

学校编码：10384

分类号_____密级_____

学号：X2012230978

UDC_____

廈門大學

工 程 碩 士 學 位 論 文

基于数据仓库的财政综合数据分析系统
设计与实现

Design and Implementation of Financial Comprehensive
Data Analysis System Based on Data Warehouse

赵重光

指导教师姓名： 廖明宏 教授

专业名称： 软件工程

论文提交日期： 2015 年 6 月

论文答辩日期： 2015 年 7 月

学位授予日期： 2015 年 9 月

指 导 教 师： _____

答 辩 委 员 会 主 席： _____

2015 年 6 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):

2015 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

近年来财政部门对一些挖掘技术的掌握运用以及数据仓库的构建都还有相当多的工作要去做。多数财政部门的数据分析仍然主要以报表等传统方式为主，相当于数据的初次利用，是一种“平面化”的分析。而数据挖掘的主要方式既包括报表统计，也包括数据的建模分析和潜在知识的挖掘，相当于数据的二次利用或多次利用，是一种“立方体式”的分析。

在大数据背景下，财政信息化应朝着以数据为核心的决策支持系统建设与应用的方向发展，着重开展数据的提取、整合分析、利用工作并将数据分析的范围从“已知”扩大到“未知”，从“过去”推向“未来”，既针对过去开展描述性分析以揭示规律，又面对未来开展预测性趋势分析，以发现趋势，从而实现从数据到信息、从信息到知识的转化，为更加科学的财政管理决策发挥其应有的价值。

本文根据财政发展的需要，运用财政多年积累的数据资源，研究如何将数据转换为知识的过程，研究用户在实际工作中对数据分析的需求，对环境资源的需求，研究我们如何借助工具实现用户的需求。

关键词：财政综合数据；分析系统；数据仓库；

Abstract

In recent years, the financial sectors have a lot of work to do, to master the use of data warehouse construction as well as some of the mining technology. Most financial sectors still mainly use Excel report or other traditional mode to analysis data, it's the initial use of data, and it's a "Planar" analysis. While data mining mainly includes both report statistics and data modeling analysis and potential knowledge mining, it's the two use of data or multiple use, is a "Cube" analysis.

In the background of Big Data, The development of the financial informationization should be towards the construction and application of decision support system based on data as the core , Focuses on the extraction of data, integration and analysis, the use of data analysis work will range from "known" expanded to "unknown", from "the past" to "the future", Both in the past to carry out descriptive analysis to reveal the rules, and facing the future development forecast of trend analysis, in order to find out the trend, so as to realize the information from the data, from the information to the knowledge, play its due role of financial management decision-making more scientific.

This paper according to the needs of the development of the finance uses the accumulation of years of financial data resources to study the process how to transfer the data to the knowledge, and study the user's needs for data analysis and resources in practice, study how to use tools to achieve user needs.

Keywords: Financial Comprehensive Data; Analysis System; Data Warehouse

目 录

| | |
|-------------------------|-----------|
| 第一章 绪论 | 1 |
| 1.1 论文研究背景和目的 | 1 |
| 1.2 研究现状与存在的问题 | 3 |
| 1.3 论文主要研究内容 | 4 |
| 1.4 论文结构安排 | 5 |
| 第二章 关键技术介绍 | 7 |
| 2.1 GOLDENGATE 介绍 | 7 |
| 2.2 COGNOS 介绍..... | 11 |
| 2.3 本章小结 | 14 |
| 第三章 系统需求分析 | 15 |
| 3.1 功能需求分析 | 15 |
| 3.1.1 常规分析 | 15 |
| 3.1.2 自定义分析 | 18 |
| 3.1.3 数据预测 | 20 |
| 3.1.4 孤立点分析 | 21 |
| 3.1.5 绩效评价 | 22 |
| 3.2 非功能性需求分析 | 22 |
| 3.2.1 性能指标要求 | 22 |
| 3.2.2 主机设备 | 23 |
| 3.2.3 数据仓库系统 | 23 |
| 3.2.4 OLAP 工具 | 23 |
| 3.2.5 报表及前端展现工具 | 24 |
| 3.2.6 ETL 工具 | 24 |

| | |
|----------------------------|-----------|
| 3.3 本章小结 | 25 |
| 第四章 系统设计 | 26 |
| 4.1 系统体系架构 | 26 |
| 4.1.1 逻辑视图 | 26 |
| 4.1.2 展现层 | 26 |
| 4.1.3 数据集成 | 27 |
| 4.1.4 业务数据（数据访问层） | 27 |
| 4.2 物理视图 | 27 |
| 4.2.1 数据源 | 28 |
| 4.2.2 财政数据仓库 | 28 |
| 4.2.3 可视化的展示工具 | 29 |
| 4.3 技术结构 | 29 |
| 4.3.1 技术介绍 | 29 |
| 4.3.2 技术规范 | 29 |
| 4.4 系统设计 | 29 |
| 4.4.1 系统总体结构 | 29 |
| 4.4.2 模块设计 | 30 |
| 4.4.3 数据库设计 | 33 |
| 4.5 对象模型及描述 | 34 |
| 4.5.1 系统领域模型 | 34 |
| 4.5.2 活动图 | 35 |
| 4.5.3 序列图 | 36 |
| 4.6 本章小结 | 36 |
| 第五章 系统详细设计与实现 | 37 |
| 5.1 模块设计 | 37 |

| | |
|------------------------|-----------|
| 5.1.1 用例设计 | 37 |
| 5.1.2 高级查询 | 38 |
| 5.1.3 模板查询 | 41 |
| 5.1.4 模板配置 | 47 |
| 5.2 数据建模设计 | 52 |
| 5.2.1 逻辑数据模型 | 52 |
| 5.2.2 维度列表 | 53 |
| 5.2.3 业务系统与维度关系 | 54 |
| 5.2.4 维表设计 | 55 |
| 5.2.5 实时表设计 | 59 |
| 5.3 组件接口设计 | 62 |
| 5.4 系统运行效果 | 63 |
| 5.5 本章小结 | 66 |
| 第六章 系统测试 | 67 |
| 6.1 测试规划 | 67 |
| 6.2 测试内容 | 67 |
| 6.3 测试结果 | 73 |
| 6.4 本章小结 | 74 |
| 第七章 总结与展望 | 75 |
| 7.1 总结 | 75 |
| 7.2 展望 | 75 |
| 参考文献 | 76 |
| 致 谢 | 77 |

Contents

| | |
|---|-----------|
| Chapter 1 Introduction..... | 1 |
| 1.1 Research background and objectives of the paper..... | 1 |
| 1.2 Research status and existing problems..... | 3 |
| 1.3 Main research contents of the paper..... | 4 |
| 1.4 Introduction to the paper structure..... | 5 |
| Chapter 2 Introduction to Key Technology..... | 7 |
| 2.1 GOLDENGATE introduction..... | 7 |
| 2.2 GOGNOS introduction..... | 11 |
| 2.3 Summary of this chapter..... | 14 |
| Chapter3 Analysis of System Requirements..... | 15 |
| 3.1 Analysis of functional requirements..... | 15 |
| 3.1.1 Conventional analysis..... | 15 |
| 3.1.2 Custom analysis..... | 18 |
| 3.1.3 Data prediction..... | 20 |
| 3.1.4 Outlier analysis..... | 21 |
| 3.1.5 Performance evaluation..... | 22 |
| 3.2 Analysis of non-functional requirements..... | 22 |
| 3.2.1 Requirements of performance indicators..... | 22 |
| 3.2.2 The host device..... | 23 |
| 3.2.3 Data warehouse system..... | 23 |
| 3.2.4 OLAP tools..... | 24 |
| 3.2.5 Report forms and front display tools..... | 24 |
| 3.2.6 ETL tools..... | 24 |
| 3.3 Summary of this chapter..... | 25 |
| Chapter 4 System Design..... | 26 |
| 4.1 System architecture..... | 26 |

| | |
|---|-----------|
| 4.1.1 Logical view..... | 26 |
| 4.1.2 Presentation layer (interface interaction)..... | 26 |
| 4.1.3 Data integration (logical control layer)..... | 27 |
| 4.1.4 Business data (data access layer)..... | 27 |
| 4.2 Physical view..... | 27 |
| 4.2.1 Data sources..... | 28 |
| 4.2.2 Financial data warehouse..... | 28 |
| 4.2.3 Visual presentation tools..... | 29 |
| 4.3 Technical structure..... | 29 |
| 4.3.1 Technology introduction..... | 29 |
| 4.3.2 Technical specifications..... | 29 |
| 4.4 System design..... | 29 |
| 4.4.1 The overall system structure..... | 29 |
| 4.4.2 Module design..... | 30 |
| 4.4.3 Database design..... | 33 |
| 4.5 Object model and description..... | 34 |
| 4.5.1 System domain model..... | 34 |
| 4.5.2 Activity diagram..... | 35 |
| 4.5.3 Sequence diagram..... | 36 |
| 4.6 Summary of this chapter..... | 36 |
| Chapter 5 Detailed Design and Realization..... | 37 |
| 5.1 Module design..... | 37 |
| 5.1.1 Case design..... | 37 |
| 5.1.2 Advanced query..... | 38 |
| 5.1.3 Template query..... | 41 |
| 5.1.4 Template configuration..... | 47 |
| 5.2 Data modeling design..... | 52 |

| | |
|---|-----------|
| 5.2.1 Logical data model..... | 52 |
| 5.2.2 Dimension list..... | 53 |
| 5.2.3 Relationship between business system and dimension..... | 54 |
| 5.2.4 Dimension table design..... | 55 |
| 5.2.5 Real-time table design..... | 59 |
| 5.3 Component interface design..... | 62 |
| 5.4 System operation effect..... | 63 |
| 5.5 Summary of this chapter..... | 66 |
| Chapter 6 System Testing..... | 67 |
| 6.1 Test plan..... | 67 |
| 6.2 Test content..... | 67 |
| 6.3 Test results..... | 73 |
| 6.4 Summary of this chapter..... | 74 |
| Chapter 7 Summary and Outlook..... | 75 |
| 7.1 Summary..... | 75 |
| 7.2 Outlook..... | 75 |
| References..... | 76 |
| Acknowledgements..... | 77 |

第一章 绪论

1.1 论文研究背景和目的

随着财政改革的深入，财政信息建设的不断完善，尤其是完成财政应用一体化建设后，彻底改变了数据管理的方式，财政数据由分散式管理向集中式管理转变。从财政工作实际情况来看，要进一步提高财政科学化精细化管理，就是要及时监控到财政各项资金的具体执行情况，每一笔收支的详细信息，例如谁支付款、哪个银行账户付款、哪个开户银行、哪个银行账号、支付金额多少、支付用途是什么、预算科目是什么、收款人银行、收款人是谁、收款人银行账号、支付方式什么、结算方式什么、结算时间等都在掌握之中。同时，随着“金财工程”的整体推进，财政部门掌握的数据也越来越明细，规模也越来越大。财政拥有近 20 年的决算数据，近 10 年的部门预算及预算执行数据，10 年非税收入数据，10 年的工资、人口供养，资产等业务数据和大量的管理数据。

充分利用和挖掘这些数据，为财政决策服务，提升财政管理和决策水平，如果能将这些底层数据蕴含的丰富信息和知识挖掘出来，将能够将财政部门手中的数据资源成为不可多得的信息“金砖”。我们将把大数据分析引入我们到工作中，利用数据挖掘手段找寻管理对象的关联性，随即进行更为深入细致的研究，并预知其规律、预估政策效果。通过财政综合查询分析系统实施，可以对财政的政策制定及业务工作提供多方面的技术支持，从而实现：

1. 健全公共财政管理体系。科学的预算执行制度是公共财政管理体系的重要组成部分。从 2002 年推行国库集中支付改革以来，我国新型预算执行管理体系逐步建立、现代国库管理制度不断完善，但在实际操作层面仍存在一些问题，影响着现代财政管理体系制度的完整性和稳固性。借助数据仓库与数据挖掘技术，发掘财政资金的运行和管理行为等规律，解决预算执行管理中量化衡量标准不清的问题，为健全公共财政管理体系提供新的切入点。

2. 提高财政决策水平。国库改革在改变预算执行管理的方式的同时，确定了新模式预算执行数据的生成规则，依托国库集中支付体系和财政大平台系统，各

级预算单位的每一笔收支数据，都可利用系统自动生成，且实时将数据传输到各级财政部门，各级财政收支情况信息的及时性、准确性和真实性有了保障。对于各级财政部门大量的收支数据进行研究分析、挖掘其内在规律性，将大大提高各级财政决策支持水平。

3. 完善财政信息化管理。开展财政数据挖掘，财政部门需要建立一个可以有效收集、监测和分析所获得大量数据的系统。这些年来按照财政金财工程规划，逐步建设财政一体化系统，财政数据丰富且数据庞大。这些业务系统的数据管理也存在信息冗余、效率低下、口径不一致等问题。采用数据仓库和数据挖掘技术，使用异常侦测可以进一步发现数据错误，使用挖掘主题研究可以进一步发现并补充欠缺的但十分有益的信息，删除失去指导意义的信息等，使财政数据更加及时有效。

4. 构建新疆预算执行数据动态统计分析中心，实现从交换中心和全疆一体化系统中实时提取数据，分析“原汁原味”的生产数据，全面详实的反映各级财政收支和单位资产情况、负债情况、人员编制情况等信息，方便各级财政部门相关处室统计分析，为领导决策提供数据支持。

5. 能够促进财政基础业务分类体系建设。财政管理的过程实际上是一个循环周期，从上半年的预算编制到第二年的预算审计可以被看作是一个最基本的财政管理周期。一般来说，新一年度的预算编制授权执行工作可以很好地通过这样一个完整的财政管理周期作为参照。财政信息化就是通过运用现代技术手段统筹管理财政管理周期的各个环节，最终实现“规范化、科学化、信息化、”管理。基础业务信息科学化分类管理是财政业务管理与财政信息化间的桥梁，是进行数据仓库建设和挖掘的根基。用信息化的视角来看，对财政基础业务信息进行分类，需要实现的是信息技术与科学合理的财政业务管理模式的深度融合，信息技术可以实现细之又细的财政基础业务信息的记录和处理，从而实现更广泛而又深入的相关业务统计、分析和挖掘，基础业务数据的可塑性更强、应用范围更加宽泛。所以说，不能简单地来看基础业务信息分类工作，它伴随着财政管理水平的一个有效提升而来的，使财政管理周期中各个环节复杂的业务基础数据有效的成为一个

有生命力的整体。

财政信息化建设逐步深入、财政数据信息日趋规范，深入挖掘分析财政数据，使更好组织和利用好财政信息的必要选择；随着财政改革的不断深化，数据挖掘也朝着精细化的方向推进，从而进一步为财政部门发现财政收支内在规律、揭示有关问题背后的关键诱因、提升管理水平提供依据。

研究如何从近年来财政信息化建设中沉淀的大量、低密度、低价值的数据来源中提取出，其数据内在有价值的信息和知识，为相关处室及各级领导提供财政运行过程中宏观经济预测、财政收支分析、财政资金运行监督监控、项目建设绩效评价等方面的辅助，将给财政的精细化、科学化管理奠定坚实的基础。

数据仓库是一个面向主题的、集成的、时变的和非易失的数据集合，用以支持管理部门的决策制定过程”。数据仓库的特点决定了它是实现分布式数据整合和信息利用的最佳解决方案^[1]。目前我们所使用的业务数据库主要关系型数据库，处理日常事务，是为具体业务日常事务处理设计的，在存储结构、数据表设计、事务处理和使用上与数据仓库不相同，不能满足海量数据处理和复杂查询及数据分析的需要（例如 oracle 数据库在大数据量下执行复杂查询所涉及的多表关联、表的累加、数据项排序、大量数据检索等操作将影响业务系统的运行效率）。为了解决大数据量处理问题，搭建数据仓库系统势在必行，依托数据仓库研发财政综合数据分析系统，是充分挖掘和发现各级财政海量数据资源中重要知识的有效手段。

1.2 研究现状与存在的问题

历经各级财政多年的数据积累，当前各级财政部门已经沉淀了大量的信息资源，这就是常说的“大数据”，这些资源无疑是各级财政亟待开发的“宝藏”。据财政部国库司统计^[2]，截至目前，中央财政每年的支付数据达到了 70 多万笔，每年非税收入数据达到几千多万笔，每年税收收入数据更是达到了 2 亿笔（目前暂未接入数据库）地方财政部门也积累了大量的财政数据，以新疆本级为例，每年的财政支付数据达到了 20 多万笔，每年包括税收在内的财政收入数据达到 70

多万笔。如果把地域范围扩展到全国，把时间范围上溯到 2001 年改革之初，把管理范围扩大到工资决算、采购、国债、执行分析等信息系统，这个数字将非常惊人。大数据时代的来临对财政部门而言都是一次难得的历史机遇和挑战，财政部门只能迎头赶上，用科学的数据挖掘方法深挖各类信息，为中国转型发展提供不竭动力。

就具体部门而言，我国财政部门对数据挖掘的认识还不够深刻。以中央财政为例，只有少数司局在运用宏观计量经济模型来进行宏观经济运行分析并对财政收支进行预测分析。财政数据最集中的国库司也是近年来才开始运用数据挖掘，而且还处于起步阶段，因此，对一些挖掘技术的掌握运用以及数据仓库的构建都还有相当多得工作要去做。多数财政部门的数据分析仍然主要以报表等传统的工具为主，相当于数据的初次利用，是一种“平面化”的分析^[3]。而数据挖掘的主要方式既包括报表统计，也包括数据建模分析和潜在的知识挖掘，相当于数据的二次利用或多次利用，是一种“立方式”的分析。

1.3 论文主要研究内容

这些年根据财政改革和财政发展需要，财政以金财工程建设为主线，各级财政信息化所取得了显著成效，财政信息系统建设与应用基本涵盖业务管理的各个方面，应用中积累了大量财政数据和管理基础信息，在支撑财政改革和提升财政管理水平等方面都发挥了重大作用。大多数应用系统都是以业务流程为核心，以处理各项具体业务为目的的运营式离散事务系统，数据在其中的作用只是一个个管理事项的单纯记录，各系统的数据采用不同的编码方式和组织结构，彼此之间相互独立处于老死不相往来的状态，其应用也仅限于查询。

以往的查询系统也称为报表系统，只是简单的数据的汇总和信息的罗列，没有对数据进行深入的分析，并未挖掘数据其内在的规律、相互关系、相互依赖等重要信息。综合查询分析系统给用户更大的发挥空间，让用户借助于自己的专业知识，结合系统提供的专业分析手段，辅助用户挖掘出超出想象的结果。

综合查询分析需要提供五大分析手段：统计预测分析、常规统计分析、绩效评价分

析、财务自动分析报告和孤立点分析。为了提高分析的性能问题，综合查询分析需要建立在数据仓库系统基础上，分析系统本身不存储数据，仅提供各种分析、预算算法和分析结果的展示。

在大数据背景下，财政信息化应朝着以数据为核心的决策支持系统建设与应用的方向发展，着重开展数据的提取、整合分析、利用工作并将数据分析的范围从“已知”扩大到“未知，从“过去”推向“未来”，既针对过去开展描述性分析以揭示规律，又面对未来开展预测性趋势以发现趋势，从而实现从数据到信息、从信息到知识的转化，为更加科学的财政管理决策发挥其应有的价值。

1.4 论文结构安排

本文将从项目的研究背景及目的、使用关键技术、系统建设需求、系统设计思路、研发、测试、总结等七部分，分别介绍整个项目的研究过程。

第一章绪论。这章主要阐述财政综合数据分析系统项目研究背景及意义，分析各级财政目前对数据资源的使用情况，分析存在的问题。

第二章关键技术介绍。在这章重点介绍了为了满足数据收集过程中，在不影响原系统性能的情况下，保证数据的实时性，所采用的 GoldenGate 工具。介绍了解决展示灵活性、分析多样化，所使用的 Cognos 工具。

第三章系统需求分析。这章重点介绍了用户对数据查询分析的要求，分别介绍了常规分析、自定义分析、数据预测、孤立点分析、绩效评价等功能的需求。最后对介绍了项目对硬件资源和环境的需求。

第四章系统设计。本章是论文的重点内容之一，阐述了系统的整个总体设计过程，重点介绍了财政综合查询分析系统的体系整体架构设计、数据系统安全、数据模型设计等。

第五章详细设计与实现。本章是在上章对系统整体设计的基础上，通过体系架构设计、系统安全、模型设计，详细描述了数据架构设计、模型设计、ETL 算法设计与实现、应用逻辑层设计与实现。

第六章系统测试。本章介绍了系统测试的方法和过程。

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.