

学校编码: 10384

分类号 _____ 密级 _____

学号: X2010230643

UDC _____

厦 门 大 学

工 程 硕 士 学 位 论 文

基于数据仓库的增值税监管预警系统的
设计与实现

Design and Implementation of Monitoring Management
Early Warning System for Value-added Tax Based on DW

廖秀琴

指导教师: 刘昆宏 副教授

专业名称: 软 件 工 程

论文提交日期: 2012 年 10 月

论文答辩日期: 2012 年 11 月

学位授予日期: 年 月

答辩委员会主席: _____

评 阅 人: _____

2012 年 9 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

随着税务部门信息化建设水平的不断提高,增值税电子化管理得到了不断的完善。但由于涉税信息的急剧增加,缺乏有效的集成,很难为管理层提供有效的决策分析支持。利用数据仓库技术可以对现有海量的数据资源进行加工、处理,发现其背后隐含的信息和规律,从而实现对增值税情况的科学分析和评估,为税收管理人员提供具体的工作指导。因此,构建满足基层高效管理、功能强大、安全可靠的增值税管理监控平台,具有十分重要的现实意义。

本文结合国税系统的实际需求,进行基于数据仓库的增值税监管预警系统的研究与实践。文章首先阐述了数据仓库的相关技术,随后进行了系统的总体设计,介绍了数据仓库的设计与实现策略。针对国税现有数据的存储情况和现有数据仓库系统的构建方法,给出了数据仓库的数据结构和性能优化方案,并详细介绍了增值税监控分析主题模型的设计和ETL的具体设计,完成了基于存储过程的ETL过程实现。为了消除系统的安全风险隐患,针对网络、操作系统、数据库三个方面做了详细的安全性设计,对系统面临的安全威胁进行了分析,提出了安全防护的具体措施,并做了有关安全性测试的分析。最后还展示了增值税监管预警系统实现结果,对系统运行测试结果做了相关分析,并对论文工作进行了总结,对下一步工作进行了展望。

该系统的研发,给税务管理人员提供了一个对增值税监控管理和分析服务的平台,对国税内部数据规划和业务决策起到了积极的作用,从而将税收管理提高到一个新的应用层次。

关键词: 数据仓库; ETL; 增值税

Abstract

With the rapid progress of the tax department information technology, the Value-added Tax (VAT) electronic management technique is also improved. However, due to the sharp increase in tax-related information along with the lack of effective integration, it is difficult to provide effective data support to top decision maker for analysis. Using data warehouse and OLAP techniques for machining and processing vast amounts of data resources is an effective solution to find the rules hidden behind information. With these techniques, scientific analysis and access of current situation of VAT can be achieved, so as to give proper instructions for the VAT management. It is very important to build an efficient, powerful, secure and reliable management and monitoring platform of VAT.

This thesis considers the actual needs of the national tax system, studies supervision and early warning system of VAT based on data warehouse. Firstly, it elaborates data warehouse technology, and then completes overall design of the system, and introduces data warehouse design and implementation strategy. For the current situation of data storage and construction method of data warehouse system in national tax system, it gives the data structure and suggestion of optimizing performance. It analyzes of the model design and ETL strategies for the theme of VAT monitoring and analyzing, and it also studies of the process of ETL based on the stored procedure implementation. In order to eliminate the potential risks of the system safety, it proposes detailed design schemes to keep the network, operating system and database in safe. It also includes specific measures to keep security along with security testing analysis. At last, this thesis shows and analyzes the test result of the VAT supervision early warning system, reviewed the work and prospected for further work.

The system introduced in this thesis can provide a services platform about the VAT monitoring and management for the tax administrators, and play a positive role in the national tax internal data planning and business decisions. In this way, it can improve tax administration to a new application-level.

Key Words: Data Warehouse(DW); ETL; Value-added Tax

厦门大学博硕士学位论文摘要库

目 录

第一章 绪论	1
1.1 研究背景与意义	1
1.2 研究目标和内容	3
1.3 论文内容的组织	4
第二章 相关技术综述	5
2.1 数据仓库	5
2.1.1 数据仓库概念与特性	5
2.1.2 数据仓库体系结构	6
2.1.3 数据仓库相关技术	7
2.2 ETL技术简介	10
2.2.1 ETL的相关概念	10
2.2.2 数据抽取	11
2.2.3 数据转换和清洗	11
2.2.4 数据装载	13
2.3 本章小结	13
第三章 系统需求分析	14
3.1 业务需求	14
3.2 系统要求	15
3.3 功能需求	16
3.4 非功能性需求	18
3.4.1 环境需求	18
3.4.2 系统性能需求	19
3.4.3 安全性需求	20
3.4.4 其他需求	21
3.5 本章小结	21
第四章 系统设计	22
4.1 系统架构设计	22
4.1.1 系统架构的目标和约束	22
4.1.2 系统总体架构	23
4.2 系统功能模块设计	25
4.3 数据仓库设计	28
4.3.1 数据仓库数据架构设计	28
4.3.2 主题数据的规划	30
4.3.3 数据仓库模型设计	32
4.3.3.1 概念模型设计	32
4.3.3.2 逻辑模型设计	33

4.3.3.3 物理模型设计	35
4.3.4 数据仓库性能优化设计	35
4.4 ETL设计	38
4.4.1 抽取	38
4.4.2 转换	39
4.4.3 清洗	40
4.4.4 加载	41
4.5 本章小结	42
第五章 系统安全设计	43
5.1 税务应用软件安全性技术要求	43
5.2 系统的安全性设计	45
5.2.1 网络安全	45
5.2.2 操作系统安全	46
5.2.3 数据库安全	48
5.3 系统的安全性测试设计	51
5.3.1 系统运行环境安全的测试	51
5.3.2 系统应用程序安全的测试	52
5.3.3 系统数据库安全的测试	53
5.4 本章小结	54
第六章 系统实现与测试	55
6.1 系统实现	55
6.1.1 技术路线	55
6.1.2 ETL功能实现	56
6.1.3 系统的主要用户界面与功能模块的实现	59
6.2 系统测试	64
6.2.1 系统功能测试	64
6.2.2 系统性能测试	66
6.2.3 系统安全性测试	67
6.2.4 评价准则	67
6.2.5 测试结果与分析	68
6.3 本章小结	69
第七章 总结与展望	70
7.1 总结	70
7.2 展望	70
参考文献	72
致 谢	74

Contents	
Chapter 1 Introduction.....	1
1.1 Background and Significances	1
1.2 Objectives and Contents	3
1.3 Chapters Arrangement.....	4
Chapter 2 Main Techniques	5
2.1 Data Warehouse.....	5
2.1.1 Concept and Characteristics	5
2.1.2 Architecture.....	6
2.1.3 Related Technologies.....	7
2.2 ETL Technology Introduction.....	10
2.2.1 Concept.....	10
2.2.2 Data Extraction.....	11
2.2.3 Data Transformation and Cleaning	11
2.2.4 Data Loading	13
2.3 Summary	13
Chapter 3 System Requirements Analysis.....	14
3.1 Business Requirements.....	14
3.2 System Requirements	15
3.3 Functional Requirements	16
3.4 Non-functional Requirements Analysis.....	18
3.4.1 Environmental Requirements	18
3.4.2 System Performance Requirements	19
3.4.3 Security Requirements Analysis.....	20
3.4.4 Other Requirements.....	21
3.5 Summary	21
Chapter 4 System Design.....	22
4.1 System Architecture Design	22
4.1.1 System Architecture of the Objectives and Constraints	22
4.1.2 System General Architecture.....	23
4.2 System Function Structure Design	25
4.3 Data Warehouse Design	28
4.3.1 Data Warehouse Architecture Design	28
4.3.2 Theme Data Selection.....	30
4.3.3 Data Warehouse Model Design	32
4.3.3.1 Conceptual Model Design	32
4.3.3.2 Logic Model Design.....	33

4.3.3.3 Physical Model Design.....	35
4.3.4 Data Warehouse Performance Optimization Design.....	35
4.4 ETL Design.....	38
4.4.1 Extracting	38
4.4.2 Transformation	39
4.4.3 Cleaning.....	40
4.4.4 Loading	41
4.5 Summary	42
Chapter 5 System Security Design	43
5.1 Tax Application Software Security Technical Requirements.....	43
5.2 System Security Design.....	45
5.2.1 Network Security	45
5.2.2 Operating System Security	46
5.2.3 Database Security.....	48
5.3 System Security Test Design.....	51
5.3.1 System Running Environment Security Test.....	51
5.3.2 System Application Security Test.....	52
5.3.3 System Database Security Test	53
5.4 Summary.....	54
Chapter 6 System Implementation and Test	55
6.1 System Implementation	55
6.1.1 Technical Route.....	55
6.1.2 ETL Function Implementation.....	56
6.1.3 System Major Interface and Function Modules Implementation	59
6.2 System Testing	64
6.2.1 System Function Test.....	64
6.2.2 System Performance Test	66
6.2.3 System Security Test.....	67
6.2.4 Assessment Criteria	67
6.2.5 Test Results and Analysis	68
6.3 Summary.....	69
Chapter 7 Conclusions and Prospect	70
7.1 Conclusions.....	70
7.2 Prospect.....	70
References	72
Acknowledgments	74

第一章 绪论

本章主要讨论增值税监管预报警系统的研究背景、研究意义，在分析研究基础上提出了对于这个系统的研究目标与内容，最后介绍本文的主要组织结构。

1.1 研究背景与意义

自1994年税制改革以来，增值税成为我国第一大税种，受到前所未有的重视，从近几年的《中国统计年鉴》分析来看，自2004年以来，增值税占全部税收收入的比重一直在1/3以上，但是由于主观与客观等多方面的原因，我国增值税税源管理还是粗放型的，仍存在许多亟待加强的薄弱环节，增值税流失依然严重。随着税收信息化程度的日益提高和税收管理员制度全面推行，增值税税源全面监控在税收管理员日常工作的重要作用已日益凸现。通过对税收管理员责任的细化和对增值税纳税评估工作的统筹安排，配合增值税新征管模式的建立，完善金税工程和中国税收征管信息系统CTAIS（China Taxation Administration Information System）系统，按照科学化、精细化管理的要求，促使增值税税源监控管理工作步入良性轨道，形成各司其职、齐抓共管、相互监督制约的工作氛围，实现对增值税一般纳税人“事前、事中、事后”的动态跟踪管理和全环节的监控，是当今国税系统的研究热点。

目前市级国税系统已经具备了稳定畅通的网络基础设施和良好的软硬件、数据环境。2003年CTAIS1.0在市级国税系统上线运行以来，市级国家税务局领导抓住市级数据大集中这一有利契机，逐渐重视信息资源开发利用，把数据资源利用开发放在信息化建设的核心地位，提出了建设适合市级局管理现状的数据应用系统，并提出从满足基层需求入手，逐步深入地将数据利用开发工作做到实处，实现国税数据的最大增值，满足市局、区县分局、科所、税收管理员四个不同层面数据利用的需求，进一步提高税收科学化、精细化管理水平。2003年以来，某市国税局就利用市局数据大集中的优势相继开发了全程阳光服务，定额管理、增值税免税、增值税预警、管理员小助手等非常有实际运用价值的系统，这些软件系统的推广应用大大提高了税收监控能力，提高了税收征管质量和效率，减轻了基层工作压力和负担。

2006年按照总局数据省级集中的要求，参照总局金税三期应用系统的建设标准，新的环境向国税应用系统提出了新的需求：

1、如何综合利用省级集中的数据指导基层工作，加强税源管理，成为当前加强税收征管工作的一个新课题。我们所要研究的增值税监管预警系统是建立在原来市局级集中数据资源基础上开发出来的增值税免税、增值税预警、管理员小助手应用系统的基础上，重新整合功能需求，设计新的开发平台，重新设计数据抽取程序，适合现在环境应用。

2、虽然基础业务系统产生了大量的业务明细事务数据，这其中包含着大量对预警有价值的信息，但由于它们的组织方式是面向业务应用的，而不是面向决策主题的，所以很难从中抽取有价值的信息。而且，这些数据以不同的数据结构，存放在不同的数据库中分布存放，备份文件格式相异。但随着OLAP (On-Line Analytical Processing, 联机分析处理)和数据仓库技术的不断成熟，通过建立基于数据挖掘的信息系统来充分利用已有数据，挖掘出其中的信息，为税务预警信息系统的发展和全面应用提供了理论技术基础，建立适合国税实际监控管理需要的税收数据仓库，在数据仓库平台上，利用联机分析处理、数据挖掘等技术，进行深层次的数据分析和预测，为实施精细化管理和领导决策提供平台和工具。

3、随着基于计算机和网络的各种税务管理信息系统规模的不断扩大和应用，数据容量急剧增长，安全问题日益严重。税务管理信息系统安全的根本目的是保护信息的正确性、完整性、机密性和可用性^[1]。如何确保系统的安全性是现在开发软件的关键步骤，开发一个安全的信息系统，必须从一开始就考虑安全性问题，并在开发的全过程中始终把系统的安全与系统的实施有机结合起来。

针对这几方面的情况，为了满足市局管理部门以及基层一线税收管理的迫切需要，达到税收风险预警和辅助管理相结合的目的，提出了基于数据仓库的增值税监管预警系统的开发，建立一个市、县、分局、税管员四级架构应用的增值税监控管理系统上，从用户体验的角度入手形成一套完整的用户交互框架，实现用户身份的统一认证和分层管理，进一步提供个性化服务和安全可控的交互平台，减轻基层税务人员操作各类软件的工作量，提高税收分析和纳税评估的质量和效率。

在增值税监管预警系统里面，税务部门通过根据总局、省局规定、结合本地实际建立的预警指标，纳税人申报情况和税负情况的关键数据，可按单位按时间按行业按类型按进项等多种方式进行全方位立体式查询，并进行统计比较

分析，为进行税源分析、税源管理和税收决策提供参考依据，预测企业出现偷税、骗税行为的几率，合理测定税务部门对不同企业的管理风险，确定对各类企业在不同阶段的管理重点，科学调配税收管理资源，必要时实施重点监控。另一方面该系统能够对事物（如税收收入）的规模、构成、分布、发展速度、平均水平、平衡程度等特征以及增长变化规律和发展趋势、事物之间（如 GDP 与税收收入）的相关关系、强度及平衡性等问题进行分析，并提供待办事宜提示、管理提示以此提高税务工作的管理效率，加强对纳税人经济活动的全面监督分析，达到税务风险预警和辅助领导决策的目的。因此，研究开发增值税监管预报警系统，符合实际发展需要，具有现实意义。

1.2 研究目标和内容

在国税系统建立安全可靠的数据仓库并开展数据分析已是当务之急。本课题目标是基于现有税务业务系统中积累的各类税务数据，在市级税务局建立综合数据仓库，实现从省局的自动抽取、集中、加载业务数据，完成对全市有关增值税数据的风险监控和税负分析，满足税收管理的需要，达到增值税风险预警和管理税源的目的。建设增值税监管预报警系统的首要 and 关键工作是税务数据仓库的建设。实施数据仓库是一项复杂的系统工程，将数据仓库产品应用到税收领域中，需要进行深入的应用研究。本课题以某市国税的具体需求为背景，研究数据仓库的关键技术，探索适合某市国税实际的数据仓库体系结构、设计方法、实施策略及应用技术，并分析了系统中数据信息的安全威胁，给出了系统的安全设计与测试，并提出了增强信息安全性的各项要求，在此基础上，完成增值税监管预警系统的设计与实现。

其主要研究内容如下：

- 1、分析和研究数据仓库的数据组织结构和 ETL (Extract-Transform-Load, 抽取-转换-加载) 的应用机制，结合国税应用的实际业务需求，设计与实现数据仓库模型，并做好相关 ETL 策略研究。

- 2、通过对增值税监管预警系统的总体业务需求分析，提出预警系统的总体设计思路，按照数据仓库理论对预警系统的框架、功能模块进行设计。

- 3、在以往系统设计中忽视了对系统安全内容的研究，增强系统的安全性是值得研究加强的地方，本系统需要在网络、操作系统、数据安全方面进行设计

和有关的安全性测试，搭建高安全、高可靠的税务应用软件。

4、按照增值税监管预警系统的设计要求，采用总局购买的 ORACLE 数据库系统（非数据仓库版本），把它策略成数据仓库，实现数据仓库的数据管理。技术人员通过编写存储过程的形式来实现从操作型数据库中抽取、净化、转换和加载数据到数据仓库，利用数据分析和数据展示代码开发与实现增值税监管预警系统软件。

总之，本文通过探讨数据仓库技术、系统安全技术与当前税收工作的结合，将先进的技术应用于实际工作中，对有关增值税方面的工作进行预测预报，提高税管员的办事效率，为各级税务部门提供执法的信息和决策的依据。本文提出了一整套的实现方案，在国内税务系统具有较高的先进性和实用性，从而有效地指导了税收工作，充分发挥税收这一经济杠杆的调节作用。

1.3 论文内容的组织

全文共分七章，具体结构安排如下：

第一章绪论，简要介绍项目背景、研究意义，课题的研究内容、论文内容的组织。

第二章相关技术综述，对实现系统需要用到的几个关键技术进行论述，包括数据仓库的相关技术，ETL 技术等。

第三章系统需求分析，主要论述系统的总体业务需求、建设目标、系统功能性、非功能性需求。

第四章系统架构与模块设计，基于数据仓库的增值税监管预警设计，研究系统设计要求和指标、总体架构及功能模块设计，数据集成、数据仓库平台的设计、ETL 设计。

第五章系统安全设计，简述税务应用软件安全性技术要求，并根据相关要求做好系统性能优化。研究网络、操作系统、数据安全方面的设计和安全性测试的设计。

第六章系统实现与测试分析，讲述了增值税监管预警系统的实现，涵盖开发技术路线、系统实现的主界面和部分代码、ETL 功能实现部分代码、软件测试内容及结果分析。

第七章总结与展望，本章对研究工作总结，并展望将来进一步的工作。

第二章 相关技术综述

本章讨论增值税监管预报警系统涉及的主要技术, 分别有数据仓库的相关技术和 ETL 技术。

2.1 数据仓库

2.1.1 数据仓库概念与特性

1、数据仓库的基本概念

所谓数据仓库是指一个面向主题的、集成的、不可更新的、且随时间不断变化的数据集合, 它用来支持管理人员的决策^[2]。通过人们多年对大量数据管理系统的研究, 在统计归纳多个部门信息的经验后, 基本上对数据仓库的特点达成共识: “数据仓库是一个面向主题的、集成的、相对稳定的、反映历史变化的数据集合, 用于支持决策分析与管理”。与数据库系统不同的是, 数据仓库是一种过程, 对各个业务系统的数据进行整合提取, 把有用的数据做进一步加工处理, 为事务数据库与高级终端查询提供衔接, 为不同的人群提供不同的决策信息。数据仓库技术则是通过建立数据仓库, 把各种动态的异构源数据抽取集成到一起, 进行清洗、转换等处理之后加载到数据仓库中, 并在一定策略下进行刷新, 为用户提供统一规范清晰的数据视图, 为数据决策分析提供一个有价值的数据库源。

数据仓库对业务系统在生产过程中生成的海量数据(包括大量的历史数据和操作型数据)进行加工集成和规范管理, 以便有效地为某个部门的决策应用提供高质量的数据源^[3]。数据仓库可以把不同类型、不同结构的数据集成, 对不同的数据来源进行统一数据结构和编码, 进而转换为统一的分析决策型数据环境, 为异构数据提供统一的数据视图。数据仓库能够实现对决策主题数据的存储和综合^[4], 它的特点在于它能把内部网络中孤立的信息岛上的各种管理业务数据集中起来, 存储在集成的数据库平台中, 利用现代各种相关技术对数据进行全方位分析, 为单位里的各个部门提供一个共享平台, 为相关决策提供有价值的信息。

数据仓库的主要作用是为用户提供一个支持多数据源的全局查询平台。其中主要包括: (1) 信息源中各种动态的数据是有规律变化, 或可以按某种趋势预测的, 如纳税申报数据是分时间段、分户产生的。(2) 可以应用历史的数据, 或延迟数据, 如纳税户档案管理、决策分析、数据挖掘等。(3) 应用要求查询性能较高, 而降低粒度。

2、数据仓库的特性

从Inmon关于数据仓库的定义中，从中发现数据仓库具有面向主题性、集成性、随时间变化和稳定性等一些重要的特性。

(1)数据仓库是面向主题的。数据仓库中的数据是从原来操作型数据库组织面向事务处理任务的数据转换成按照一定的主题域进行组织。主题按标准进行数据归类，根据一个宏观的分析领域可对应一个主题，在决策分析中重点规划的方面，一个主题也可以从多个操作型信息系统中组织。

(2)数据仓库是集成的。数据仓库的集成特性是指在数据进入数据仓库之前，必须经过数据加工和集成，这是建立数据仓库的关键步骤，首先要统一原始数据中的矛盾之处，把不规范的数据进行清洗，还要将原始数据结构做一个从面向应用向面向主题的转变。对存储不同地方的操作型数据，不同结构的数据源进行数据结构的统一和编码，利用集成工具将从原来的数据中抽取所需数据，进行加工处理之后才能进入数据仓库。

(3)数据仓库是稳定的。数据仓库包含大量的历史数据而不是日常事务处理产生的数据，数据经加工和集成进入数据仓库后是极少或根本不修改的^[5]。数据仓库涉及的操作主要是数据的查询，为某种决策分析提供数据来源，而这种分析往往是在大量的历史数据中寻找规律，因此不需要经常进行更新处理。数据仓库的操作除了进行查询以外，还可以按策略对数据进行定期的加载，把各种数据源发生的数据进行追加，通常不再修改追加后的数据，因此在数据仓库中可以通过使用预先计算、索引等方式。

(4)数据仓库是随时间而变化的。数据仓库内的数据有时限，如5年到10年之间，因此数据的键码包含时间项，标明数据的历史时期，这适合DSS进行时间趋势分析。而数据库只包含当前数据，即存取某一时间的正确的有效的数据^[6]。

2.1.2 数据仓库体系结构

数据仓库不是单纯的一种技术，而是一种体系结构。数据仓库从不同的数据源提取数据和信息，经过净化处理，把它们转换成统一规范的数据模型并可追加到数据仓库中。当用户对仓库提出查询请求时，由于数据矛盾、字段表达不一致等问题都已经解决，因此马上就可得到需要的信息，这使得决策查询更容易、更有

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库