

学校编码: 10384

分类号 \_\_\_\_\_ 密级 \_\_\_\_\_

学号: X2011230711

UDC \_\_\_\_\_

厦 门 大 学

工 程 硕 士 学 位 论 文

基于数据仓库的新疆财产保险业务分析与  
决策支持系统的设计与实现

Design and Implementation of the System of Xinjiang  
Property Insurance Business Analysis and Decision Support  
Based on Data Warehouse

吴俊

指导教师: 杨律青副教授

专业名称: 软件工程

论文提交日期: 2013年10月

论文答辩日期: 2013年11月

学位授予日期: 2013年 月

指导教师: \_\_\_\_\_

答辩委员会主席: \_\_\_\_\_

2013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):

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

随着保险业的飞速发展，保险公司如何通过对业务数据的归类存储、对各种要素的自动判别分析，从已存在的、大量的数据中挖掘出保险经营管理人员、业务分析人员所需要的信息，以支持决策分析成为一个迫切需要解决的问题。

数据仓库是随着关系数据库、并行处理和分布式技术的飞速发展而提出的解决使用数据的一种新技术、新概念，它是目前已知的最为广泛采用的解决方案。它不是对传统数据库的替代，而是在传统数据库的基础上对数据进行重新组织。利用数据仓库整合保险企业内部所有分散的原始的业务数据，并通过便捷有效的数据访问手段，可以支持企业内部不同部门、不同需求、不同层次的用户随时获得自己所需的信息，并能将网络中分布的商业数据集成到一起，以提供各种类型的数据分析。

本文结合人保财险新疆分公司的业务特点，提出了财产保险业务分析决策支持系统的设计及实现方案。此方案综合数据仓库、ETL 架构、OLAP 分析等技术，从不同角度对业务资料分析，对业务流程进行梳理利用 Cognos 的系列产品。通过联机分析处理工具，实现以 Web 方式进行的数据分析与表现，由此方便地进行钻取、切片和切块、多模式显示等操作，使用户从多种角度观察数据，发现潜在的信息。初步实现了省级保险数据仓库决策支持系统的要求。

**关键词：**数据仓库；数据挖掘；保险业务

## Abstract

Along with insurance business rapid development Insurance companies that how to support the decision analysis by service data of classification memory and each kind of essential factor automatic distinction analysis and also excavation existent and massively information which is insurance management and operation personnel service analysis staff need, becomes question which urgent needs to solve.

The data warehouse is one kind of new technology and the new concept for solves the service data which is proposed along with the relational database, the parallel processing and the distributional technical rapid development. It is the solution which is widely used now. It is not to the traditional database substitution but carries on in the traditional database foundation to the data organizes. Using data warehouse entire and safe enterprise interior all disperser primitive service data ,and through convenient effective data accessing method may support the information which the enterprise interior different department, the different demand and the different level user obtains as necessary oneself needs and also can integrate the commercial data which distributes the in network together, provides each kind of type the data analysis.

This article unifies the Chinese Property insurance Limited liability company Xinjiang Branch's service characteristic proposed the property insurance data warehouse design and the realization plan. This plan and summarized the information warehouse, ETL construction, OLAP analysis technologies, analyze service material analysis from different angle, carries on to the service flow combs. Realizes data analysis and performance which the Web way carries on by on-line analysis processing tool, From this carries on conveniently drills takes, slice and cutting operations, multi-pattern demonstration causes the user from many kinds of angle observed data discover the latent information Initially realized the provincial level insurance data warehouse decision support system.

**Keywords:** Data warehouse, Data mining, Insurance Business

# 目 录

<b>第一章 绪论</b> .....	<b>1</b>
<b>1.1 问题的提出与研究意义</b> .....	<b>1</b>
1.1.1 问题的提出 .....	1
1.1.2 研究的意义 .....	1
<b>1.2 国内外研究现状</b> .....	<b>2</b>
<b>1.3 本文研究的目标及论文结构安排</b> .....	<b>3</b>
<b>第二章 数据仓库的概念及其体系架构</b> .....	<b>4</b>
<b>2.1 数据仓库的概念</b> .....	<b>4</b>
2.1.1 数据仓库的作用 .....	4
2.1.2 数据仓库的基本组成 .....	5
<b>2.2 数据仓库的规划</b> .....	<b>7</b>
2.2.1 规划原则 .....	7
2.2.2 规划内容 .....	9
<b>2.3 数据仓库架构</b> .....	<b>13</b>
2.3.1 自顶向下与自底向上 .....	13
2.3.2 数据仓库架构的发展 .....	16
<b>2.4 COGNOS 的 BI 产品及特点</b> .....	<b>22</b>
2.4.1 Cognos 主要产品介绍 .....	23
2.4.2 Cognos 业务智能解决方案的主要特点 .....	25
<b>2.5 本章小结</b> .....	<b>27</b>
<b>第三章 系统的需求分析</b> .....	<b>28</b>
<b>3.1 项目背景</b> .....	<b>28</b>
<b>3.2 保险业务流程分析</b> .....	<b>28</b>
<b>3.3 功能需求分析</b> .....	<b>30</b>
<b>3.4 数据源需求分析</b> .....	<b>33</b>
3.4.1 来自业务系统的实时数据 .....	33
3.4.2 汇总数据 .....	34
3.4.3 数据粒度 .....	34

<b>3.5 非功能需求分析</b> .....	<b>35</b>
3.5.1 安全性 .....	35
3.5.2 可靠性 .....	36
3.5.3 易用性 .....	36
<b>3.6 本章小结</b> .....	<b>36</b>
<b>第四章 系统的设计</b> .....	<b>37</b>
<b>4.1 设计理念和原则</b> .....	<b>37</b>
4.1.1 设计理念 .....	37
4.1.2 设计原则 .....	37
<b>4.2 基础数据平台结构设计</b> .....	<b>38</b>
4.2.1 基础数据仓库 (Baseline) .....	38
4.2.2 数据集市 (Data Mart) .....	39
4.2.3 数据基础定义 .....	39
<b>4.3 系统的模型设计</b> .....	<b>42</b>
4.3.1 概念模型设计 .....	43
4.3.2 逻辑模型设计 .....	43
4.3.3 数据库表的设计 .....	45
<b>4.4 数据抽取、转换、加载过程 ETL</b> .....	<b>46</b>
4.4.1 数据抽取 .....	46
4.4.2 数据转换 .....	50
4.4.3 数据加载 .....	54
<b>4.5 系统的功能模块设计</b> .....	<b>54</b>
<b>4.6 本章小结</b> .....	<b>55</b>
<b>第五章 系统的实现与测试</b> .....	<b>56</b>
<b>5.1 系统的开发流程与环境</b> .....	<b>56</b>
5.1.1 硬件环境 .....	57
5.1.2 软件环境 .....	57
<b>5.2 数据分析与表现层的实现</b> .....	<b>58</b>
5.2.1 建立数据的链接 .....	58

5.2.2 生成 PowerCube .....	59
<b>5.3 数据前端展现.....</b>	<b>61</b>
5.3.1 实现方法.....	61
5.3.2 实现效果.....	62
<b>5.4 系统测试.....</b>	<b>68</b>
5.4.1 测试环境 .....	68
5.4.2 系统功能测试 .....	69
5.4.3 系统性能测试 .....	69
5.4.4 测试方法 .....	70
5.4.5 测试用例 .....	70
5.4.6 测试结果 .....	72
5.4.7 测试结论 .....	73
<b>5.5 本章小结.....</b>	<b>73</b>
<b>第六章 总结与展望 .....</b>	<b>74</b>
6.1 总结 .....	74
6.2 展望 .....	74
<b>参考文献.....</b>	<b>76</b>
<b>致 谢.....</b>	<b>78</b>



# CONTENTS

<b>Chapter 1 Introduction.....</b>	<b>1</b>
<b>1.1 The Problem and the Research Significance .....</b>	<b>1</b>
1.1.1 Put forward the Problem .....	1
1.1.2 The Significance of the Study .....	1
<b>1.2 Research Status in Nation and Abroad.....</b>	<b>2</b>
<b>1.3 Objectives and Contents of the Paper .....</b>	<b>3</b>
<b>Chapter 2 The Concept and Architecture of Data Warehouse .....</b>	<b>4</b>
<b>2.1 The Concept of Data Warehouse .....</b>	<b>4</b>
2.1.1 Role of Data Warehouse .....	4
2.1.2 The Basic Components of Data Warehouse .....	5
<b>2.2 Data warehouse planning .....</b>	<b>7</b>
2.2.1 Planning Principles .....	7
2.2.2 The Content of Planning.....	9
<b>2.3 Data Warehouse Architecture .....</b>	<b>13</b>
2.3.1 The Top-down and Bottom-up .....	13
2.3.2 The Development of Data Warehouse Architecture .....	16
<b>2.4 Product of BI of Cognos and it's Character .....</b>	<b>22</b>
2.4.1 Introduction of main products of Cognos.....	23
2.4.2 The main characteristic of BI solutions of Cognos.....	25
<b>2.5 Summary .....</b>	<b>27</b>
<b>Chapter 3 Requirement Analysis .....</b>	<b>28</b>
<b>3.1 Project Background .....</b>	<b>28</b>
<b>3.2 Insurance Business Analysis .....</b>	<b>28</b>
<b>3.3 The function requirement analysis .....</b>	<b>30</b>
<b>3.4 Analysis of the Data Source .....</b>	<b>33</b>
3.4.1 Real time Data from Business Systems.....	33
3.4.2 Tabulate Data.....	34
3.4.3 Data Size.....	34
<b>3.5 Analysis of Non Functional Requirements .....</b>	<b>35</b>
3.5.1 Security.....	35

3.5.2 Dependability .....	36
3.5.3 Ease of use.....	36
<b>3.6 Summary .....</b>	<b>36</b>
<b>Chapter 4 System Design.....</b>	<b>37</b>
<b>4.1 The Design Concept.....</b>	<b>37</b>
4.1.1 The Design Concept.....	37
4.1.2 Design Discipline.....	37
<b>4.2 Basic Data Platform .....</b>	<b>38</b>
4.2.1 Basic Data Platform .....	38
4.2.2 Data Mart.....	39
4.2.3 Data Definition.....	39
<b>4.3 Design Method of PICC Xinjiang Branch Data Warehouse .....</b>	<b>42</b>
4.3.1 Conceptual Model Design .....	43
4.3.2 The Logic Model Design .....	43
4.3.3 The Design of Database Tables.....	45
<b>4.4 Data Extraction, Transformation, the Loading Process of ETL .....</b>	<b>46</b>
4.4.1 Data extraction .....	46
4.4.2 Data conversion.....	50
4.4.3 Data loading .....	54
<b>4.5 Design of System Function Module .....</b>	<b>54</b>
<b>4.6 Summary .....</b>	<b>55</b>
<b>Chapter 5 The System Implementation and Testing .....</b>	<b>56</b>
<b>5.1 Data Show .....</b>	<b>56</b>
5.1.1 Implementation method.....	57
5.1.2 Implementation effect.....	57
<b>5.2 The Realization of Data Analysis and Presentation Layer .....</b>	<b>58</b>
5.2.1 Establishing data chaining.....	58
5.2.2 Generate PowerCube .....	59
<b>5.3 The Development Process and Environmental System .....</b>	<b>61</b>
5.3.1 Hardware environment .....	61
5.3.2 Software environment.....	62
<b>5.4 System Test .....</b>	<b>68</b>

5.4.1 Testing Environment.....	68
5.4.2 System Functional Test.....	69
5.4.3 System Performance Test.....	69
5.4.4 Testing Method.....	70
5.4.5 Testing Cases.....	70
5.4.6 Testing Results.....	72
5.4.7 Testing Conclusions.....	73
<b>5.5 Summary.....</b>	<b>73</b>
<b>Chapter 6 Conclusions and Prospect.....</b>	<b>74</b>
<b>6.1 Conclusions.....</b>	<b>74</b>
<b>6.2 Prospects.....</b>	<b>74</b>
<b>References.....</b>	<b>76</b>
<b>Acknowledgements.....</b>	<b>78</b>

## 第一章 绪论

### 1.1 问题的提出与研究意义

随着保险业的飞速发展，各保险公司体系架构逐步完善，对信息化水平的要求也越来越高。目前，基础业务分析已远远满足不了保险精细化管理的要求，保险行业需要利用数据仓库技术提高信息化水平，通过数据仓库来管理和运用已存在的、大量的数据，建立决策支持系统，从而增强企业的竞争力。保险行业的原始业务数据频繁变更，不符合传统的数据仓库的理论，所以在实现保险行业数据仓库的过程中有很多的困难<sup>[1]</sup>。

#### 1.1.1 问题的提出

目前，新兴的财产保险公司不断加入参与竞争，市场的变化要求各家保险公司不断改变经营策略，制定科学的决策方案，而没有对业务的精细化、历史的分析，就谈不上精算，更谈不上决策。如今，仅有极少数的大型保险公司才能在原运行的事物处理系统建立业务数据分析系统，但运行状况不尽人意。主要原因在于对数据的基础工作不够细致，忽略了数据质量的重要性。为了更好的解决这个问题，并结合保险业务实际，预采用面向决策层和一线管理层共用一个数据仓库，通过数据抽取、数据清洗、数据转换等过程提高数据质量，用数据仓库工具进行主题开发。

在很长的一段时间里，很多保险公司通过建立数据库软件系统来处理各种业务，它为财务、业务的关联分析打下了基础，现在需要通过数据仓库技术将现有数据变成企业经营的有用信息。

#### 1.1.2 研究的意义

对于保险公司，特别是象拥有几十年历史的人保财险公司来说，其原有历史数据存储较为分散，经过需求的不断变化，其基础数据架构也发生了很大变化。

而随着保险行业管理体系架构的不断完善,经营策略日趋多样化,公司各管理层,各业务部门对在现有系统上的联机分析有了更高的要求,其中包括各种报表和数据汇总操作,这不但影响了现有事物处理系统的性能,而且这样分析又不能满足现有管理及应用的要求,其结果往往是事倍功半。

而将数据仓库技术引入保险公司应用数据处理与分析,从数据管理角度来说,联机事务处理和数据仓库在应用中相对独立,这样使联机事务处理系统从繁重的应用分析中解脱出来,同时使统计分析更加灵活便利。对于现代保险企业来说,数据仓库技术的运用,使经营决策更科学、更及时<sup>[2]</sup>。

## 1.2 国内外研究现状

世界的历史是一门对事业进行投资以产生更多财富的学科。在农业时代,人们在土地上投资,以产生更多财富;工业革命之后,生产物品的能力变得比钱更重要,人们开始投资于工厂;而如今,人们更懂得使用信息来获得金钱或节省金钱。数据仓库就是一种增加利润的信息系统,可以用来减少费用、避免将来的花费或通过为公司决策人员提供信息以增加收入。

然而数据仓库不是一个简单的系统,国内近 10 年来,各个行业都逐步建立起了数据仓库或者数据仓库近似的系统,然而就实际的使用情况来看,很多项目从技术上都不能说很成功,更不用说是业务上的成功。截止到目前为止,国内几乎还没有一个数据仓库项目被业界公认为“技术上成功、业务上成功”,即使业界顶级的数据仓库专家也只能说我们技术上很成功。国内诸多数据仓库项目偃旗息鼓,不了了之,业界从早期的狂热到今天几乎已经到了谈“数据仓库”色变的程度,很多数据仓库应用都挂上了统计分析系统之类的名目,企图绕开数据仓库这块礁石<sup>[3]</sup>。

遗憾的是,业界对数据仓库规划少有关注,很多项目草率上马,很多项目做完一期,做二期,做完二期,做三期,每期几乎都是完全推倒重建,造成巨大的经济浪费,而且也阻碍了数据仓库的业务价值实现。

业界众多的数据仓库项目用户体验不佳,众多项目虽然采用了 BI 的工具平台和体系架构,事实上却只完成了基础报表,更多的项目 BI 应用上线后使用者寥寥,不能形成有效的互动,更不要说创造分析价值。关于这个结果的原因,业界多有讨论,从我们的观点来看,很多问题的根源可以追溯到数据仓库的实施方

方法论层面。

业界不少厂商提出了各种数据仓库实施方法论，但是从实践的效果来看，至少在业务层面上还没有真正成功的案例，而且这些方法论的出发点几乎都是数据仓库实施厂商，通常关注的也是数据仓库项目的实施，而非从企业战略的角度来规划一个整体的数据仓库/企业信息能力构建综合战略，关注项目的成功而非数据仓库的成功实现业务价值<sup>[4-5]</sup>。

### 1.3 本文研究的目标及论文结构安排

本文的目的主要是在数据仓库的设计、构建和数据挖掘进行研究的基础上，结合人保财险新疆分公司实际需求，构建数据仓库模型，并在此基础上构建多维分析模型。利用 cognos 的系列产品、IIS 服务及 Weblogic 实现分析数据的展现。

本文的结构安排主要内容包括以下几个方面：

第一章分析现有数据分析的缺陷，介绍了数据仓库技术在现代保险经营管理中的优势及其重要性。概述了数据仓库技术在国内外企业中应用的现状及数据仓库技术应用中需要注意的事项。

第二章介绍了数据仓库的基本概念及其构成要素，数据仓库的设计原则及其体系架构。

第三章保险业务数据仓库的项目分析，阐述了项目有关背景，从保险业务及数据源角度对保险业务数据仓库应用进行分析。

第四章详细讲述数据仓库在财产保险业务中的应用，为人保财险新疆分公司构建保险业务数据仓库设计、开发、部署及工程方法，对构建模型的研究，及如何通过数据质量管理架构的建设保证仓库数据的正确性和有效性。

第五章基于数据仓库的应用分析系统设计与构建，通过对 cognos 系列产品进行研究，实现可钻取的，灵活的应用分析及前台展现。

第六章对全文进行概括总结。回顾项目的设计与实现，阐明下一步所应该做的主要工作及应用技术应完善的方面。

## 第二章 数据仓库的概念及其体系架构

### 2.1 数据仓库的概念

数据仓库之父 Bill Inmon 在 1991 年出版的“Building the Data Warehouse”一书中所提出的定义被广泛接受—数据仓库 (Data Warehouse) 是一个面向主题的 (Subject Oriented)、集成的 (Integrated)、相对稳定的 (Non-Volatile)、反映历史变化 (Time Variant) 的数据集合, 用于支持管理决策 (Decision Making Support)<sup>[6]</sup>。

1、面向主题: 操作型数据库的数据组织面向事务处理任务, 各个业务系统之间各自分离, 而数据仓库中的数据是按照一定的主题域进行组织的。

2、集成的: 数据仓库中的数据是在对原有分散的数据库数据抽取、清洗的基础上经过系统加工、汇总和整理得到的, 必须消除源数据中的不一致性, 以保证数据仓库内的信息是关于整个企业的一致性的全局信息。

3、相对稳定的: 数据仓库的数据主要供企业决策分析之用, 所涉及的数据操作主要是数据查询, 一旦某个数据进入数据仓库以后, 一般情况下将被长期保留, 也就是数据仓库中一般有大量的查询操作, 但修改和删除操作很少, 通常只需要定期的加载、刷新。

4、反映历史变化: 数据仓库中的数据通常包含历史信息, 系统记录了企业从过去某一时点 (如开始应用数据仓库的时点) 到目前的各个阶段的信息, 通过这些信息, 可以对企业的发展历程和未来趋势做出定量分析和预测<sup>[7]</sup>。

数据仓库系统是一个信息提供平台, 他从业务处理系统获得数据, 主要以星型模型和雪花模型进行数据组织, 并为用户提供各种手段从数据中获取信息和知识。

#### 2.1.1 数据仓库的作用

企业中最重要资源就是它所拥有的信息。这种信息资源通常保存为两种形式, 即操作型记录系统和数据仓库。而数据仓库可以实现在业务系统和数据仓库之间形成一个隔离层, 转移一部分业务系统细节查询的功能。数据仓库要比操作

型记录系统具有更多的需求、客户、结构和节奏。

数据仓库在企业中的作用一般表现在以下几个方面：

(1) 支持全局应用许多企业在其发展过程中逐渐形成了各自独立的计算机应用(子)系统,如CAD/CAM系统、生产计划管理、库存管理、质量管理、财务管理和人事管理等系统。这些子系统有些可能是独立的,其中的数据源往往是异构的,有文件系统、层次数据库、关系数据库或面向对象数据库等等。企业信息化常常需要建立企业范围内围绕某些主题的全局应用,直接在许多分散的、不统一的数据上实施是很困难或不可能的。而数据仓库提供企业范围内的全局模式,其中存储的是经过集成的信息,来自各数据源的相关数据被转换成统一格式,方便了全局应用系统的开发。

(2) 支持决策分析在信息技术不断发展的今天,人们对信息的使用也越来越复杂。企业高层管理者需要从积累的丰富数据中提取有用信息,进行各种复杂分析,如长期趋势分析和数据开采等,以力图找出规律性的知识规划,更科学地作出决策<sup>[8]</sup>。

企业信息化中存在两类不同的数据处理:操作型数据处理和分析型数据处理。操作型数据处理也叫事务处理,如业务人员对数据库联机地进行日常操作,通常是对一个或一组记录的查询和修改。对此,人们关心的是响应时间、数据安全性和完整性。分析型数据处理则用于管理人员的决策分析,经常要访问大量来自多方面的历史数据、高度概括的数据,其数据容量非常大(到Tb级)。显然,这两者间的巨大差异使得操作型处理和分析型处理的分离成为必然。传统数据库只适用于操作型处理,数据仓库适用于分析型处理。为了满足企业信息处理需要,企业的数据库环境应发展为一种由操作型环境和分析型环境共同构成的体系化环境。数据仓库是企业数据体系化环境的组成部分,是建立决策支持系统(DSS)的基础。

### 2.1.2 数据仓库的基本组成

一个典型的企业数据仓库系统通常包含数据源、数据存储与管理、OLAP 服务器以及前端工具与应用四个部分<sup>[9]</sup>,如图2.1所示。



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