

学校编码: 10384

分类号_____密级_____

学号: X2013232424

UDC_____

厦 门 大 学

工 程 硕 士 学 位 论 文

辽宁国税数据分析一体化平台的
设计与实现

Design and Implementation of Data Analysis Integrated
Platform for Liaoning Taxation

张晓明

指导教师姓名: 廖明宏 教授

专业名称: 工程硕士

论文提交日期: 2015年9月

论文答辩时间: 2015年10月

学位授予日期: 2015年12月

指导教师: _____

答辩委员会主席: _____

2015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

目前，随着税收信息化建设的不断深入，主要信息系统实现总局、省级集中，系统内部目前积累了大量、丰富的涉税业务数据，数据资源丰富，但是应用系统之间相互独立、分散运行的状况比较普遍。从功能上看，还缺乏对税收业务的分析、监控等数据利用方面的应用系统。因为税务系统软件平台的多样性，业务需求、功能设置、技术标准等还没有统一，数据质量问题一定程度上制约了查询、统计等基本的数据利用业务。数据分析一体化平台的开发就是在现有大集中的数据基础上，通过商务智能、数据仓库、数据抽取和挖掘的先进技术，整合为稳定、高效、安全、开放的一体化平台，最大限度的利用多样数据，查找税收工作中的漏洞和错误，寻找问题的根源，及时准确的解决问题，实现数据的集中分析、深入挖掘和信息共享。

本课题按照税务系统数据分析一体化平台的需求分析和设计要求，在国家税务总局税务信息化建设总体框架和标准规范的指导下，采用数据仓库的技术理念，整合国税内部各孤立应用系统中的异构信息，经过过滤、清洗、抽取加工和重新规划，形成基于统一技术框架，数据覆盖税收征管 Ctais 系统、防伪税控、稽核、进出口等业务系统的基础数据源，充分利用海量的业务数据，深入挖掘，以分析、预警、查询统计、报表等为应用主导，运用图表结合的丰富展现形式，既能辅助领导层和管理层进行宏观税收分析，也能服务于基层，统计业务数据，监控疑点纳税人。

论文从项目的开发背景和意义开始，介绍了数据分析一体化平台设计的背景以及利用价值，然后详细介绍了项目开发的过程，详细阐述了整个项目的需求分析，具体介绍了数据分析一体化平台的整体架构设计以及具体实现。最后，针对数据分析一体化平台进行了详细的测试和评价，并提出了改进和希望。

关键词： 税务信息； 数据分析； 一体化平台

Abstract

With the continuous development of the tax information system, the tax data started to be centrally managed and maintained at provincial level or even at national level. As a result, the current tax information system has gathered large amount of data which provides a good resource for government and research needs. However, the current application system lacks the ability to analyze and monitor the overall tax collection procedure due to the fact that the current application system still remains separated and decentralized and thus the business rules, system functions, and techniques vary case by case. Therefore, the data quality is problematic that has limited data mining and statistical analysis functions. The goal of this project is to develop a new tax data analysis platform that stores and manages the large amount of tax data for efficient search, analysis and extraction, identify the problem and weakness in the tax collection process, and ultimately provide insights for managing tax collection and facilitate decision making process.

In response to the needs for a new tax data analysis platform as stated above, and based on the information system design standards made by the national tax bureau, this project intends to present a new data warehouse model that centrally stores a variety of data sources that cover revenue management information system, anti-forgery tax control system, audition , and import/export activities by cleaning, filtering and combining the separate data sources in the current tax application systems. At application level, the new platform will provide a variety of enhanced functions such as search, statistical analysis, projection, and charts. The new platform is intended to facilitate the government and enterprise's decision making by providing statistics and monitoring fraud.

The thesis first introduced the background and the goal of designing a new tax data analysis platform. Next, the method and steps are addressed, focusing on the design and the implementation of the new platform. The final part of the thesis evaluates the new platform and describes the features to be enhanced in future.

Key words: Tax Information; Data Analysis; Integrated Platform

目 录

第一章 绪论	1
1.1 课题的研究背景及意义	1
1.2 国税系统信息化现状分析	2
1.3 本文的研究内容和目标	3
1.4 本文的组织结构	4
第二章 数据分析一体化平台的需求分析	5
2.1 系统概述	5
2.2 系统功能性需求分析	5
2.2.1 业务功能模块分析	5
2.2.2 系统管理流程	7
2.2.3 系统用例分析	8
2.3 系统非功能性需求分析	9
2.4 本章小结	11
第三章 数据分析一体化平台的设计	12
3.1 系统架构设计	12
3.1.1 平台总体架构	12
3.1.2 平台技术架构	12
3.1.3 平台体系架构	14
3.1.4 平台应用架构	15
3.1.5 平台网络架构	16
3.2 系统功能模块设计	17
3.2.1 总体功能模块设计	17
3.2.2 平台子系统的关系	24
3.2.3 外部接口	25
3.3 系统数据库设计	26
3.3.1 数据库逻辑设计	26
3.3.2 主要数据库结构设计	27
3.4 本章小结	32
第四章 数据分析一体化平台的实现	33
4.1 系统功能模块的实现	33
4.1.1 内网门户	33

4.1.2 综合数据平台	42
4.2 数据仓库设计	61
4.2.1 数据存储与管理	61
4.2.2 ETL 设计	63
4.2.3 主题数据设计	66
4.2.4 元数据设计	67
4.3 数据质量检查和出错处理	68
4.3.1 数据质量检查	68
4.3.2 出错处理设计	68
4.4 本章小结	69
第五章 数据分析一体化平台的测试	70
5.1 系统测试环境	70
5.1.1 硬件环境	70
5.1.2 软件环境	70
5.2 系统测试概述	71
5.3 功能测试	73
5.4 结果分析	74
5.5 本章小结	74
第六章 总结与展望	75
6.1 总结	75
6.2 展望	75
参考文献	77
致 谢	78

Contents

Chapter 1 Preface.....	1
1.1 Background and significance of the research.....	1
1.2 Present situation analysis of informationization	2
1.3 Content and purpose of this dissertation.....	3
1.4 The structure of this dissertation.....	4
Chapter 2 Analysis of system requirements	5
2.1 System overview	5
2.2 Analysis of system functional requirements	5
2.2.1 Analysis of business function module.....	5
2.2.2 System management process	7
2.2.3 Analysis of the system use case	8
2.3 Analysis of system non-functional requirements.....	9
2.4 Summary.....	11
Chapter 3 System overall design	12
3.1 The design of the system architecture	12
3.1.1 The overall application architecture.....	12
3.1.2 Technology implementation architecture.....	12
3.1.3 System architecture	14
3.1.4 Application architecture	15
3.1.5 Network architecture.....	16
3.2 The design of the system function module.....	17
3.2.1 The design of the system function module	17
3.2.2 The subsystem interactive functions.....	24
3.2.3 External interface	25
3.3 The design of the system database.....	26
3.3.1 Logical design of database.....	26
3.3.2 Structural design of database	27
3.4 Summary.....	32
Chapter 4 Implementation of the integrated platform	33
4.1 Implementation of the system function module	33
4.1.1 Intranet portal.....	33

4.1.2 Integrated data platform.....	42
4.2 The design of the data warehouse.....	61
4.2.1 Data storage and management	61
4.2.2 ETL design.....	63
4.2.3 Subject data design	66
4.2.4 Metadata design	67
4.3 Data quality checking and error handling.....	68
4.3.1 Data quality checking	68
4.3.2 Error handling design.....	68
4.4 Summary.....	69
Chapter 5 System testing and the running results.....	70
5.1 The environment of system test	70
5.5.1 Hardware environment	70
5.5.2 software environment.....	70
5.2 Summary of the system test	71
5.3 Functional test.....	73
5.4 Results analysis.....	74
5.5 Summary.....	74
Chapter 6 Summary and prospect	75
6.1 Conclusions of the dissertation	75
6.2 Future works	75
References.....	77
Acknowledgements	78

第一章 绪论

1.1 课题的研究背景及意义

计算机信息技术的迅猛发展，使得整个世界的政治、经济、科技乃至管理格局都焕然一新。随着网络和信息技术的日新月异的变化，税收信息化建设也取得了巨大的成果。税收信息系统和征管数据实现省级、总局大集中，大量分散的涉税数据资源集中在一起，为数据分析的宽度扩展和广度纵深提供了各种可能。近几年，税务系统开发了大量的税务应用软件，各应用系统之间相互独立、分散运行的状况比较普遍。从功能上看，还缺乏对税收业务的分析、监控等数据利用方面的应用系统。因为税务系统软件平台的多样性，业务需求、功能设置、技术标准等还没有统一，严重制约了查询、统计等基本的数据利用业务。对税收信息整理和分析，通过先进的信息技术手段进行深入挖掘，发现问题，寻找规律，并将结果应用在税收分析的各个领域，使散乱的数据经过整合后能更好的为税收领导决策所用，这才是目前我们信息化工作的重要内容^[1]。税收分析是税收工作的拓展，开发完善的分析管理信息系统，查找税收工作中的漏洞和错误，寻找问题的根源，及时准确的解决问题，切实加强税源管理和税款征收工作。

省局总局的数据大集中，为我们提供了数据整合挖掘的先决条件。一是收集了完整准确的税收数据，来源于各个分散独立的系统平台，对于数据分析的深层次拓展和挖掘奠定了基础；二是实现建立起统一的数据平台，实现了信息资源的共享；三是大数据一体化平台的建立和完善，对上下联动进行执法管理评价提供了可能^[1-2]。

现在我们面临的问题是：系统中存在海量的数据；现有的信息资源系统并不能有效的将分散在各个平台上的数据转化成有价值的战略信息。这是我们在信息高速发展的今天面临的信息危机。我们所说的信息危机并不是因为缺乏足够的信息数据，而是指这些信息没有经过筛选、过滤和重组并应用到战略决策的范畴，数据的散乱分布，对于领导决策分析和目标制定，作用甚微小^[2]。我们的数据库正迅速的膨胀起来，但是这些信息独立的分散在各地，我们的管理员和领导者除了简单的查询统计之外，并不能利用这些宝贵的数据资源进行分析和决策。大量的数据储存下来被广泛使用，并且急需将它们转换成有价值的信息资源。

数据分析一体化平台,是在现有大集中的数据基础上,通过商务智能、数据仓库、数据抽取和挖掘的先进技术,整合为稳定、高效、安全、开放的一体化平台,最大限度的利用多样数据,实现数据的集中处理和信息共享。

数据仓库^[5]是一个总的概念,面向主题的数据集合,在这里面涵盖了数据丰富多样的表现形式。作为一个子集,数据集市^[5]是一种数据存储形式,面向部门的特定主题设置,能够满足事务操作要求,可以完成数据分析的功能。这几个概念之间的分类是按照应用的不同而各有侧重,概念间的边界有重叠。

关系数据库之父 E. F. Codd 提出了联机分析处理(OLAP)的概念,基于管理者的角度将这一概念定位为提高信息浏览分析的能力。OLAP 系统架构主要分为ROLAP(Relational OLAP)、MOLAP(Multidimensional OLAP)、HOLAP(Hybrid OLAP)三种。

我们所说的数据挖掘是指从巨量散乱、模糊的、破碎的、独立异构的数据中,追寻数据之间的关联和规律,按照特定挖掘公式,将这些信息整合为完整有用的资源和知识^[4-5]。

1.2 国税系统信息化现状分析

中国税收信息化建设随着电子科技的迅猛发展,经历了一系列的历史变革。从最初完全手工的录入模式到手工电子相结合,由传统纸质转向现代网络,到金税三期“一个平台、两级处理、三个覆盖、四个系统”,如今,税务系统信息技术的利用得到了较大的拓展,信息资源已成为税收工作不可或缺的重要内容。

税收信息化是通过各种信息网络和数据库的计算机技术,结合征管管理业务流程的优化,实现整个税务系统计算机网络技术和税收业务的完美结合,将散乱独立的各个系统中的数据收集在一起,经过整合和过滤,利用数据仓库提取和分析,得到可以实现分析决策管理的有用资源,提升数据分析质量和效率,这是技术和管理两个创新联动的过程^[3]。数据的分析利用对于税收工作具有极大的实践意义,随着国税三大业务系统数据的集中,客观上具备了数据分析的基本条件,对于数据分析应用,是一个不断改进、不断提高的过程。

目前国税系统软件繁多,为满足各项业务需求,我们开发了很多应用分析软件,在收集信息、整理数据、软件应用中都存在很多问题。

一是税收数据不完整。从手工记录到现代化办公,税收信息化建设经历了一

系列变革，税收信息系统的开发也经历了从 DOS 系统到 Windows 系统软件、从单一简单到复杂完善的过程。早期由于技术和管理的限制，大多都是各地自行开发软件，开发平台和系统口径都是五花八门各自为政，而且软件的开发没有充分考虑到系统业务的多变性和拓展性，新政策新业务出现，就独立开发新系统，直接导致了大量的软件交叉分布在税务系统内部，基础数据散乱分布，由于没有完整统一的综合数据平台，无法进行深入分析挖掘。

二是系统中数据失真。税收系统中的数据涉及面广，纳税人基础信息、认定信息、申报征收信息、发票使用情况、稽核协查信息等等，信息量巨大，而且因为业务变化的原因纳税人信息变更频繁，系统中的状态与实际情况不能及时同步。同时，系统原因导致的错误数据，税收管理员误操作造成的数据错误和垃圾数据，因为不能及时发现和改正，这些错误数据越来越多，严重影响了数据的准确性和完整性^[2]。

三是数据利用率不高。不同的税收业务部门有自己的分析口径，包括不同的时间区域和项目细则，因此相同的征管指标，在系统中查询的结果却大不相同，无法实现横向和纵向的比对分析。要实现应用层次的拓展，还要从数据分析提高数据质量入手。

1.3 本文的研究内容和目标

本课题按照税务系统数据分析一体化平台的需求分析和设计要求，在国家税务总局税务信息化建设总体框架和标准规范的指导下，采用数据仓库的技术理念，整合国税内部各孤立应用系统中的异构信息，经过过滤、清洗、抽取加工和重新规划，形成基于统一技术框架，数据覆盖税收征管、稽核协查、进出口等业务系统的基础数据源，充分利用海量的业务数据，深入挖掘，以分析、预警、查询统计、报表等为应用主导，运用图表结合的丰富展现形式，既能辅助领导层和管理层进行宏观税收分析，也能服务于基层，统计业务数据，监控疑点纳税人。

数据分析一体化平台按照系统架构分为内网门户、综合数据平台两部分。系统按业务逻辑分成管理服务、金税发票、申报征收、税收会计、稽查法制五大部分，按功能分成查询统计、预警、分析监控、报表四大类。分析监控运用分析工具进行图形展现；报表根据业务需求提供的表样用电子形式展现；查询统计和预警采用先统计，后出清册的展现原则，清册粒度到纳税人。

大量的涉税数据经过数据分析一体化平台有效的整理合并，宏观上将税收数据同经济发展直接关联起来，又通过分析每一个独立纳税人的具体涉税行为，将视角微观化。

1.4 本文的组织结构

本论文共分为六章，论文首先分析了全省国税系统信息化的现状，阐明了课题的研究背景以及开发意义。

论文具体安排如下：

第一章 简单介绍了信息化孤岛现象导致数据分散无法综合利用的现状。针对辽宁省国税局数据利用分析的使用情况，提出了课题的背景、目标和研究价值。

第二章 概述数据分析一体化平台的系统需求分析。

第三章 详细介绍了数据分析一体化平台系统的总体设计。

第四章 介绍了数据分析一体化平台的实现。详细阐述了数据分析一体化平台的背景和总体目标。

第五章 介绍了数据分析一体化平台的系统测试和运行结果。

第六章 总结数据分析一体化平台系统研发所做的工作，指明了信息化建设未来的发展方向。

第二章 数据分析一体化平台的需求分析

2.1 系统概述

辽宁国税数据分析一体化平台，整合了辽宁国税内部各孤立应用系统中的异构信息，经过过滤、清洗、抽取加工和重新规划，形成具备统一数据口径和业务接口的信息数据源，为数据的增值应用提供了数据基础。以数据平台为基础，充分利用海量的业务数据，深入挖掘，以分析、预警、查询统计为应用主导、图表结合的丰富展现形式，既能辅助领导层和管理层进行宏观税收分析，也能服务于基层，统计业务数据，监控疑点纳税人。

2.2 系统功能性需求分析

2.2.1 业务功能模块分析

数据分析一体化系统主体共包括两部分：内网门户、综合数据平台。内网门户是数据分析一体化平台的入口，完成注册用户的统一认证和登录检验功能。单点登录后，列出所有可以应用的数据平台和业务系统链接，即可进入相应软件平台操作。图 2-1 详细阐述了数据分析一体化平台的业务功能结构。

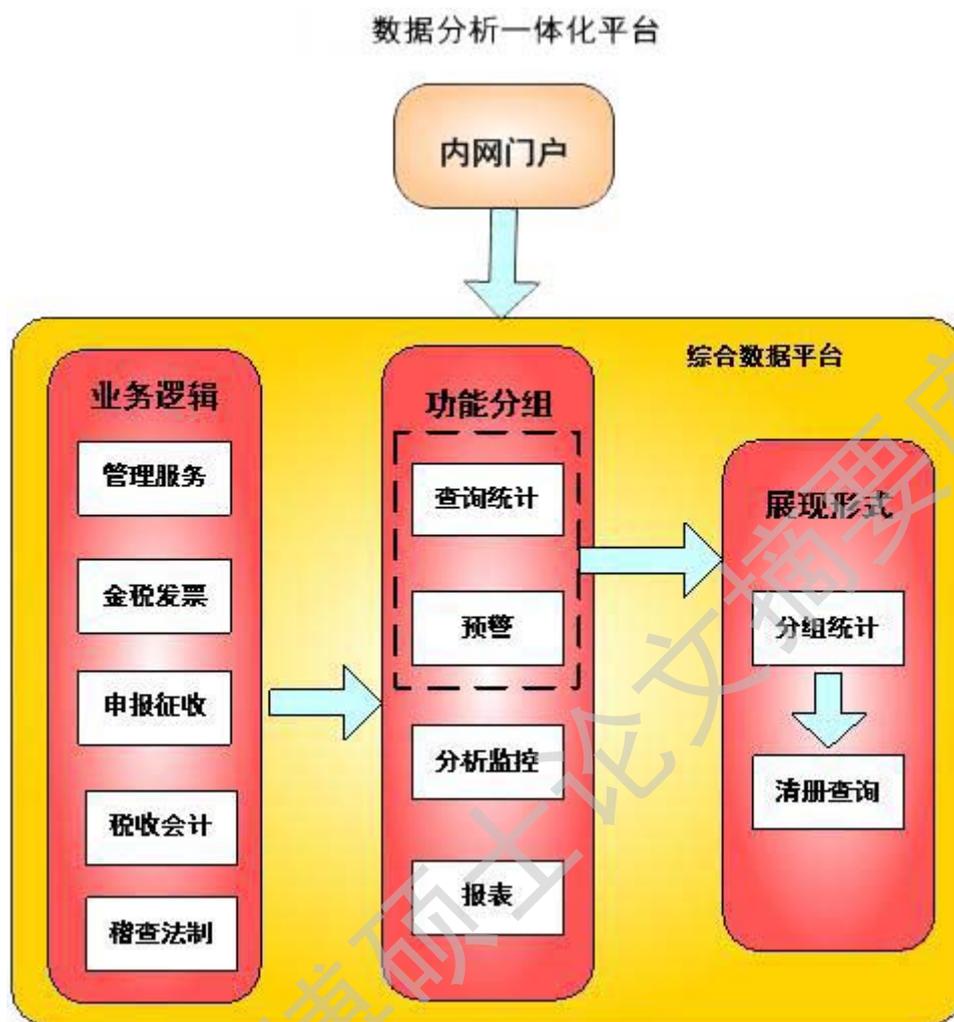


图 2-1 数据分析一体化平台功能结构图

综合数据平台是整个数据分析一体化系统的基础信息统计分析平台。按照业务需求，分成管理服务、申报征收、税收会计、金税发票、稽查法制五个功能模块。按应用结构，主要分为系统管理、查询统计、预警、分析和报表五大类，采用表单、饼图、对比图等丰富的电子表现形式，根据业务部门提出的指标口径生成监控分析和预警数据，为企业监控、税源分析、税收预测提供了准确详尽的数据分析资料。各部分具体概念如下：

查询统计：根据一定的条件，从一定的维度，如属期、税务机关、行业、登记注册类型等，把各项查询的指标进行分类汇总统计，同一查询功能的各维度可自由切换，并且支持钻取；从统计指标可以点击进去查询具体纳税人清册。

预警：根据一定的条件和预警阈值，筛选出超过预警阈值的企业，按照一定的类别（如属期、税务机关、行业等），形成统计；从统计指标可以点击进去查询具体纳税人清册。

分析：根据一定的税收规律和常用分析方法，分别从决策管理的宏观方向、操作使用的微观角度，对一定的税收指标进行归类统计，指导业务规律。

报表：按照格式固定、定期给上级税务机关上报数据的载体，原为纸质，在系统内体现为电子文档，主要包含总局报表和省局报表。

2.2.2 系统管理流程

从数据分析一体化平台的使用者来划分，大致可以分为决策层、管理层、操作层 3 个层面。对于决策层是指各级领导如省局局领导、处领导、各市局局领导等，这一层比较关注宏观的问题，包括趋势分析、同期比较、重点税源变化等情况；管理层包括决策层部分人员，还有各基层的干部，这一层面是处于决策层与操作层之间，比较关注某一方面的分析，而且内容非常具体时效性也要求较高；操作层则是指所有利用综合数据平台来解决实际问题的使用者，包括各个层面的人，但是操作层主要是完成某一项具体操作，会对细节数据比较关注，而且往往是反复操作，对系统的易用性、准确性要求较高。图 2-2 用直观的视图模式阐述了数据分析一体化平台不同的使用者关注的不同问题。

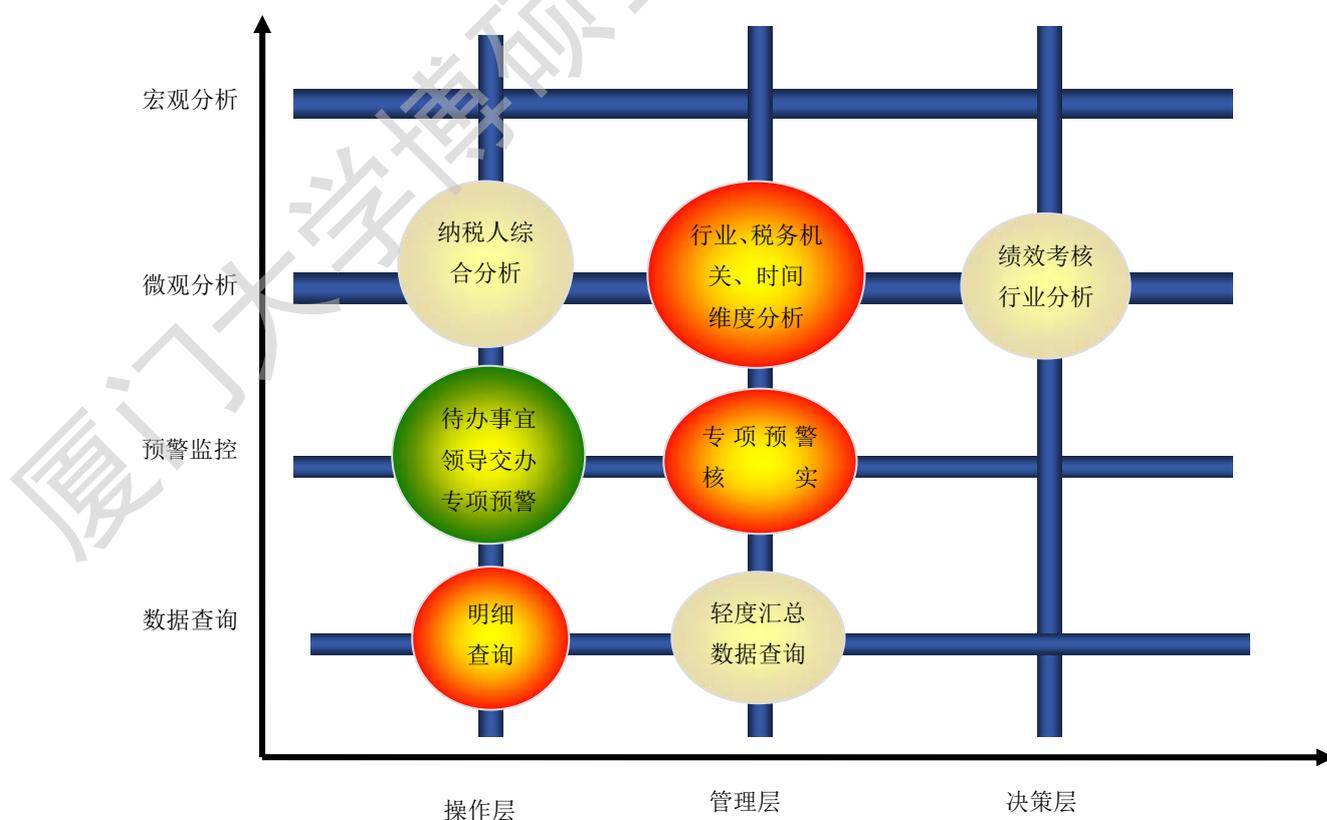


图 2-2 数据分析一体化平台使用者关注问题

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts 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.