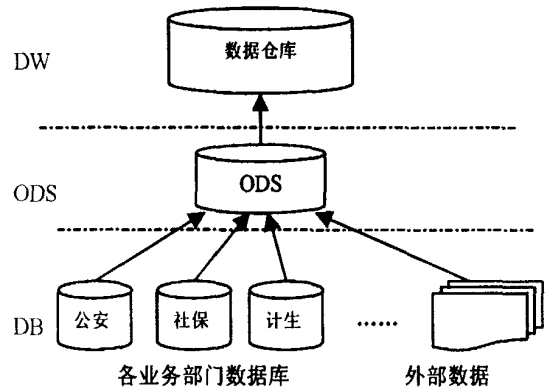


图 1 DB-ODS-DW3 层体系的电子政务系统

Fig. 1 DB-ODS-DW three tier government system

数据仓库的面向长期决策的分析处理, 在功能层次上也可以更好地划分系统. 而且, 在联机事务处理中数据能够及时得到更新, 保证数据是最新的而且是一致的, 极大地方便业务使用者.

对于数据仓库系统而言, 我们只能通过清洗和转换保证进入数据仓库的数据是一致的, 而业务数据库系统中的数据仍然保持原来的不一致状态, 在日常业务工作中, 使用者面对的仍然是不一致的数据. 数据一致性问题并没有得到根本性的解决. 为了实现企业级的 OLTP, 在业务中取得一致的数据, 必须在 ODS 与数据库之间建立一种双向映射关系, 以保持双方的一致性. 一方面, ODS 记录系统定义表达的是 ODS 与数据库的数据抽取关系, 或者说是各数据库中 ODS 的敏感区, 在记录系统上所做的任何修改操作都会反映在相应的 ODS 记录中; 另一方面, ODS 系统中还存放着一些参考表, 它所反映的关系是 ODS 全局更新时所必须反映到的所有数据库中相关记录的信息. 通过定义各级记录系统和参考表, 通过相互间的映射 ODS 系统可以实现同步更新, 保证各业务数据库数据的实时一致. 操作型环境中各分散的 DB 记录经过过滤后形成了 ODS 系统的记录系统, 向 ODS 系统中提供数据. 记录系统定

表 1 从数据库系统到数据仓库系统的记录系统定义

Tab. 1 Record definition from data base to data ware house

主题	表名	属性字段	源数据库	源表名	源字段	数据产生时间
企业基本信息	企业情况表	企业名称	工商部门数据库	企业信息	企业名称	02/03/14
企业基本信息	企业情况表	营业执照注册号	工商部门数据库	企业信息	营业执照号	02/03/14
企业基本信息	企业情况表	税务登记号	税务部门数据库	企业税务登记	税务登记号	02/03/04
.....

义了原有分散的 DB 中哪些数据送往 ODS,并指明与 ODS 数据相应的数据表。(表 1)通过 ODS 的定义可以描述如何把分散于应用的 DB 中的数据抽取到 ODS 中,形成了 ODS 中的面向主题的记录。而从 ODS 向 DB 方向则通过在 ODS 系统中的参考表来反映 ODS 全局更新时要反映到的所有 DB 中的相关记录。ODS 的每张表中都有时间属性的字段用于标识数据进入 ODS 的时间。

DB 中变化的数据要反映到 ODS 中。ODS 捕捉这些数据变化大致可以通过 3 种途径:

(1) 在应用级捕获,即在操作数据库 DB 中运行一个应用程序来感知数据的变化,ODS 中规定好哪些数据从哪个 DB 中取得,当应用程序发现那个数据变化则实时地向 ODS 写入相应记录。

(2) 使用 DELTA 文件,这时各变化数据的记录都被记入到 DELTA 文件中,ODS 只要定期读取文件内容则可更新数据。但这种方式的数据更新可能不是实时的,需要预先决定读取 DELTA 文件的周期。

(3) 在 DBMS 级获取。即由 DBMS 来感知 DB 数据的变化,并将后映像(After Image AI)记录写入到一个 DELTA 文件中,再由 ODS 处理。

在政府各职能部门之间,实际上某些相同的信息本身存在一定的权威与从属之分。因此对于相同信息的重复使用也可以定义一定的主从关系,以某一部门数据库的数据为权威,其它职能部门与之对照,建立映射。当主要数据库中的数据因业务操作产生变化时,从属数据库对照新的数据更新原来内容,保证数据的一致。例如,我们可以通过 ODS 在公安、计生、社保各部门间实现数据一致的体系。当公安部门的数据因业务操作而发生更改时,计生和社保的数据内容也应当随之变更。

这一过程通过记录系统的定义和参考表来实现(如图 2)。

记录系统说明如何从 DB 向 ODS 进行数据抽取的。数据从 DB 进入 ODS 的过程以元数据的形式记录下来,根据这些内容,当 DB 中的数据发生变化时就可以反应到 ODS 中。而参考表则是从 ODS 数据项到 DB 数据项的反向映射,记录的是 ODS 中数据的变化与哪些 DB 相关,这样,当 ODS 中的数据变化时,通过参考表将这些变化反映到 DB 中,引起其它 DB 内容的变化。从而达到多个关系紧密的数据库保持数据一致的目的。过程(1)公安数据库中的

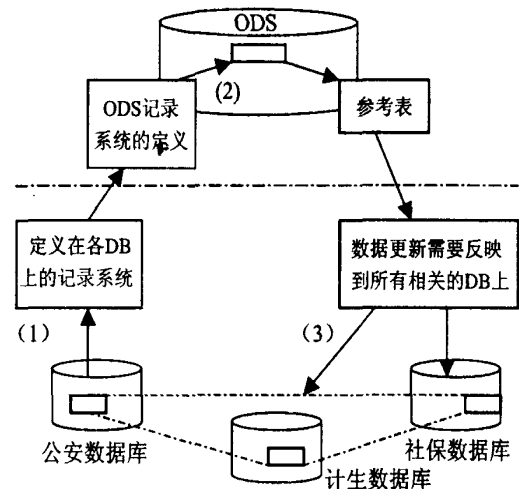


图 2 ODS 参考表

Fig. 2 ODS reference table

数据变化在过程(2)中通过记录系统反应到 ODS 中,通过参考表则在过程(3)中修改与公安数据库密切相关的计生和社保数据库,使 3 者的数据保持一致。

通过 ODS 的全局联机事务处理,将原有分散的数据库紧密地结合起来。一点动则全局变,保持多点数据一致,方便政府职能部门业务工作。

但是,应用基于 ODS 的电子政务系统以实现数据全局一致还需要政府各职能部门之间统一思想、协调关系,各业务数据库理清层次、分配权限,才能在具体业务操作过程中进行数据库内容的实时更新,在实际意义上达到数据一致的效果。

ODS 作为一种数据的共享集成方案,不仅要为信息使用者提供共享的统一数据,还要提供相应的数据分析管理工具来辅助用户的短期决策^[7]。TAS 技术(tactical analysis system,战术分析系统),集成了当前的 DSS,ODS,DM 和人工智能技术,提供了象 ODS 一样的近于实时的数据分析和短期决策支持,如通过对社保、医保帐户的异常大额支出与公民个人基本信息状况的综合分析来预测盗领或冒领现象。但这种决策分析是以短期数据和跨部门综合数据为基础的,面向短期性、综合性的政府分析决策,与基于数据仓库的长期数据分析与决策支持相辅相承,解决不同时间级别的趋势分析和预测问题。

4 DB-ODS-DW 结构

在 DB-ODS-DW 体系结构中,DB 作为业务数据

库,提供实时的、细节的业务数据。ODS 提供政府部门间的数据共享机制,可以提供准确的、权威的、综合的基础数据,为各部门业务系统消除数据重复、数据不一致起到了重要的作用。同时,ODS 以其高效的数据同步更新机制来保证共享数据的同步更新,保证数据的及时性和有效性。

ODS 保存的是短期数据,而电子政务的发展同样需要提供长期决策分析与支持,需要大量的历史数据,这就需要在 ODS 的基础上建立面向电子政务的数据仓库。ODS 所提供的基础共享数据完成了从各部门异构数据库中的数据采集、统一抽取,可以直接抽取到数据仓库中。

ODS 中存储短期数据,以数据的存储期限为边界确定 ODS-DW 数据抽取原则。DW 直接从 ODS 中抽取数据,ODS 从各部门的业务数据库中抽取相关的细节数据,在数据的存储周期内保存在 ODS 中,当超过数据的存储周期时,从 ODS 抽取到 DW,同时,清除 ODS 中原有的数据,抽取新的数据。这样,ODS 中只保存在数据存储周期内的数据,超过数据周期的数据转换到数据仓库中,形成由 DB-ODS 和 ODS-DW 的两级数据存储、处理机制,最终将数据保存在数据仓库中。数据仓库中保存的大量历史数据作为长期趋势分析和预测的基础数据,通过对网络各种经济资源的挖掘,确定未来经济的走势,从而制定出相应的宏观经济调控政策,为政府宏观管理的提供决策支持。

目前,电子政务的工作正在各个城市开展,其中首要的工作是城市基础信息建设。而基础信息库的建设要立足现状、展望未来,合理规划短期目标和长期发展计划,有效地整合信息、利用信息,将信息变为促进经济和社会发展的生产力,促进政府从行政

职能向服务职能的转变,更好地服务于公众,推进“Citizen-centered”的电子政府建设^[8]。

5 小 结

在电子政务工作中,数据的共享是电子政务建设的基础,然而现实生活中各级政府各职能部门间存在着大量重复却不一致的数据。基于 ODS 的电子政务系统不仅可以提供面向主题的、集成的数据,而且可以实现面向全局的联机事务处理,使全局数据始终保持一致状态,从根本上解决数据不一致问题。同时,DB-ODS-DW 结构为电子政务建设提供数据共享、数据分析基础,为政府的宏观管理和电子政务的远期发展提供有力的支持。

参考文献:

- [1] 顾文星. 电子政务信息资源库建设[J]. 上海统计, 2002, 11: 30 - 32.
- [2] Bill Inmon. Designing the operational data store[J]. DM Review Magazine, 1998, 7.
- [3] Claudia Imhoff. The operational data store: hammering away[J]. DM Review Magazine, 1998, 7.
- [4] Claudia Imhoff. A new class of operational data store[J]. DM Review Magazine, 2000, 7.
- [5] 林宇. 数据仓库原理与实践[M]. 北京: 人民邮电出版社, 2003.
- [6] 王三明. 基于 ODS 技术的政务信息系统方案探讨[EB/OL]. <http://tech.mitux.com/>. 2003-1-15.
- [7] Timothy K Hawes. The ODS she ain't what she used to be[J]. DM Direct Newsletter, 2004, 8.
- [8] Implementing the President's Management Agenda for E-Government E-Government Strategy. [EB/OL]. <http://www.whitehouse.gov/omb/ebov>. 2003-4.

ODS Solution of the Coherence Matter in E-government

ZHANG Yu¹, XUE Yong-sheng¹, ZHANG Yan-song¹, LIN Kur-hui², HAN Shao-bing¹

(1. Department of Computer Science, Xiamen University,

2. School of Software, Xiamen University, Xiamen 361005, China)

Abstract: In this paper, ODS was used to solve the coherence matter in E-government solution, the three-tire structure of DB-ODS-DW will be more suitable to the E-government solution. Using the record system and the reference table, data in different DB would be dynamic update. The ODS system also provides different levels of update strategies and tactical analysis system to support analysis and decision.

Key words: e-government; ODS; DB; DW; reference; table