

学校编码: 10384

分类号 _____ 密级 _____

学号: X2011231085

UDC _____

廈門大學

工 程 碩 士 學 位 論 文

基于数据仓库技术的税收分析系统的设计
与实现

Design and Implementation of Tax Analysis System Based on
Data Warehouse Technology

吴长燕

指导教师: 夏侯建兵副教授

专业名称: 软件工程

论文提交日期: 2013年5月

论文答辩日期: 2013年11月

学位授予日期: 2013年12月

指导教师: _____

答辩委员会主席: _____

2013年11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

伴随税务信息化建设工作的深入发展，税务机关积累了越来越多的数据信息，“科技管税”在税收征管中作用越来越明显。为进一步提高税收管理的科学化、精细化、规范化水平，如何充分利用一系列税收征管系统产生的海量数据进行税收分析，已经渐渐成为税务信息化工作的重点。如何突破传统的统计、查询和报表分析模式，需要在现有征管系统的基础上，利用数据仓库技术和丰富的数据资源进行税收数据深度分析，从而更好地为税收管理和决策支持服务，是当前税收分析工作必须考虑的问题。

为此，本文讨论建设税收分析系统，以便能够更好地为税务管理人员和中高层领导服务，提高其管理效率和决策质量。本文首先结合国内外税收数据分析现状，对税收分析系统的可行性进行分析；其次从数据仓库、联机分析处理（OLAP）及数据仓库的实现策略设计的基本概念和理论出发，阐述了典型的数据仓库系统的构建技术和构建方法；第三是从业务需求、用户需求、功能需求及非功能需求方面进行税收分析系统的功能和性能分析，提出了基于数据仓库技术的税收分析系统的总体设计思路和技术架构；最后重点阐述了税务日常查询、多维税务分析和税收宏观分析等系统分析模型关键点的详细设计，并通过一个具体的例子，将整个 OLAP 分析实现过程加以描述。

本文通过应用数据仓库技术，建设一个全省地税部门的税收分析系统，能够实现四个目标：一是提高税收分析的质量和效率，防范税收管理风险；二是及时发现税收征管中的异常情况，找出逾期未申报纳税人和偷税漏税纳税人，并对其跟踪；三是有效减少税收预测误差，合理制订税收计划，确保各项税款及时足额入库；四是提高税务管理整体水平，为领导决策提供科学依据。

关键词：数据仓库；OLAP；税收分析

Abstract

With further development of Taxation informationization construction work, the tax authority has accumulated more and more data information, "science and technology tube tax" in tax collection and management is more and more obvious.

In order to further improve the level of the scientific, refinement, standardization of revenue management, how to make full use of a series of tax collection and management system to produce huge amounts of data to tax analysis, has gradually become the focus of the tax informatization. How to break the pattern statistics, query and report the analysis of the traditional, need to be based on existing tax system, tax data depth analysis by using data warehouse technology and abundant data resources, so as to better support services for tax management and decision-making, is the current tax analysis must consider the issue.

Therefore, this paper discusses the construction of tax analysis system, in order to better for the tax management and leadership services, improve the management efficiency and decision quality. First, in light of the current situation of domestic and international tax data analysis, to analyze the feasibility of tax analysis system; Secondly, from the data warehouse, on-line analysis processing (OLAP) to achieve the basic concepts and theories of strategy design and data warehouse theory, elaborated the construction technology and construction method of data warehouse system typical of; Third, function and performance analysis of tax analysis system from the business requirements, user needs, functional requirements and non functional requirements, and puts forward the overall design ideas and technology architecture tax analysis system based on data warehouse; Finally, focuses on the detailed design model of key point of tax, tax analysis and daily query multi-dimensional macroscopic analysis of the tax revenue system, and through a specific example, the whole OLAP implementation process described.

In this paper, by applying the data warehouse technology, building a tax analysis system of the whole province land tax section, to achieve four objectives: One is to improve the quality and efficiency of tax analysis, prevent the tax management risk; The second is to find abnormal situation of tax collection and administration in time, find the overdue undeclared, taxpayers, taxpayers and tax evasion, and carry on the track; The three is to effectively reduce revenue prediction error, formulat reasonable

tax plan, to ensure tax Treasury in a timely manner; Four is to improve the overall level of the tax administration, Provide scientific basis for leadership decision-making.

Key Words: Data Warehouse; OLAP; Tax Analysis

厦门大学博硕士论文摘要库

目 录

第一章 绪论	1
1.1 选题背景及意义	1
1.1.1 选题背景.....	1
1.1.2 选题意义.....	2
1.2 国内外税收数据分析现状.....	3
1.2.1 国外发展现状.....	4
1.2.2 国内发展现状.....	4
1.3 本文的主要工作	5
1.4 本文章节结构安排.....	6
第二章 相关理论及技术简介	7
2.1 数据仓库	7
2.1.1 数据仓库的概念及特点.....	7
2.1.2 数据仓库体系架构.....	11
2.1.3 数据仓库的数据组织.....	13
2.1.4 ETL 工具	16
2.1.5 数据集市.....	18
2.1.6 维度.....	19
2.2 联机分析处理（OLAP）	19
2.2.1 OLAP 的概念和特点	19
2.2.2 OLAP 与 OLTP 的区别.....	21
2.2.3 OLAP 与数据仓库的关系	22
2.2.4 多维数据分析.....	23
2.3 数据仓库的实现策略设计	24
2.3.1 数据仓库的实现策略.....	25
2.3.2 数据仓库模型设计.....	27
2.4 本章小结	29
第三章 需求分析	30
3.1 税收分析系统的可行性分析.....	30

3.2 业务需求分析.....	32
3.3 用户需求分析.....	33
3.4 功能需求分析.....	35
3.5 非功能需求分析.....	30
3.6 本章小结.....	38
第四章 系统总体设计.....	39
4.1 系统体系结构.....	39
4.2 数据源.....	40
4.2.1 数据抽取.....	40
4.2.2 元数据管理.....	41
4.2.3 呈现和分析.....	42
4.3 应用框架.....	42
4.4 数据层次.....	44
4.4.1 操作数据.....	44
4.4.2 原子数据.....	45
4.4.3 汇总数据.....	45
4.4.4 历史数据.....	45
4.4.5 元数据.....	45
4.5 数据流程.....	45
4.5.1 数据迁移.....	46
4.5.2 数据汇总.....	47
4.5.3 数据分割.....	47
4.6 数据描述.....	48
4.7 多维数据立方体生成.....	49
4.8 本章小结.....	50
第五章 系统详细设计与实现.....	51
5.1 税务日常查询.....	51
5.2 多维税务分析.....	51
5.2.1 纳税人主题分析.....	51

5.2.2	文书主题分析.....	52
5.2.3	发票主题分析.....	53
5.2.4	票证主题分析.....	54
5.2.5	税款主题分析.....	54
5.3	税收宏观分析	61
5.3.1	税收负担分析.....	61
5.3.2	税收与消费分析.....	62
5.3.3	税收与 GDP 分析.....	63
5.3.4	税收预测分析.....	64
5.4	OLAP 实现.....	65
5.5	本章小结.....	71
第六章	总结与展望	72
6.1	总结.....	72
6.2	展望.....	72
	参考文献.....	74
	致 谢.....	76

Contents

Chapter 1 Introduction	1
1.1 Topic Background and Significance	1
1.1.1 Topic Background	1
1.1.2 Topic Significance.....	2
1.2 Domestic and Foreign Situation of Tax Data Analysis	3
1.2.1 Foreign Situation.....	4
1.2.2 Domestic Situation	4
1.3 Main Topics	5
1.4 Structure and Arrangement.....	6
Chapter 2 Introduction of Relevant Theory and Techniques... 7	7
2.1 Data Warehouse	7
2.1.1 Concept and Characteristics of Data Warehouse	7
2.1.2 Architecture of Data Warehouse	11
2.1.3 Data Organization of Data Warehouse	13
2.1.4 ETL	16
2.1.5 Data Mart	18
2.1.6 Dimension.....	19
2.2 OLAP	19
2.2.1 Concept and Characteristics of OLAP	19
2.2.2 Difference between OLAP and OLTP	21
2.2.3 Relation between OLAP and Data Warehouse	22
2.2.4 Multidimensional Data Analysis	23
2.3 Design and Implementation Strategy of Data Warehouse	24
2.3.1 Implementation Strategy of Data Warehouse.....	25
2.3.2 Model Design of Data Warehouse	27
2.4 Summary	29
Chapter 3 Demand Analysis	30
3.1 Feasibility Analysis of Tax Analysis System.....	30
3.2 Business of Demand Analysis	32
3.3 User of Demand Analysis	33
3.4 Function of Demand Analysis.....	35

3.5 Non-function of Demand Analysis	30
3.6 Summary	38
Chapter 4 Overall Design of System.....	39
4.1 System Architecture	39
4.2 Data Source	40
4.2.1 Data Extraction	40
4.2.2 Metadata Management	41
4.2.3 Presentation and Analysis	42
4.3 Application Frame work	42
4.4 Data Hierarchy	44
4.4.1 Operational Data	44
4.4.2 Atomic Data	45
4.4.3 Tabulate Data	45
4.4.4 Historial Data	45
4.4.5 Metadata.....	45
4.5 Data Process	45
4.5.1 Data Migration	46
4.5.2 Data Aggregation	47
4.5.3 Data Segmentation	47
4.6 Data Description	48
4.7 Multidimensional Data Cube.....	49
4.8 Summary	50
Chapter 5 System Detailed Design and Implementation.....	51
5.1 Daily Tax Query	51
5.2 Multidimensional Tax Analysis	51
5.2.1 Taxpayers Subject Analysis	51
5.2.2 Documents Subject Analysis.....	52
5.2.3 Invoices Subject Analysis	53
5.2.4 Tickets Subyeect Analysis	54
5.2.5 Tax Subject Analysis	54
5.3 Tax Macro Analysis	61
5.3.1 Tax Burden Analysis	61
5.3.2 Tax and Consumption Analysis.....	62

5.3.3	Tax and GDP Analysis	63
5.3.4	Tax Forecasting Analysis	64
5.4	OLAP Implementation.....	65
5.5	Summary	71
Chapter 6	Conclusions and Outlook	72
6.1	Conclusions	72
6.2	Outlook	72
References	74
Acknowledgements	76

厦门大学博硕士学位论文摘要

第一章 绪论

地税部门作为一个特殊的政府机构，现在正逐步向业务专业化、精细化、规范化、为纳税人提供优质高效服务的方向转变。新的环境需要新的需求，地税系统除了要加强自身管理外，还需要利用高新科学技术，以计算机、网络为辅助工具，实现全面的、高效的税务电子化，为纳税人提供更优质的、高效的、多渠道的服务，提高税务人员的工作效率，还要为业务骨干以及中高层领导提供准确的辅助决策信息。

1.1 选题背景及意义

自 1994 年税制改革，国地税分设以来，地税部门一直是国家和政府部门十分关注的重要部门，其信息化建设和发展一直得到了当地政府部门的高度重视。本结主要阐述了选题研究的背景和意义。

1.1.1 选题背景

自地税系统分设以来，系统内部高度重视“科技加管理”的重要意义和作用，为迎接新一轮征管改革的到来，云南省地税系统充分利用现有计算机技术和网络信息化技术，开发了一系列税收征管业务软件系统和多个子系统，不断完善税收征管手段，强化税源管理。

然而，在满足一系列税收征管业务需求后，如何充分利用系统产生的海量数据，为强化征管服务、为组织收入服务、为领导决策服务，成了摆在地税分析决策人员面前的一大难题，因此，分析决策人员对税务信息化提出了更高质量的要求。随着信息化建设的深入推进，地税部门的信息系统，已经不能适应当前税收数据分析和领导决策的需要，种类繁多、数据分散、数据集成度低的现象更为突出，不能成为服务企业和地方经济的有用的知识。只有充分发挥数据集中整合能力、数据分析和数据挖掘能力，才能更好地为地方经济服务，达到支撑地方财政的作用。

因此，需要在现有征管信息系统的基础上，采用数据仓库技术，并

充分利用多维分析工具，对大量的税收数据进行多角度的分析。而在目前的征收管理信息系统中，在最初规划设计时，就充分考虑到未来的应用和扩展，从而为数据仓库的建立提供了良好的基础。基于数据仓库的建立及其应用，特别是多维数据分析的应用，不仅可以满足当前信息消费的需求，而且也是未来全省地税建立安全和稳定的税收网络信息的基础，并具备功能齐全、协调高效、信息共享、严密监控等特点。

1.1.2 选题意义

随着计算机技术和网络信息技术的不断进步，地税系统的信息技术已经融入到税收工作的各个方面，税收征管业务系统所产生的数据，包括了大量有用的信息，为管理层提供决策服务。然而，由于这些数据组织是面向业务应用的，而不是面向决策主题的，因此从这些数据中很难获得更好地服务于税收决策的有用信息。随着税收征管业务系统的不断升级、扩大和改善，这些数据将会具备以下特点^[1]：

(1) 数据结构异构，海量数据。数据来自从不同的数据库，分布式存储、备份到各种存储器中，功能全面，综合的把握税收信息，这些数据成为了全面综合地把握税收信息的一个门槛。

(2) 复杂的业务系统，不统一的技术标准，不一致的指标口径，从而易容形成信息孤岛，信息难以共享。与此同时，各业务系统的数据又比较分散，给监控和管理带来很大难度，同时在基层单位上报数据资料时需要花去大量的工作时间。

(3) 基于在线事务处理（OLTP）技术的的查询系统，面对如此巨大的数据，在进行大量的计算、查询、分析和监控等方面，效率极其低下。如果想要从多个查询系统获得有价值的信息，需要花长时间去进行手工汇总加工，这将不可避免地降低了数据的有效性和准确性。

建设一个全省集中统一的税务数据如果对这些数据结构进行重组，并实现对这些数据的深度利用，不但能够满足目前的税务信息系统的特点和发展的需要，而且更加有利于税收决策分析。因此，在全省地税系统建立一个省局集中的、统一的、集成的数据仓库系统，是为了实现现有征管业务系统的数据信息

高度共享，让税收数据成为税收征管、税收计划、税收预测和税务决策的宝贵资源，也是提高管理者和各级领导决策水平的有效途径。

仓库系统，通过对各类分散的数据进行汇总、集成、统一规划和加工整理，及时发现税收征管中的异常情况和趋势。并充分利用可视化的数据，全面、多角度地分析税收问题，研究税收数据，挖掘有用的信息，更好地帮助税务人员指导税收征管活动，并最终实现以下目标：

(1) 建设全省地税集中统一的数据访问平台。为满足“科技管税”的需要，需要建立一个数据仓库系统，将不同系统不同规则业务数据和外部数据迁移出来，利用 ETL 工具对数据进行统一的数据抽取、转换、清洗和加载，再进行统一的集成、整合，最终建立起一个统一集成的、高效安全的、兼备扩展性好的税收分析系统，为所有地税系统人员提供准确可靠的决策支持数据，更好地为提高税收征管水平服务。

(2) 为全省地税系统提供决策支持智能分析。应用数据仓库技术和多维数据分析，对日常税收征管数据进行深度分析和挖掘，从而发现税收征管工作中存在的问题和异常情况，寻找税收发展的主要规律，并对变化情况进行趋势预测和预警判断，进而形成有价值的各种决策战略分析，为开展好下步工作作指导。它也可以帮助管理者轻松地跟踪指标变化，做到早发现问題，及时防范税务执法风险。

(3) 可以快速生成各种报表和进行各种分析查询。经过建立数据仓库系统，对税收数据进行充分整合和集成，可以提供灵活的动态报表生成和分析查询功能，并从多种角度对各种报表和查询结果进行深层次分析。

(4) 为准确制订税收计划、税收预测提供科学依据。为领导作出正确的税务决策提供科学依据，不断提高税收管理的科学化、精细化和专业化水平，并最终提升纳税服务质量水平。

1.2 国内外税收数据分析现状

随着计算机和网络技术的飞速发展，基本数据仓库的各种技术和应用开发技术也在不断进步，并在行政管理和企业生产中发挥着越来越重要的作用。伴随着传统税收业务系统累积的涉税数据越来越多，这些数

据如何被进一步的深度利用，成为了国内外税务部门及相关公司关注的焦点。数据仓库技术的使用是一种有效的方式来解决这个问题。

1.2.1 国外发展现状

在国外，数据仓库技术发展比较早，并在许多行业得到了广泛应用，先进的例子还很多，尤其在税务管理工作中并取得了明显成效。

比如，日本的计算机技术发展比较快，很早就在税务系统应用计算机信息技术，加强税务管理。日本的税务系统通过 Oracle 公司来帮助实施数据仓库项目，并在此基础上研究数据挖掘技术，对提高税收预测的准确性和税收征管的效率起到了非常明显的作用。自系统投入使用以来，日本政府得到的回报率也非常高，税收成本大大下降。

近年来，为更好地保证税务信息的真实性，美国税务机关又把数据挖掘技术应用到了新的税收征管软件中，申报率得到了很大提高，偷逃税率也明显下降了。同时，数据仓库和数据挖掘技术的应用，给美国税务机带来非常高的效益：德克萨斯州税务机关在实施数据仓库项目以来，每月的收入可达 600 万美元，年收益高达 8000 万美元^[2,3]。

1.2.2 国内发展现状

目前，国外的税收数据分析大多已经应用到数据挖掘技术，不再是简单的统计分析。随着信息技术产业的深入发展，国内税务系统也越来越高度重视税收数据分析工作。

在国内，税务机关和研究人员已纷纷投于税收数据分析领域研究，并发表了许多研究成果和著作。国家税务总局于 2006 年出版了两本著作：

《税源监控管理及数据应用分析》和《微观税收分析指标体系及方法》，这两部作品比较全面地从宏观到微观的角度介绍了数据分析模型、数据分析指标体系和数据分析方法，为各级税务机关进行数据分析，强化税源管理和监控，提供了理论支持和方法指导。

另外，一些专门从事税收数据分析的学者主要集中于信息技术应用领域的税收分析研究，并在税收数据分析的应用中首次探讨了数据仓库技术、多维数

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