

学校编码: 10384
学号: X2014230004

分类号 _____ 密级 _____
UDC _____

厦 门 大 学

工 程 硕 士 学 位 论 文

基于数据仓库的证券公司业务报表系统的
设计与实现

**Design and Implementation of Business Reporting System in
Securities Company Base on Data Warehouse Technology**

黄 晔

指导教师姓名: 王 备 战 教 授

专 业 名 称: 软 件 工 程

论文提交日期: 2017 年 10 月

论文答辩日期: 2017 年 11 月

学位授予日期: 2017 年 12 月

指 导 教 师:

答 辩 委 员 会 主 席:

2017 年 10 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。)

本人声明该学位论文不存在剽窃、抄袭等学术不端行为,并愿意承担因学术不端行为所带来的一切后果和法律责任。

声明人 (签名):

指导教师 (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

证券公司一般都有着众多的 IT 系统和繁杂的信息来源，于此相对的，证券公司对于数据统计、数据分析的依赖成都非常高。传统的数据报表面对日新月异的数据需求显得力不从心，拖累了整体业务的发展。为了适应这种发展的需求，解决传统报表面对新需求时展现的种种弊端，需要设计一套报表系统作为解决方案。本文试图通过一种基于数据仓库模式的解决方案来解决这个问题。

首先，将原先的“数据——报表”的模式细化为“源数据——整合数据——指标——报表”的模式。通过数据仓库 ETL 工具作为调度工具，将数据从各个证券公司不同时期、不同背景，面对不同的应用、不同的开发商等各种客观前提下建立的各个业务系统中整合到数据中心里面，形成了报表系统的数据基础。

接着，引入“业务逻辑层”，增加了“指标”的概念，将报表需求转变为对指标的选取，将报表需求颗粒化，可度量化。指标之间存在上下级关系，在 ETL 程序调度的时候根据指标之间的计算优先级完成各层指标的计算和保存。

然后，业务报表的需求将从原先业务部门的字面描述，转变为业务部门对既有指标的选择和对新指标的统计需求。指标与指标之间的核对因为业务逻辑层的存在也变得很方便，这就减少了以往报表需求中最麻烦的由于不同报表之间勾稽导致的返工。提高了工作效率和需求响应速度。

在 ETL 的整体调控下整个数据报表系统可定期从各类系统中抽取数据，并根据不同的维度形成各自的统计指标，形成了以日期为形式的不同时间段的报表，这充分展现了数据仓库“围绕主题、集成数据、数据时变性及非易失性”的特点。

关键词：数据仓库；报表系统；ETL

Abstract

The securities industry has IT systems, more complex information source, high degree of dependence on data statistics analysis of the characteristics. The traditional collection of statements, statistical way of these drawbacks, great drag on the development of the securities business of the company as a whole. In order to meet the needs of this development, solve the problems in the use of statistics report mode, need to design a set of reporting system solutions. This article attempts through the model based on data warehouse is a kind of solution to provide a solution for this problem.

First of all, refine the original data report mode to the source data integration model data index statements. Through data warehouse ETL tools as a scheduling tool, the data from each securities firm in the different times background, in the face of different application developers, such as various kinds of objective premise to establish various business system integration to the data center, formed the reporting system data base.

Then, introduced the business logic layer, increase the concept of index, transforming the reporting requirements for the selection of index, the report needs a granular, parent-child relationships between quantitative indicators can be degrees, at the time of the ETL process scheduling according to the index calculation priority to complete each layer between the index of calculation and save.

Then, the needs of the business report will from the business department of the literal description, into a business for both the choice of indicators and statistics on the targets of the new requirements. Index and index between the check for the existence of the business logic layer becomes very convenient also, reducing the previous reporting needs the most trouble caused by the different report checked the rework Improve the working efficiency and demand response speed.

Under the control of the entire report system in ETL scheduler, regularly collect data from various systems, and according to the different dimensions of statistical

indicators, again the report according to the date form each time Fully reflects the data warehouse oriented thematic data integration of data when degeneration, nonvolatile characteristics of data.

Key Words: Data Warehouse; Reporting System; ETL

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	1
1.2 研究现状	2
1.3 论文研究意义与内容	3
1.4 论文结构	3
第二章 相关技术介绍	5
2.1 数据仓库技术	5
2.1.1 数据仓库的特征.....	5
2.1.2 数据仓库的数据组织.....	8
2.2 数据抽取、转换与加载	9
2.3 本章小结	13
第三章 系统需求分析	14
3.1 系统建设背景和目标	14
3.1.1 建设背景.....	14
3.1.2 建设目标.....	16
3.2 系统特征	16
3.3 功能性需求	17
3.3.1 需求概述.....	17
3.3.2 指标设计与计算.....	19
3.3.3 报表设计与数据计算.....	22
3.4 非功能性需求	26
3.5 本章小结	27
第四章 系统设计	28
4.1 系统总体架构	28
4.2 数据模型设计	28
4.2.1 调度层数据模型.....	29

4.2.2 业务逻辑层数据模型.....	36
4.2.3 报表逻辑层数据模型.....	38
4.3 本章小结	40
第五章 系统实现	41
5.1 报表需求管理模块	41
5.2 指标管理模块	43
5.3 任务调度模块	44
5.4 报表数据生成	46
5.5 报表数据展现	47
5.6 本章小结	49
第六章 总结与展望	50
6.1 总结	50
6.2 展望	50
参考文献	51
致谢.....	53

Contents

Chapter 1 Introduction.....	1
1.1 Project Background.....	1
1.2 Research Status.....	2
1.3 Significance and Contents.....	3
1.4 Outline of Thesis	3
Chapter 2 Related Technologies.....	5
2.1 Data Warehouse Technology	5
2.1.1 Features of Warehouse	5
2.1.2 Organizations of Data Warehouse.....	8
2.2 Data Extracting, Transforming and Loading.....	9
2.3 Summary.....	13
Chapter 3 System Analysis	14
3.1 Background and Target	14
3.1.1 Background.....	14
3.1.2 Target	16
3.2 System Characteristics	16
3.3 Functional Requirement Analysis	17
3.3.1 Summary of Requirement.....	17
3.3.2 Index Design and Calculation.....	19
3.3.3 Report Design and Data Calculation.....	22
3.4 Non-Functional Requirement Analysis.....	26
3.5 Summary.....	27
Chapter 4 System Design	28
4.1 General Architecture of System.....	28
4.2 Data Model Design	28
4.2.1 Scheduling Layer Data Model	29
4.2.2 Business Logic Layer Data Model.....	36

4.2.3 The Reporting Logic Layer Data Model.....	38
4.3 Summary.....	40
Chapter 5 System Implementation.....	41
5.1 Report Requirements Management Module	41
5.2 Index Requirements Management Module	43
5.3 Task Scheduling Module.....	44
5.4 Report Data Generation	46
5.5 Report Data Display.....	47
5.6 Summary.....	49
Chapter 6 Conclusions and Prospect	50
6.1 Conclusions.....	50
6.2 Prospect	50
References	51
Acknowledgements	53

第一章 绪论

1.1 研究背景

近些年中国证券业的急速发展意味着券商的经营规模及需要应不断扩大,此外券商的经营模式亦面临前所未有的改革变化,目前已从单一营业部的管理模式逐渐过渡到整个券商公司级的管理模式。有学者用“圈地”战略转变成“耕地”战略的形象说明来解释目前中国证券经济业务的变化形式。随着此行业竞争日趋激烈,随着佣金价格涨浪潮而来的是各证券公司的服务战,在此阶段,满足客户需求,引导及适应客户个性化意愿,脱离同质化竞争,增加公司的盈利幅度已然成为各证券公司的主要战略方向,而要在激烈的同行竞争中脱颖而出,优化及改革业务模式是核心,努力提高管理水平的同时提升公司竞争力^[1-3]。

在打造适合面向公司级别的数据分析中心各类必要条件中最关键的环节即证券业务集中交易系统的稳定运行,准确、稳定的分析型数据中心的成功建设对券商的竞争力提升的核心环节。目前国内券商将信息系统视为单一的、独立的成本中心,虽然此类观点具有一定的直观性,但其忽略了客户、信息系统、管理部门、客户服务等环节的统一管理,因此亦失去了上述价值链带来的盈利空间。在西方发达国家的证券市场里,准确的、权威的数据信息研究报告往往源于证券公司,计算机信息处理人员的数量甚至达到公司总人数的 1/3,而我国证券公司核心交易系统架构远远落后于西方发达国家,仍处于十余年前的水准,提供的仅仅是基本信息、咨询服务而已,毫无品牌意识及核心竞争力。

通过大量的资料^[4-6]研究我们发现目前国外券商管理措施及业务程序均源自高科技支持运行,其广泛运用各类完善的管理系统,因此当金融市场低迷或公司发展遇到瓶颈是,券商可借助现金的技术增强管理体系,以此更进一步为证券业务流程提供更强有力的支持,以提升公司竞争力。

大量的历史数据在近十多年来被中国证券市场所积累发展,而此类数据含有大量的有用信息,目前中国证券面临的主要问题是,如何从大量的历史数据提取有价值的信息并成功利用。虽然历史数据可不断积累,但对其的开发利用是有一定限度的,目前国内不少券商信息库里存在大量数据,但却对数据的提取束手无策,

故无法客观地、充分地了解自身经营情况，从而影响如何加强管理和提高服务质量决策的制定及实施。

目前中国证券市场的内部经营模式逐渐发生变化，从卖方占主导地位逐渐转变为买方市场，从业务角度着手，我们发现在日趋激烈的国内券商竞争中经纪业务是竞争围绕的重点，因为经纪业务是券商利润的主要环节。粗放经营是我国证券行业的主要经营模式，仅需为客户提供交易渠道，并未对客户需求进行研究、了解，因此对客户的真正的需求并不知情。自从 80 年代中国证监会对佣金比例放开管理，各大券商才真正意识到了解客户的重要性，因为客户是企业生存、发展的基础，透彻了解客户的真正需求是所有券商商业活动的重要基础。如若券商对客户需求不了解，则容易在日趋激烈的竞争中流失客户。数据仓库可集中地、实时地处理大量的历史数据，并将上述数据转化为有价值的知识信息，对券商快速发展发挥重要作用，故我们认为应用数据仓库技术对有价值性的客户进行挖掘、拓展是可行的，同时还可满足有限资源的合理配置和利润最大化，增加券商的行业竞争力。

1.2 研究现状

目前我国金融行业持续发展，不论是业务规模还是业务种类都一直在扩展，证券公司的经济业务部门在日常工作中需要处理的数据问题也越来越多，从单纯的数据详表、数据清单到数据统计表、数据分析表，对于业务报表的需求也随之越来越大^[7,8]。从最基本的检查匹配特殊业务场景的数据，到根据汇总统计表确认经营状况、考核业绩指标，甚至于基于分析数据对未来发展做决策都离不开报表系统的支撑。在业务和需求都高速发展的背景下，相应的报表系统却在大部分时候都落后于现状，目前证券公司在制作报表过程中存在各种各样的问题，比如报表分散，数据源不统一，结果勾稽困难等，而解决的方案通常是大量的人力投入去解决，造成效率低下，数据准确性不足等等衍生的问题。

本论文的目的在于探讨一种解决方案，能够尽可能通过 IT 系统来解决以上的问题。

1.3 论文研究意义与内容

证券业数据包括股票历史数据、股票的基本面数据、营业部的交易结算数据、营业部的客户信息数据、企业内部财务数据、上市公司信息、宏观经济社会信息等等。随着证券行业的发展迅速，证券公司的业务部门必须有更加快速有效准确的方式来获取，整合这些数据，从中提取中可用于决策的信息。这使得报表需求是一个持续性增长的需求，业务部门对它的要求也早已跨过了“有无”的层次，越来越多的要求多维度，多系统数据整合的报表，对报表需求的响应速度也有了越来越高的要求。为了适应这种发展的需求，解决目前使用的报表统计模式的种种问题，需要设计一套报表系统的解决方案，它应该能够实现以下目标：

1. 能够从多个不同的系统获取数据，集成证券公司的相关数据。能够对各个营业部交易数据、行情数据、其他业务系统数据进行定期的抽取，把异构系统的数据经过抽取转换，集中到总部的系统中，为后续分析提供全局、一致的数据基础。
2. 业务系统的变动对报表系统的影响小，避免出现业务系统一有变动，报表系统就需要跟着变动的情况。
3. 减小报表统计口径歧义，能够有效的减小沟通成本。
4. 简单的业务需求能够快速形成报表。

1.4 论文结构

第一章，绪论。通过对当前国内证券业数据应用情况，以及目前国内主要的报表解决方案的介绍，阐述了本文的选题原因和意义。同时也对论文的结构进行了说明。

第二章，相关技术介绍。介绍数据仓库的技术特征以及数据组织特征，并且对与数据仓库密切相关的数据抽取、转换与加载技术进行了说明。为后续系统设计奠定技术基础。

第三章，系统需求分析。分析了系统建设的目标和系统特征，并且中功能性需求和非功能性需求两方面对系统需求进行了分析。

第四章，系统设计。从系统的总体架构和数据模型两个层面对系统的实现方

案进行设计。

第五章，系统实现。按照系统的设计的模块，逐一完成对系统整体架构的实现。

第六章，总结和展望。总结论文各章节的内容，对于系统中未完善部分提出未来展望。

厦门大学博硕士论文摘要库

第二章 相关技术介绍

2.1 数据仓库技术

“数据仓库”这一技术源自 20 世纪 80 年代, W.H.Inmon 在其著作《Building the Data Warehouse》首次提出数据仓库这一概念, 并被后人誉为“数据仓库之父”。他在书中针对“一方面拥有大量数据, 另一方面有用信息却很缺乏”这一现象提出了一套综合整合的解决方案, 而这个综合解决方案即数据仓库技术, W.H.Inmon 在书中写到: “20 世纪 90 年代数据仓库将会被视为信息技术领域的新焦点, 在该技术中集成化和历史化数据将得到充分体现, 将不同种类的应用系统通过一定的、合理的手段进行集成, 数据仓库对数据从不同事务发展或历史角度对数据进行重新组织及存储, 并在不同的应用系统中进行集合, 以进一步实现数据信息化及处理。”因此 W.H.Inmon 认为数据仓库的定义应是: 其属于一个管理层的决策过程, 此过程乃数据的集合过程, 具有主体性、集成性、非易失性, 并且随着时间变化而变化^[9-11]。

W.H.Inmon 利用简短有全面的字眼描述了数据仓库的主要特征, 主体性、集成性、非易失性以及及时变性将实现数据仓库与其他存储系统之间的差别。

2.1.1 数据仓库的特征

根据 W.H.Inmon 对于数据仓库理念的阐述, 我们在此对数据仓库那四类重要特征进行详细阐述:

1. 面向主体性(subject-oriented)

主体这一说法目前仍被视为抽象的概念, 在数据仓库领域的主题指的是将企业信息系统中的数据在一个较高的层次进行不断地综合、归类以及分析, 在逻辑意义层面上认为主题应该是针对企业中某一宏观分析领域所囊括的对象。综上所述, 面向主体的数据组织方式即站在一个较高的高层次水平对对象数据进行完整的、一致的刻画, 并由此可完整、统一、准确的描述企业的各项数据以及各个数据环节彼此间的联系, 而其中所谓的高层次水平则是一个相对的概念, 其相对的

对象是数据仓库所面向的数据组织方式而言，数据仓库的主体性使其具有更高的数据抽象级别。

在数据仓库的数据组织中面向主体性可谓其核心原则，在此仓库中的所有数据均围绕某一主体进行开展，此时的主体即是主题，而数据仓库的使用者多数为企业的管理者，是抽象的、较高层次分析数据对象的主要使用者。譬如：企业的产品、产品生产的供应商、客户均可视为主体，从信息管理角度认为，面向主体的数据组织方式即站在一个较高的高层次水平管理对象数据，对分析对象数据进行分析、归纳、分类；从数据组织角度认为，主体就是对上述数据进行完整、统一、准确的描述，其中包括数据本身以及数据间彼此的关联。

因此可以这么认为，数据仓库的创立是围绕主体形成的，我们在创立数据仓库前应准确了解决策分析，以此来设定主体，从中抽取主题，所抽取的主体具体包括哪些分析对象，分析对象应如何组织，均必须根据决策分析进行，正如，企业销售管理中的主体指的就是产品与客户两者。

2. 数据集成性(integrated)

数据集成性亦是数据仓库一个重要特征，指的是使用集成手段将分散于不同领域的源数据进行抽取、整合、筛选等。

数据仓库所集成的数据无法像业务部门处理数据那般简单获取，而是从业务处理发生直接联系的业务处理系统那里获取。而所获取的业务数据与业务处理是紧密联系的，数据服务于日常业务处理，并不服务于决策分析。从业务系统中提取的源数据无法直接加载至数据仓库中，在此之前应进行一系列的预处理，正如我们前面所阐述，及对源数据进行抽取、整合、筛选等集成。换句话说，从业务系统获得的源数据应根据一定标准进行统一，如命名、单位等，将不统一的现象消除掉。

3. 数据的时变性(time-variant)

数据的时变性指的是数据仓库中的数据并非一成不变，而是随着时间的推移而逐渐发生变化。尽管上数据仓库集成性中我们提到数据仓库的数据无法实时反映业务处理的情况，但仓库里的数据亦是不断变化的，设想一下，如果我们将10年前数据仓库的数据进行分析，制定10年后的决策，那产生十分严重的后果。由此我们必须间断性的捕捉数据仓库内不断变化的源数据，将产生变化的数据追

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.

厦门大学博硕士学位论文摘要库