

学校编码: 10384

分类号 \_\_\_\_\_ 密级 \_\_\_\_\_

学号: X2009230174

UDC \_\_\_\_\_

厦 门 大 学

硕 士 学 位 论 文

基于 ODS 的银行统一数据管理平台的研究与设计

Research and Design of the Bank Unified Data Management

Platform Based on ODS

陈梦冰

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: 2011 年 10 月

论文答辩时间: 2011 年 11 月

学位授予日期: 2011 年 月

答辩委员会主席: \_\_\_\_\_

评 阅 人: \_\_\_\_\_

2011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。)

声明人(签名): 陈梦冰

2011年11月30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）：陈梦冰

2011年11月30日

## 摘要

在金融业竞争日益激烈的趋势下，带来服务产品的多样化。同时业务的发展和变化使得各类信息数据量急剧增加，给数据的传输和存储带来了许多新的问题，同时也对企业的管理信息和数据服务建设提出了新的要求。银行业务系统繁多，信息数据随着业务系统的独立呈现数据孤岛的状况，由于没有一个统一的、高效的数据管理平台的支撑，各应用系统的数据获取方式大部分还是采取独立从源系统获取的方式，这很容易造成众多的竖井式数据架构出现。此外，随着业务的发展，外部监管机构、银行管理层和业务部门对决策信息的依赖程度在不断提高，经常会有一些突发的数据统计和分析需求。这些决策类的信息需求一般是跨多个业务系统，若仅通过手工方式来实现，是难以满足这些需求的时效性要求。

本课题的研究是基于以上现有金融行业业务系统的弊端，结合具体银行项目实施案例分析其信息系统需求，在 ODS 的设计原理基础上建立银行统一数据管理平台，建设全行统一的数据视图，解决银行的各种数据问题。针对银行信息系统的现状问题，阐述统一数据管理平台建设的必要性。详细描述了基于 ODS 设计理念的统一数据管理平台的总体架构和数据层次模型划分，以及各个数据层次对应的数据存储周期和历史数据中心的数据保存策略。研究 ETL 体系架构设计，设计与数据模型相结合的标准化 ETL 数据处理流程。介绍了元数据管理的功能和数据质量检核机制，使整个统一数据管理平台管理更具高效性和严谨性。最后探讨统一数据平台的建设对银行综合报表应用的集成和优化作用。

**关键词：**ODS； ETL； 数据管理

## Abstract

In the trend of increasing competition of the financial industry, development and changes of business make all kinds of information data increased dramatically, and bring up new problems about the data transmission and storage. It also brings up new requirements for the enterprise's management information and data services. There are many banking systems, and information data presents as islands of data because of the separateness of the business systems. As the absence of a unified, efficient data management platform, most data acquisition methods of each system are the way that accesses data from the source system independently. In addition, as business development, external regulators, bank management and business department improve the dependence on the decision-making information, and always bring up some unexpected data statistics and analysis needs. The decision-making information generally exist multiple business systems. If achieving the information manually, it is difficult to meet these timeliness requests of requirements.

Based on the above drawbacks and combined with specific case of Bank project, this dissertation would study of this topic analyzes the requirements of the information system, design the banking unified data management platform based on the design principle of the ODS, and solve the bank's various data issues. This dissertation analyzes the status of the banking information system, and brings up the necessity for the construction of a unified data management platform. This dissertation details the overall architecture and the division of data hierarchical model of the unified data management platform based on the design principle of the ODS. It also details data storage period and data retention policy of historical data of each data level. Researched of ETL architecture design, and bring up standard data processing combined with data model. Also described the function of metadata management and mechanism of data quality validation, and make the unified data management platform more efficient and rigor. Finally, the effects of the integration and optimization of Bank Comprehensive Report by unified data platform are discussed.

**Keywords:** ODS; ETL; Data management

## 目 录

<b>第一章 绪论</b> .....	<b>1</b>
1.1 研究背景及选题意义 .....	1
1.2 主要研究内容 .....	3
1.3 本文的结构安排 .....	3
<b>第二章 系统分析与总体设计</b> .....	<b>4</b>
2.1 系统的主要目标 .....	4
2.2 系统需求及现状 .....	4
2.2.1 系统需求概述 .....	4
2.2.2 系统现状分析 .....	5
2.3 建议的实施方案 .....	6
2.4 系统可行性分析 .....	8
2.4.1 项目风险分析 .....	8
2.4.2 技术可行性分析 .....	9
2.4.3 经济可行性分析 .....	9
2.5 总体架构设计 .....	10
2.5.1 逻辑架构设计 .....	11
2.5.2 数据架构设计 .....	13
2.5.3 技术架构设计 .....	13
2.5.4 物理架构设计 .....	13
2.6 小结 .....	15
<b>第三章 数据模型设计</b> .....	<b>16</b>
3.1 模型设计目标 .....	16
3.2 模型设计规范 .....	16
3.3 模型架构设计 .....	17
3.3.1 源数据增量层 .....	18
3.3.2 标准化增量数据层 .....	19
3.3.3 基础数据层 .....	19
3.3.4 共性加工数据层 .....	20
3.4 数据存储策略 .....	20
3.4.1 数据存储周期 .....	20
3.4.2 历史数据存储 .....	21
3.5 小结 .....	22
<b>第四章 ETL 体系设计</b> .....	<b>23</b>
4.1 ETL 整体架构 .....	23
4.2 ETL 数据处理 .....	24
4.2.1 传统 ETL 处理 .....	24
4.2.2 标准化的 ETL 处理 .....	24

4.3 数据质量统计.....	26
4.4 异常情况处理.....	27
4.5 ETL 调度监控.....	27
4.6 ETL 接口设计.....	30
4.6.1 与外围源系统的接口.....	30
4.6.2 与通用报表管理平台的接口.....	31
4.7 数据卸载.....	31
4.7.1 卸数原则.....	31
4.7.2 ETL 处理流程.....	31
4.8 小结.....	35
<b>第五章 元数据管理及数据质量控制.....</b>	<b>36</b>
5.1 元数据分类.....	36
5.2 元数据管理.....	36
5.2.1 元数据管理平台功能.....	36
5.2.2 数据字典.....	37
5.2.3 权限管理.....	38
5.2.4 指标服务.....	39
5.2.5 其他派生功能.....	39
5.3 元数据分析.....	39
5.3.1 数据剖析.....	39
5.3.2 数据血统.....	40
5.4 数据质量检核.....	40
5.4.1 业务数据质量检查.....	40
5.4.2 元数据质量检查.....	41
5.4.3 数据质量检核框架.....	41
5.4.4 数据质量跟踪.....	43
5.4.5 数据质量量化分析.....	44
5.4.6 数据质量反馈机制.....	44
5.4.7 数据质量报告.....	45
5.4.8 脏数据处理.....	46
5.5 小结.....	47
<b>第六章 综合报表应用的集成和优化.....</b>	<b>48</b>
6.1 报表应用现状.....	48
6.2 报表应用集成目标.....	49
6.3 综合报表应用优化.....	50
6.3.1 报表数据架构.....	51
6.3.2 报表展现.....	56
6.4 小结.....	60
<b>第七章 总结与展望.....</b>	<b>62</b>
7.1 总结.....	62

7.2 展望 .....	62
参考文献 .....	63
致 谢.....	64

厦门大学博硕士学位论文摘要库



## Contents

<b>Chapter 1 Introduction .....</b>	<b>1</b>
1.1 Background and Significance.....	1
1.2 Main Research and Contributions of the Dissertation .....	3
1.3 Outline of the Dissertation.....	3
<b>Chapter 2 System Analysis and Overall Design .....</b>	<b>4</b>
2.1 Goals of the Project .....	4
2.2 System Requirements Analysis and Present.....	4
2.3 The Proposed Technical Scheme .....	6
2.4 Feasible Analysis .....	8
2.5 System Architecture Design .....	10
2.6 Summary .....	15
<b>Chapter 3 Data Model Design .....</b>	<b>16</b>
3.1 Target of Data Model.....	16
3.2 Design Specification .....	16
3.3 Data Model Architecture Design .....	17
3.4 Data Storage Strategy .....	20
3.5 Summary .....	22
<b>Chapter 4 ETL Design.....</b>	<b>23</b>
4.1 Introduce of ETL Tools .....	23
4.2 ETL Processing.....	24
4.3 Data quality statistics .....	26
4.4 Abnormalities Handing .....	27
4.5 ETL scheduling monitoring .....	27
4.6 ETL Interface Design.....	30
4.7 Data Offload.....	31
4.8 Summary .....	35
<b>Chapter 5 Metadata Management and Data Quality Control .....</b>	<b>36</b>
5.1 Classification of Metadata .....	36
5.2 Metadata Management .....	36
5.3 Metadata Analysis .....	39
5.4 Data Quality Checklist.....	40
5.5 Summary .....	47
<b>Chapter 6 Reports Integration and Optimization .....</b>	<b>48</b>
6.1 Application Status Report .....	48

6.2 Target of Reports Integration .....	49
6.3 Comprehensive Report Applications Optimized.....	50
6.4 Summay.....	60
<b>Chapter 7 Conclusions and Future Work .....</b>	<b>62</b>
7.1 Conclusions .....	62
7.2 Future Work .....	62
<b>References .....</b>	<b>63</b>
<b>Acknowledgements .....</b>	<b>64</b>

厦门大学博硕士学位论文摘要库

## 第一章 绪论

### 1.1 研究背景及选题意义

随着社会信息化的飞速发展，金融业的竞争日益激烈，信息技术带来了多样化的服务产品，然而如何通过对企业已存在的大量数据中挖掘出企业经营管理者及业务分析人员所需要的信息，以支持决策分析是企业信息技术发展的当务之急。目前，信息化管理已成为银行业的必由之路。各银行已转变了经营理念和管理策略，逐渐趋向决策管理信息化。在此形势下，数据仓库的建设在国内商业银行中逐步得到了重视。从本世纪初，对于四大国有商业银行和股份制银行来说，数据的物理大集中已初步实现，银行数据整合的主题逐步转变成围绕管理主题的数据深加工。截至 2005 年，各家商业银行基本完成数据处理中心的建设和业务数据集中处理<sup>[1]</sup>。从而支持企业的联机分析处理、数据挖掘和决策管理。

数据仓库的建立是基于 DB-DW 两层体系结构，由于商业银行数据量的规模差异，以及对数据处理层次需求的不同，原有的两层结构并不能很好地满足所有银行的数据处理要求。针对这种情况，Bill Inmon 又提出了 ODS(Operational Data Store) 的概念，形成 DB-ODS-DW 三层体系结构<sup>[2]</sup>。ODS 作为数据仓库的可选部件，是建设城市商业银行数据服务的重要组成部分，可以将其定位成统一的数据存储和交换平台。作为数据存储平台，ODS 统一存储了经过简单标准化、规范化和模型化后的各个业务源系统中的原始数据。作为数据交换平台，ODS 将作为各个系统间统一的数据交换和分发平台，ODS 存储业务系统的源数据，业务系统之间的数据需求以及各个主题应用的数据集市统一从 ODS 获取数据。建立一个规整有效的 ODS 平台，构建面向分析应用的数据模型，按照主题重新进行组织、集成，保存高质量的数据存储，形成良好的银行信息体系架构。同时 ODS 相对 DW (Data Warehouse) 的硬件设备投资较小，并且可以在短时间内见成效，满足内外监管部门的数据需求<sup>[3]</sup>。

课题源自某商业银行的统一数据管理平台项目建设。该平台是基于 ODS 技

术，定位为提供全行性的统一的数据交换和存储平台，为时效性要求大于 T+1 的应用提供数据支持，并形成历史数据中心。随着银行业务的快速发展，信息系统不断增多，业务数据量的规模也在急速膨胀，信息集成问题逐渐暴露。诸如各类业务数据随其源系统之间的孤立而呈现分散的现状、分析决策支持效率较低，传统的数据加工模式难以保证数据需求的及时性和准确性、较复杂数据服务响应慢，增加了各系统的应用负担和网络传输压力等等。建设适合的数据存储和交换平台，改变银行信息系统孤立的现状，为银行业务发展和管理提供支撑的信息共享平台，是解决信息集成问题的必由之路<sup>[4]</sup>。

建设统一数据管理平台主要意义包括以下四个方面：

1. 统一数据管理平台是典型的数据整合实践，通过对业务数据的标准化和模型化，整合和覆盖了大部分的业务系统细节，可以建立全局性的业务视图，提供统一的数据存储和交换平台。通过统一数据管理平台的建设可以大大降低不同系统间的数据搬迁、冗余、和交叉验证，减少数据物理不一致的概率，基本消除数据的逻辑不一致<sup>[5]</sup>。同时，可以大量减少重复的技术和人力投入，进而大幅降低信息系统建设方面的长期投资；
2. 通过统一数据管理平台建设，详细梳理各个主要业务系统的业务数据和元数据。以此建设全行性的统一的元数据管理平台和数据质量监控平台。元数据管理和数据质量监控是两个系统性、连续性的工作，是一个不断充实、不断完善的过程。一旦形成了完善的元数据管理和数据质量监控机制，无疑会大大促进规范化的企业级数据治理，进而大大增强银行的核心竞争力<sup>[6]</sup>；
3. 通过统一数据管理平台的建设，实现跨系统的、全行性的数据集中，这为满足突发性、多变性的业务需求提供了数据基础，可以满足目前银行业务管理部门提出的绝大部分信息提取的需求，并且为分析型应用的开发奠定基础；
4. 在建设过程中通过对全行统一的展现平台的选型和二次开发，基于统一数据管理平台实现对数据结果的统一展现。并实现全行信息应用开发平台和使用的标准化，同时可以进一步规范全行性的 IT 架构，有利于加强业务管理。

## 1.2 主要研究内容

本文的主要内容如下：

1. 分析银行统一数据管理平台系统的需求现状，分别从数据质量、数据接口、数据管理以及报表展示等方面进行分析研究；研究系统建设的可行性，进而研究统一数据管理平台整体架构，包括逻辑架构、数据架构解析；
2. 介绍统一数据管理平台的数据模型和数据存储，研究数据抽取、转换、加载（ETL），包括标准化的数据处理策略、作业调度设计、监控设计和接口设计；
3. 元数据管理和数据质量检核的控制，通过统一数据管理平台的建设最终形成全行统一的元数据定义；
4. 综合报表系统的分析与应用。

## 1.3 本文的结构安排

本文共分为七章，组织结构如下：

第一章 绪论。介绍课题的研究背景及意义、研究的主要内容。

第二章 系统分析与总体设计。分析信息系统的需求现状、系统可行性、解决的技术方案，以及系统总体架构设计。

第三章 数据模型设计。研究统一数据管理平台的数据模型设计。

第四章 ETL 体系设计。重点介绍 ETL 体系设计。

第五章 元数据管理及数据质量控制。描述元数据管理和数据质量检核的平台功能和流程设计。

第六章 综合报表应用的集成与优化。基于统一数据管理平台上综合报表应用的集成和优化。

第七章 结论与展望。

## 第二章 系统分析与总体设计

本课题的研究是基于银行目前的数据存储弊端，分析业务系统的 IT 架构和数据情况，在不调整原有业务系统架构的基础上，根据银行业务主题和监管需求对数据重新统一定义和组织，形成全行统一数据视图，并建设有效的数据管理机制，保证数据的一致性和时效性。

### 2.1 系统的主要目标

1. 建立全行的数据中心，统一为全行的目标系统提供数据，屏蔽目标系统对源系统的直接数据需求，建立统一、层次合理的数据模型，集中管理全行操作型应用系统的源数据<sup>[7]</sup>；
2. 快速提供系统之间的批量数据，满足后续拓展应用系统的数据需求，合理分配数据加工功能，保证数据处理的高效性、准确性；
3. 建立标准化的数据采集、数据整合、数据转换清洗的 ETL 体系，以及 ETL 调度和监控平台，提供统一的数据批量分发机制和平台；
4. 项目将形成完善有效的数据治理机制，包括数据质量管理、元数据管理、数据标准化管理等内容，为全行提供准确、全面、可靠的数据资源，使全行的数据处于整体有序管理的状态；
5. 统一数据管理平台承接源系统批量报表、基础性业务管理报表和即时性分析应用，提供重要性和时效性较高的报表和分析应用。

### 2.2 系统需求及现状

#### 2.2.1 系统需求概述

在 80 年代直到 90 年代初联机事务处理成为数据库应用的主流<sup>[8]</sup>。但随着技术的发展和应用范围的不断扩大，传统的联机事务处理模式已经不能满足银行经营管理中新的战略要求，决策管理层需要对企业自身经营信息及整个行业动态信息进行分析，以做出适应未来发展战略决策<sup>[9]</sup>。而这些决策的制定需要对历史的、

跨业务系统的数据进行统计和分析,对这种数据的统计分析处理,称之为联机分析处理<sup>[10]</sup>。因此,势必需要通过联机分析处理技术从各类原始数据中根据业务监管和管理决策需求,在统一的数据处理平台上,对数据进行深层次分析,转化为面向业务管理者和高层决策者可理解的各项维度信息指标。

针对银行的信息系统与管理理念分析,银行的信息系统主要包括核心系统、信贷系统、反洗钱应用系统等数十个业务应用系统和管理系统。并且随着银行业务的拓展,信用卡、网上银行、交通卡、国际支付结算业务等业务的陆续推行上线,以及银行营业网点不断地增加,银行日常业务数据交易量正日益膨胀。银行管理者急需从各系统产生的大量数据中提取整合出有价值的经营信息,从而更高效地进行客户结构分析、产品结构分析、资产负债分析、员工绩效考核及风险管理与防控。因此统一数据管理平台的建设就显得十分重要。

### 2.2.2 系统现状分析

#### 1. 数据质量分析

银行的业务数据主要来自核心系统,与其它业务系统的关联绝大部分是在数据交换需求上。并且这些业务信息系统皆由不同软件公司研发,因此每个源系统都存在不同程度的数据质量问题,影响着分析与决策。

数据质量缺陷主要体现在:

- (1) 每个源系统对数据定义没有标准化,数据结构缺乏统一性,造成数据存储粒度无法满足业务部门的分析统计要求,开发人员需要付出大量精力在数据的标准化转化工作中,严重影响了数据的时效性,降低工作效率;
- (2) 由于银行的业务发展需要,核心业务系统的不断改造更新,存在着许多早期业务数据内容的不完整,数据值的缺失和不准确等问题;数据迁移造成的新旧数据格式差异;后期各项银行业务的研发带来的数据冗余等。

数据质量是风险管理工作的基础,是成功实施新巴塞尔协议的必要条件<sup>[11]</sup>。因此改善提升银行的数据质量问题是银行工作中的迫切任务。

#### 2. 数据接口分析

基于业务系统传统联机事务处理的模式下,对于各类业务系统产生的数据缺少统一的全行分发数据接口。数据传输效率低,新系统研发时需要单独对数据来

源进行数据处理层的应用开发，缺乏统一的数据处理流程机制。数据架构的拓展性较差，不足以有效地支持银行即将实施的绩效考核系统，以及未来发展中需要建设的客户关系管理、整体风险管理等主题应用。

### 3. 数据管理分析

目前银行的各个业务信息系统使用独立的传统关系型数据库进行业务数据的管理。通过数据库查询工具，使用 SQL（关系数据语言）对业务数据进行查询存储等操作。然而随着银行业务的拓展，应用系统的增多和数据量的不断增加，以数据库技术为基础的数据管理已经无法处理经营管理所需要的大量数据，很好地支持数据的管理分析。主要表现在：

- (1) 各个业务系统单独配备数据库和存储空间，这种方式势必会造成硬件存储投资和软件管理成本随系统的增加而不断增加；
- (2) 由于业务系统要求快速响应的特点，其历史数据不宜保留过长，尤其是核心业务系统通常采用磁带备份保留历史数据，各个系统缺乏统一的数据存储平台，不便于早期数据的查询；
- (3) 数据库对信息检索的响应能力不足，难以处理跨系统、跨业务、跨时间等复杂信息集成问题。数据的拼接整合需要依靠人工处理，不利于保证数据的及时性和准确性，增加了管理和营销的难度，也加大了技术人员的工作量。

通过统一数据管理平台的建设，整合归集各业务系统的数据，实现跨系统、全行性的数据集中，满足突发性、多变性的业务需求，有效管理分散在多个业务应用系统的管理数据，为分析型应用的开发奠定基础。

### 4. 数据展示分析

目前银行的报表数据展示平台较为分散，存在着核心报表平台、信贷报表平台、1104 监管报表平台、风险报表等。缺乏统一的报表数据展示平台，常常存在对其相应业务部门在同一业务需求上数据统计口径不一致、数据重复统计、对分散在各报表平台上的关联信息指标查询不便捷等问题，不利于业务部门、管理决策部门及时获取所需信息，以支持经营管理和科学决策。

## 2.3 建议的实施方案



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](mailto:etd@xmu.edu.cn) for delivery details.

厦门大学博硕士论文摘要库