

# 警务综合信息系统数据仓库的建设与实践

袁丽娜

(广州大学华软软件学院软件工程系 广州 510990)

**摘要** 在公安行业,随着信息化工作的逐步推进,仅仅使用业务数据库已经不能满足需求,数据仓库的建立已经迫在眉睫。根据公安行业的特点,详细描述了一个适合公安行业数据仓库系统的解决方案。重点阐述了数据仓库系统体系结构的设计,数据的抽取、转换和加载,多维模型的建立和数据仓库前端应用的展现。

**关键词** 数据仓库,公安行业,数据模型,ETL

**中图法分类号** TP311 **文献标识码** A

## Construction and Practice of the Data Warehouse in the Police Comprehensive Information System

YUAN Li-na

(South China Institute of Software Engineering, Guangzhou 510990, China)

**Abstract** In Public Security Industry, with the gradually promoting informationization work, only using business database to handle alert, case handling, security business work already can not meet the demand, so the construction of data warehouse is the must. According to the characteristics of public security, this paper gave a solution which is suitable for public security industry data warehouse system. In this project overall design and implementation phase, this paper expounded the data warehouse system of data extraction, conversion and loading, the establishment of multidimensional model, data warehouse front application show.

**Keywords** Data warehouse, Public Security Industry, Data model, ETL

随着全球信息化的快速发展,人类社会刑事案件、经济案件等犯罪手段也逐步高科技化,给公安部门信息化工作带来了新的挑战及机遇。怎样才能建立一个将公安行业信息化系统中的各类业务数据进行加工、处理,并转换为对公安决策有用的业务智能分析系统呢?那就是在公安信息化建设中引入数据仓库技术,因为只有通过数据仓库技术,才能够充分挖掘出深藏在繁杂的各类公安业务数据中的关键信息,并对公安相关业务进行科学合理的预测,为公安领导提供决策支持。

## 1 需求及体系结构设计

### 1.1 需求

在公安行业,随着信息化工作的逐步推进、“金盾工程”的全面实施,仅仅通过业务数据库来处理警情、案件办理、治安等业务工作已经不能完全满足需求,如何更好地利用这不断增长的数据,如何有效地建设与实施数据仓库系统,以实现公安数据资源的高效使用,已经是公安行业所关注的一个重要问题。

警务综合信息系统数据仓库主要基于警务综合信息系统业务库,包括案事件业务数据、治安管理业务数据、境外人员数据、常住人口、暂住人口等各类相关数据,构建警务综合信息系统数据仓库,可为“数据仓库应用平台”提供基础数据服务。通过警务综合信息系统数据仓库,实现警务综合信息系统等各类数据的共享、综合查询、统计分析、自定义报表、情报

导侦与外部职能部门的数据交换、决策支持分析等各种数据应用。

### 1.2 架构设计

基于以上各种用户需求的分析,及公安部统一标准、规范,使用数据仓库等方法及技术建立的系统架构如图 1 所示。

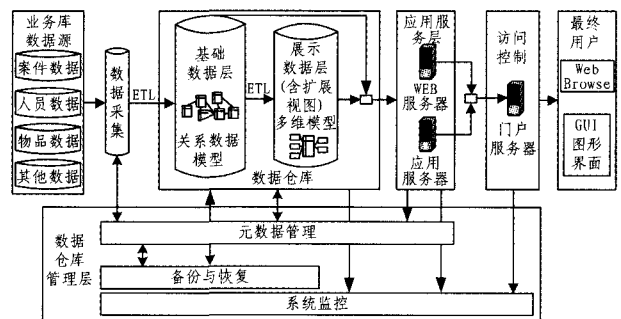


图 1 系统架构图

根据组建的基本功能,警务综合信息系统数据仓库的基本组建可以分成数据仓库基本功能层和数据仓库管理层两个层次。

数据仓库基本功能层包括实现数据仓库从数据集成到数据展现,为决策分析提供支持基本功能的基本模块。本系统的数据源主要包括案事件信息、治安管理信息、境外人员信息、常住人口信息、暂住人口、旅业信息、被盗车辆、在逃人员、违法人员及指纹信息等。

表 4 8类案件嫌疑人展示主题

宏观数据	指标数据	明细数据
案件分析	案件数	在各时间和案发地域维度的案件数
嫌疑人情况分析	嫌疑人总数	在8类案别和户籍区划维度的案件数

基于本数据仓库的应用平台采用基于 J2EE 技术的 B/S 3 层架构,即数据层、逻辑层、表现层。主要提供的服务类型包含动态分析、主题分析、情报研判、信息查询、报表等。

数据仓库管理层是围绕数据仓库的主线,是为了数据仓库安全、稳定及高效地运行而建立的模块。

基于本数据仓库,创建了警综信息系统应用平台门户网站。此门户网站为用户提供了一个统一的应用界面,是本数据仓库的统一入口。此门户系统集成所有应用,集成所有访问权限,控制每个用户仅访问到应该访问的数据,保证数据信息的安全。

### 3 数据模型设计

警务综合信息系统数据仓库总体上采用星型模型,以案件为中心,关联时间、地点、人、机构和物品。具体的一、二级分析主题设计如表 1 所列。

表 1 警务综合信息系统数据仓库一、二分析主题设计

一级主题	二级主题
人	常住人口、暂住人口、嫌疑人、在逃人员数据等
案件	刑侦案件信息、治安案(事)件信息、破案信息等
物品	涉案物品、机动车等
机构	涉外机构、治安管辖单位、涉案单位、特种行业场所等
地点	人的住址、案发地点等

基于数据仓库可面向实际的决策分析的应用需求,构建相应的多维立方体,以便于 OLAP 分析。而对于人这个主题,有常住人口、暂住人口、嫌疑人、在逃人员等二级主题,以下仅以嫌疑人的分析为例。嫌疑人的主题仍然采用星型模型存储,包含 8 类案件嫌疑人分析事实表及案别、嫌疑人户籍行政区划、案发地域、案发终值时间维表,如图 2 所示。

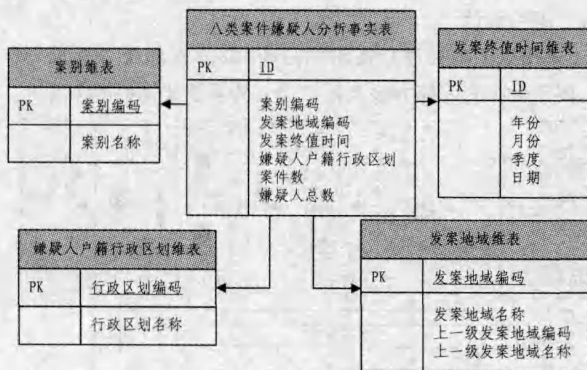


图 2 8类案件嫌疑人分析主题

8 类案件嫌疑人分析主题包含维度指标、量度指标、展示主题,分别如表 2—表 4 所列。

表 2 8类案件嫌疑人主题维度

嫌疑人分析主题	分析 8 类案件嫌疑人案发各个方面			
功能描述	维度层次			
案发时间终值	年	季度	月	
嫌疑人户籍区划	派出所	分局	市局	
案发地域	派出所	分局	市局	
案别	8 类案别			

表 3 8类案件嫌疑人主题量度

嫌疑人分析主题	分析 8 类案件嫌疑人案发各个方面	
量度	案件概况	案件数
	嫌疑人概况	嫌疑人总数

### 4 ETL 设计

至今为止,数据仓库的应用领域已经遍及各行各业,因而各大数据库厂商纷纷宣布其产品支持数据仓库,并提出了建立和使用数据仓库的产品,如 Oracle、Informaix、Microsoft、IBM 公司等。

根据其操作环境,ETL 工具可分为 3 种类型:交互型、编码型、代码生成器型。

由于公安行业业务规则繁杂,涉及的业务数据种类繁多,而且数据量也巨大,因此本系统采用编码型来实施 ETL。

为了节约成本,本系统数据采集服务器和数据仓库服务器为同一服务器。本系统将所有有用的业务数据采集到数据仓库中。再通过自编写的 SQL 存储过程中的数据清洗和数据加载功能包将采集到的数据清洗、加载到相应数据仓库事实表和维度表中。

ETL 除了包括第一次实施的数据抽取、转换和加载,还包括实施之后持续不断的数据追加和清理。

本数据仓库的数据追加和清理(即数据同步)系统利用 Oracle 自身 CDC 特性实现,CDC 特性能够实时捕捉业务库各个数据表的变化情况,利用发布和订阅机制可以记录业务库数据表在一段时间内的所有数据的操作情况,数据同步系统封装了 CDC 繁琐的配置过程,简化侦听内容,实现数据自动同步,无须人工干预。

同步 CDC 模式(Synchronous Mode)通过在源库上建立 trigger 的方式来捕获增量数据,因此可以做到实时抽取增量数据。当源库执行 commit 的时候,增量数据将生成于 change table 中。CDC 同步模式的基本架构如图 3 所示。

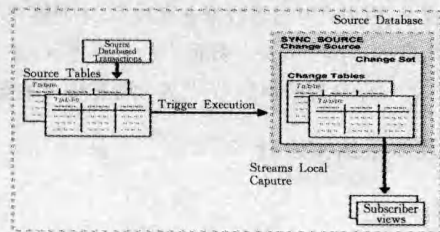


图 3 CDC 同步模式基本架构

### 5 前端应用展现

基于警务综合信息系统数据仓库,创建了警务综合信息系统应用平台门户网站。警务综合信息应用平台为一线干警和管理层提供了强有力的信息支持。警务综合信息应用平台包括系统维护、系统工具、应用工具、情报研判平台等,应用工具中包括动态分析、主题分析、报表和信息查询等。

#### 5.1 动态分析

本系统使用蓝葵商业智能工具,将数据仓库中的星型模型层的数据展示给系统用户,用户可以通过工具在数据立方体上实现上卷、下钻、切片等操作。

(下转第 308 页)

Sampling for Class Imbalance Learning[C]//Proc. Int'l Conf. Data Mining, 2006:965-969

- [17] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: Synthetic Minority Over-Sampling Technique[J]. J. Artificial Intelligence Research, 2002, 16: 321-357
- [18] Chawla N V, Lazarevic A, Hall L O, et al. SMOTEBoost: Improving Prediction of the Minority Class in Boosting[C]// Proc. Seventh European Conf. Principles and Practice of Knowledge Discovery in Databases. 2003:107-119
- [19] Breiman L. Bagging Predictors[J]. Machine Learning, 1996, 24 (2): 123-140
- [20] 沈学华, 周志华, 吴建鑫. Boosting 和 Bagging 综述[J]. 计算机工程及应用, 2000, 36(12)
- [21] Breiman. Random Forests[J]. Machine Learning, 2001, 45(1): 5-32
- [22] Freund Y, Schapire R E. Experiments with a new boosting algorithm[C]//Proceedings of the Thirteenth International Conference

on Machine Learning. The Mit Press, Cambridge, MA, Morgan Kaufmann, Los Altos, CA, 1996:148-156

- [23] Agarwal R, Joshi M V. PNRule: A new Framework for Learning Classifier Models in Data Mining (A Case-Study in Network Intrusion Detection)[R]. Department of Computer Science University of Minnesota, USA, 2000
- [24] Elkan C. The Foundations of Cost-Sensitive Learning [C]// Proc. Int'l Joint Conf. Artificial Intelligence. 2001:973-978
- [25] Ting K M. An Instance-Weighting Method to Induce Cost-sensitive Trees[J]. IEEE Trans. Knowledge and Data Eng. , 2002, 14 (3): 659-665
- [26] Wei Fan, Stolfo S, Zhang Jun-xin. AdaCost: Misclassification Cost-sensitive Boosting[C]// Proceedings of the 16th International Conference on Machine Learning. 1999:97-105
- [27] Wu Jun-jie, Xiong Hui, Chen Jian. COG: local decomposition for rare class analysis[J]. DMKD, 2010, 20(2): 1384-5810

(上接第 292 页)

图 4 展示了动态分析界面。

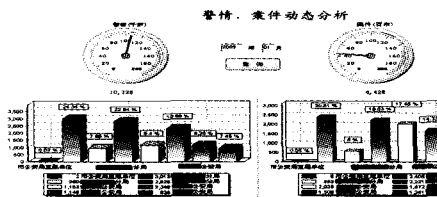


图 4 动态分析图

## 5.2 主题分析

主题分析主要是用来针对特殊的或者阶段性的分析需求而专门进行的分析作业。图 5 示出了案件主题的分析。

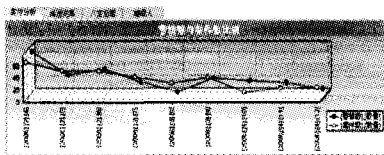


图 5 主题分析图

## 5.3 情报研判

情报研判管理主要针对捕获数据的汇总、分析。用户可结合信息查询、全文检索、BI、i2 等工具对“比对结果分析”做更深层次的分析,最终得出数据的研判评估。并将研判评估的结果发布到应用平台门户网站上。

另外,还提供了信息查询、统计报表等功能。

**结束语** 本文详细介绍了在公安行业建设与实施数据仓库系统的问题以及解决的方法,给出一个适合公安行业数据仓库的可行性解决方案。ETL 方案为数据仓库的建设与实施建立了良好的数据基础,保证了数据仓库中数据的正确性、完整性。合理地设计了系统的人、案件、物品、机构、地点等 5 大主题模型,决策层可以根据每个主题的分析结果来分析数据,掌握各个地区的各类案件的发案、立案、破案、涉案物品、涉案人员等 5 大主题的情况。

基于此数据仓库,实现了警务综合信息系统的数据共享、综合查询、统计分析、自定义报表、情报导侦与外部职能部门

的数据交换、决策支持分析等各种数据应用。通过对系统中各大主题分析结果进行相互关联分析,来更好地预防犯罪,提高破案率,提高“打”、“防”、“管”、“控”的工作效率。

本文相关研究成果已在实际应用中得到验证及完善,并在实际的应用过程中起到了较好的效果。2011 年某月份全省社会警情通报最新数据显示,杀人、绑架、强奸案件同比下降 2%、24.1%、7.8%,入室抢劫、抢汽车案件同比下降 24.8%、51.2%,放火、投放危险物质案件同比下降 8.7%、66.7%,涉枪案件同比下降 12.6%。

通过本课题的研究,得到一个适合公安行业数据仓库的可行性解决方案,并与“金盾工程”建设相结合,把公安情报信息资源作为新时期公安工作的重要战略资源。

## 参考文献

- [1] 王春雨,王延章,叶鑫,等.基于数据仓库的刑事案件决策支持系统设计[J].计算机工程与设计,2010,31(4):767-775
- [2] 麦永浩,杨超.公安数据仓库和数据挖掘应用研究[J].警察技术,2009(2):27-29
- [3] 尉宁.电信行业数据仓库建设与实践[D].重庆:重庆大学,2007
- [4] 徐珊.数据仓库在江西省公安厅综合信息分析系统中的应用[D].江西:南昌大学,2008
- [5] 杨兴凯.基于本体的政务数据仓库构建方法研究[J].计算机工程与设计,2010,31(7):1492-1499
- [6] 汪涛.医院数据仓库数据模型设计[J].计算机应用与软件,2010,5:191-194
- [7] Joy Mundy Warren Thornthwaite Ralph Kimball.数据仓库工具箱[M].北京:清华大学出版社,2007:1-265
- [8] Golfarelli M, Rizzi S.数据仓库设计——现代原理与方法[M].战晓苏,吴云浩,皮人杰,译.北京:清华大学出版社,2010:1-375
- [9] 英蒙. DW2.0 下一代数据仓库的架构[M].北京:机械工业出版社,2010:1-198
- [10] 周亮.电子政务决策支持系统中数据仓库的研究与设计[J].武汉理工大学学报,2005,27(1):31-35