

学校编码: 10384

分类号 \_\_\_\_\_ 密级 \_\_\_\_\_

学号: X2006230018

UDC \_\_\_\_\_

厦 门 大 学

硕 士 学 位 论 文

# 基于 Teradata 的银行数据仓库模型研究与优化

Research and Optimization of Data Warehouse Model in Bank

Based on Teradata

贺晓锋

指导教师姓名: 董槐林 教授

专业名称: 软件工程

论文提交日期: 2008年4月

论文答辩时间: 2008年5月

学位授予日期: 2008年 月

答辩委员会主席: \_\_\_\_\_

评 阅 人: \_\_\_\_\_

2008年4月

## 厦门大学学位论文原创性声明

兹提交的学位论文，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的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文而产生的权利和责任。

声明人（签名）：

年 月 日



## 摘要

近年来，企业产生的关键业务数据和客户数据飞速增长，对于那些希望利用这些数据获得优势的企业，数据仓库无疑是最佳的解决手段。不过在实际情况中，许多企业在数据仓库建设中投入了大量的人力与物力，但却没有得到预期的回报。其中关键的一点就是数据仓库的模型建设不能与相关的领域很好的结合，而这也是如今各个数据仓库厂商的研究热点之一。另外依据所采用的数据仓库产品，进行针对性的模型改造也成为数据仓库模型研究的一个重要课题。

在银行领域中，模型的客户化一直是数据仓库建设的一个重要研究方向。模型的客户化包括决定模型的总体架构的逻辑模型客户化和针对特定数据仓库平台的物理模型客户化，而这两方面的设计都是需要和银行领域的业务紧密结合，因此具有很好的指导意义。同时根据银行的业务特点进行模型优化，也是一个研究重点。

针对以上问题，本文重点研究了基于 Teradata 的银行数据仓库的模型建立与优化。传统的研究都是采用了 Microsoft SQL Server 或者 Oracle 等产品，大部分的模型都只适用于某一方面的业务，不具有通用性。而本文研究的商业银行采用了业界领先的 Teradata 数据仓库，使用了成熟的 Teradata 金融领域模型，因此有效的组织了来源多样的业务数据，并使用统一的逻辑语言描述银行业务，满足了不同部门的业务需求和不同的数据访问方式。同时还根据银行业务的特点，提出了逻辑模型到物理模型客户化的设计要点，并建立起了银行自己的数据仓库模型。最后，基于后期的空间容量、执行效率等方面的要求，在保证结构层次一致性和针对具体运行环境的前提下，对现有的模型优化设计的进行了探索，并取得了初步效果。

**关键词：** 数据仓库；数据模型；Teradata

## Abstract

In recent years, a substantial increasing in the business-critical data and customer data of enterprises, data warehouse is the best solution means for those who wish to make use of these data with advantages. Many enterprises had put lot of resources in the data warehouse building, but they didn't receive the expected return. The key point is that the data warehouse model construction and related fields couldn't be very good combine, which is the one of the hotspots for all the data warehouse vendors. And researching on the reconstruction of targeted model Based on Specific products has become an important issue.

In the banking field, the customer-modeling has been an important research direction in data warehouse, which includes the logic model that decides the overall structure and the physical model that based on specific data warehouse platform, and both are need to closely integrate the fields of bank business, so it's a good guide for all. Researching on the model's optimization Based on the characteristics of bank's business is also a focus of study.

For the above-mentioned problems, this paper is focused on researching and optimization of Data Warehouse Model in Bank Based on Teradata. The traditional research are based on the Microsoft SQL Server or Oracle or other products, most of the model only applicable to a business on the one hand, is not universal. In this paper, the commercial bank adopted the industry-leading Teradata data warehouse, and it used a mature financial model, organized the diverse sources of business data effectively, and used unified logic language for description of banking business to meet the different departments. So it satisfies the different Business needs and data access. Besides, in accordance with the characteristics of the banking business, we propose the design features of the logic customer-modeling to the physical customer-modeling, and to establish its own bank data warehouse model. Subsequently, because of the requirements of storage space and efficiency, we explore optimized design based on assurance for consistency and specific operating environment, and have achieved initial results.

**Key Words:** Data Warehouse; Data Model; Teradata

## 目录

<b>第一章 绪论</b> .....	<b>1</b>
1.1 研究背景 .....	1
1.2 研究动机 .....	2
1.3 研究目标与技术路线 .....	2
1.4 论文组织结构 .....	3
<b>第二章 数据仓库概述与 Teradata 技术分析</b> .....	<b>4</b>
2.1 数据仓库发展历史 .....	4
2.2 数据仓库相关概念 .....	4
2.2.1 OLAP 和 OLTP .....	4
2.2.2 数据仓库查询特点 .....	7
2.2.3 数据仓库体系结构 .....	8
2.3 数据仓库与数据集市 .....	8
2.3.1 关系模型与星型模型 .....	9
2.3.2 数据集市 .....	10
2.4 Teradata 方法论与技术特点 .....	10
2.4.1 Teradata 可扩展数据仓库方法论 .....	11
2.4.2 Teradata 可扩展数据仓库框架 .....	12
2.4.3 Teradata 技术特点 .....	13
<b>第三章 数据仓库与数据建模</b> .....	<b>17</b>
3.1 Teradata 概念模型 .....	17
3.2 数据模型 .....	18
3.2.1 数据模型的概念 .....	19
3.2.2 数据模型的意义 .....	19
3.2.3 数据模型的建立 .....	20
<b>第四章 基于 Teradata 的银行数据仓库模型</b> .....	<b>26</b>
4.1 基于 Teradata 的 LDM 与 PDM 分析 .....	27
4.1.1 LDM 的概念 .....	27
4.1.2 LDM 的作用 .....	27
4.1.3 LDM 的设计 .....	28
4.1.4 PDM 与 LDM .....	30
4.2 基于 Teradata 模型的客户化分析与设计 .....	30
4.2.1 模型客户化设计原则 .....	30
4.2.2 LDM 客户化分析 .....	31
4.2.3 PDM 客户化分析 .....	32
4.2.4 LDM 客户化到 PDM 客户化的设计 .....	35
4.3 建立银行领域的 BANK-LDM 模型 .....	40
4.3.1 BANK-LDM 概念层模型设计 .....	40

4.3.2 BANK-LDM 主题域模型设计 .....	40
4.3.3 BANK-LDM 模型详细设计 .....	42
4.3.4 BANK-LDM 成果与总结 .....	52
<b>第五章 基于 Teradata 的银行数据仓库模型优化 .....</b>	<b>55</b>
<b>5.1 物理模型优化.....</b>	<b>55</b>
5.1.1 物理模型优化设计原则.....	55
5.1.2 物理模型优化设计目标.....	56
5.1.3 物理模型优化详细设计 .....	56
<b>5.2 逻辑模型优化.....</b>	<b>60</b>
<b>5.3 模型优化统计模板设计 .....</b>	<b>60</b>
<b>5.4 模型优化部分成果 .....</b>	<b>62</b>
<b>第六章 总结与展望 .....</b>	<b>64</b>
<b>参考文献 .....</b>	<b>65</b>
<b>致谢 .....</b>	<b>67</b>

# Contents

<b>Chapter 1 Introduction .....</b>	<b>1</b>
<b>1.1 Background.....</b>	<b>1</b>
<b>1.2 Motivation.....</b>	<b>2</b>
<b>1.3 Target and Technology Roadmap .....</b>	<b>2</b>
<b>1.4 Architecture of Thesis.....</b>	<b>3</b>
<b>Chapter 2 Data Warehouse and Teradata Analysis .....</b>	<b>4</b>
<b>2.1 Development History of Data Warehouse.....</b>	<b>4</b>
<b>2.2 Relative Concept of Data Warehouse.....</b>	<b>4</b>
2.2.1 OLAP and OLTP .....	4
2.2.2 Query Characteristics of Data Warehouse .....	7
2.2.3 Architecture of Data Warehouse .....	8
<b>2.3 Data Warehouse and Data Mart.....</b>	<b>8</b>
2.3.1 Relation Schemas and Star Schemas .....	9
2.3.2 Data Mart .....	10
<b>2.4 Methodology and Technology Characteristics of Teradata.....</b>	<b>10</b>
2.4.1 Scalable Data Warehouse Methodology .....	11
2.4.2 Framework of Scalable Data Warehouse .....	12
2.4.3 Technology Characteristics of Teradata .....	13
<b>Chapter 3 Data Warehouse and Data Modeling.....</b>	<b>17</b>
<b>3.1 Concepts Model of Teradata .....</b>	<b>17</b>
<b>3.2 Data Model .....</b>	<b>18</b>
3.2.1 Concepts of Data Model .....	19
3.2.2 Significance of Data Model .....	19
3.2.3 Models Building.....	20
<b>Chapter 4 Data Warehouse Model in Bank Based on Teradata .....</b>	<b>26</b>
<b>4.1 LDM and PDM Analysis Based on Teradata.....</b>	<b>27</b>
4.1.1 Concepts of LDM.....	27
4.1.2 Functions of LDM.....	27
4.1.3 LDM Design .....	28
4.1.4 PDM and LDM .....	30
<b>4.2 Customer-modeling Analysis and Design Based on Teradata.....</b>	<b>30</b>
4.2.1 Principles of Customer-modeling .....	30
4.2.2 LDM Customer-modeling Analysis .....	31
4.2.3 PDM Customer-modeling Analysis .....	32
4.2.4 LDM to PDM in Customer-modeling.....	35
<b>4.3 Building BANK-LDM Model.....</b>	<b>40</b>
4.3.1 CDM Design of BANK-LDM .....	40
4.3.2 Theme Domain Model Design of BANK-LDM .....	40
4.3.3 Detailed Design of BANK-LDM.....	42
4.3.4 Results and Summarization of BANK-LDM.....	52



<b>Chapter 5 Model Optimization in Bank Based on Teradata.....</b>	<b>55</b>
<b>5.1 Physical Model Optimization.....</b>	<b>55</b>
5.1.1 Principles of Physical Model Optimization .....	55
5.1.2 Target of Physical Model Optimization .....	56
5.1.3 Detailed Design of Physical Model .....	56
<b>5.2 Logical Model Optimization .....</b>	<b>60</b>
<b>5.3 Template Design in Statisticof Model Optimization .....</b>	<b>60</b>
<b>5.4 Results of Model Optimization .....</b>	<b>62</b>
<b>Chapter 6 Conclusion and Prospect .....</b>	<b>64</b>
<b>References.....</b>	<b>65</b>
<b>Acknowledge .....</b>	<b>67</b>

厦门大学博硕士论文摘要库

## 第一章 绪论

决策支持系统的建设已经越来越受到重视，特别在金融领域，如何通过数据仓库技术建立更专业有效的信息管理机制已经成为一个热点问题。这里我们将对数据仓库及其模型的研究现状以及存在的问题等进行阐述，最后对本文研究内容以及本文的结构安排等进行总体概述。

### 1.1 研究背景

信息是企业的命脉，是竞争优势的关键所在。随着现代技术的发展，企业产生的关键业务数据和客户数据早已汗牛塞屋。对于那些希望利用信息来赢得优势的企业，数据仓库无疑是其建立商业智能的基础。

特别是对于国内的银行来说，如何把自身打造成现代化的银行，制胜的关键就在于金融信息化建设。近年来各银行纷纷致力于进行数据大集中，相继开发了较完善的业务系统，但在对客户信息的挖掘分析能力上与外资银行相比尚有很大差距，单纯的数据集中和系统升级并不能完全实现数据和信息的传递、共享和应用。这就要求国内银行在信息技术上加强决策支持系统的建设，通过数据仓库技术建立更专业有效的信息管理机制，为科学化管理决策和业务创新服务。

不过现实情况是，即使是一些拥有成熟的商业智能方案的领先企业，也在为如何经济高效的利用信息而困扰。无论是系统的复杂性，还是庞大的数据量和高性能要求，都使得企业在实施数据仓库、商业智能以及应对不断提升的用户需求时显得力不从心。许多企业已经在数据仓库建设中投入了大量的人力与物力，但却没有得到预期的回报。而其中很关键的一点就是数据仓库的模型建设不能与相关的领域很好的结合起来，特别是在物理模型和逻辑模型之间不能找到一个平衡点。

对于本文研究的商业银行，在一期的项目中已经建立起了企业级数据仓库基础模型，但是由于项目一期的数据量相对较小和应用相对较少，随着数据量的不断加大，应用需求的不断增多，原有的物理模型设计渐渐无法满足空间容量、执行效率等方面的要求，因此必须对现在的物理模型做针对 Teradata 物理特性和银行现有运行环境的技术调整，也就是物理模型优化。

## 1.2 研究动机

模型建设是数据仓库建设中重要部分。一般来说，模型建设分为逻辑模型设计和物理模型实现两部分，本文中研究的商业银行模型 BANK-LDM(特指根据 Teradata FS-LDM 客户化后，此商业银行的专有 LDM，下同)，其中的逻辑模型设计是以 Teradata 成熟的金融业逻辑数据模型 FS-LDM(Financial Services Logical Data Model) 7.0 为蓝图和指引，通过对银行自身数据源分析的结果并客户化而成；物理模型实现则是对客户化的逻辑型的物理实现。逻辑数据模型是与数据库平台无关的数据仓库逻辑蓝图，物理数据模型是逻辑数据模型的物理实现，是与数据库平台紧密相关，与真实数据存储、数据加载、数据查询等密切相关的。

Teradata 是一个关系数据库管理系统，能够把企业的交易转变成关系。它采用标准的 SQL 查询语言，但独特的内部结构特别适合于处理复杂查询数据仓库应用。其良好的扩展性能够随着业务的发展而发展，从 GB 级扩展到 100TB 以上。目前世界上最大的数据仓库是 SBC Communications Inc 的数据仓库，数据量达到 128TB，就是采用 Teradata 数据库实现的。

Teradata 在数据仓库实施方面有着丰富的经验，并且在金融领域也有丰富的经验，现有的数据仓库物理数据模型就是 Teradata 在结合银行企业级数据仓库逻辑数据模型设计原则和精髓的基础上设计的，使 Teradata 在数据仓库方面的先进性、大数据量查询高效性、高可扩展性和稳定性等方面的优势和特点得以全面的发挥。因此，对于现有的模型研究有非常重要的意义，对于今后各个银行在建立自己的数据仓库模型时有一定的参考价值。

另外，现有的物理模型实现是依托于客户化后的原有逻辑模型实现的，在实现过程中基本没有考虑模型优化的问题，仅仅是做了简单的逻辑模型到物理模型的一一对应。随着数据量的不断加大，应用需求的不断增多，原有的物理模型设计渐渐无法满足空间容量、执行效率等方面的要求，因此必须对现在的物理模型做针对 Teradata 物理特性和银行的现有运行环境的技术调整，也就是物理模型优化，这也是本文所关注的一个方面。

## 1.3 研究目标与技术路线

一个逻辑数据模型是建立商业智能的基础框架，也是建立一个灵活的强有力的数据仓库系统的第一步，并且是奠定现在或者将来为知识工作者提供有价值数据分析的重要基础。因为基于 Teradata 数据仓库的 FS-LDM 是按照三范式的原则设计的，是预先构建的逻辑数据模型，利用它可以直接开始数据仓库模型设计，可以运行在任何数据库和平台上，与 Teradata 数据库无关。并且遵循了灵活易扩展、面向主题的设计方法，从而使增加功能时不需要重构

整个数据仓库。

在逻辑数据模型设计方面，研究重点是逻辑数据模型设计时的主要原则和特点，从而发现在逻辑模型设计时是如何在数据建设阶段为解决业务需求而定义解决方案，和逻辑模型在设计时如何针对很多的具体的应用和系统进行详细的客户化。

在物理数据模型设计方面，设计的主要目标和关键点是将逻辑模型结合数据仓库平台的特点来物理化，优化数据存储，节约数据的物理成本开销，使数据加载和查询应用保持高效。本文研究的商业银行模型特点是以信息的集中和共享为基础、保证数据的一致性，并且设计是采用图形的展现方式和面向业务主题的方法，来对各种来源的数据进行有效的组织，并使用统一的逻辑语言对业务进行描述，从而有效的解决商业银行缺乏有效的数据管理的分析工具和交流手段的问题。

在后期的模型优化方面，研究的重点在于如何根据商业银行的业务需要，进行源系统产生的代码数据整理收集，并且对源系统代码和通用数据标准的公共代码数据进行差异分析。预期的结果是制定出各业务、各系统所需要共同遵守的数据标准。在具体的模型优化方面，如何充分利用 Teradata 数据仓库优秀的海量并行处理能力，使物理模型实现三范式设计，对今后的大型数据仓库建设有一定意义。

## 1.4 论文组织结构

第一章主要阐述了论文的研究背景、目标和基本的技术路线。

第二章深入分析了数据仓库的相关概念和体系结构，包括 Teradata 数据仓库方法论和相关的技术特点。

第三章结合 Teradata 数据仓库模型的基本概念，从数据模型的概念及意义入手，探讨了数据仓库模型的架构，并提出了建模四个阶段的划分，总结了三种基本的建模方法。

第四章根据文中研究的商业银行的具体情况和 Teradata FS-LDM 模型，讨论了 LDM 及 PDM 的客户化原则，并设计了 LDM 客户化到 PDM 客户化的原则，最后建立了 BANK-LDM 模型。并展现了建立的概念模型和主题域及相关成果(由于涉及保密问题，概念模型和 11 个主题域没有详细分析)。

第五章以上一章节的模型分析为基础，提出了详细的模型优化原则及优化路线，其中主要为针对物理模型的优化(由于涉及保密问题，没有写出具体的优化细节)。并在后期的实际生产环境中进行了初步验证，取得了预期的效果。

第六章对本文的研究内容和阶段性成果进行了概括。

## 第二章 数据仓库概述与 Teradata 技术分析

本章主要探讨了目前数据仓库的主要的技术特点，和 Teradata 提出的数据仓库方法论，并分析了 Teradata 产品的技术特点。

### 2.1 数据仓库发展历史

七十年代中期出现了 MIS(管理信息系统)，但 MIS 具有很大的局限性。首先，它基本上是按照预先定义好的流程对数据作相应的处理，因此只能对预先描述好的业务问题进行回答<sup>[1,2]</sup>；其次，由于开发工具的限制，对它的修改也不太方便，特别是当业务流程发生变化，模型需要调整时，这种修改变得更加困难；最后，由于数据的不断积累，数据量迅速增加，普通的商用数据库(即 OLTP 数据库)难以处理，系统的扩展存在很大的限制。

在这种情况下，MIS 逐步发展到了数据仓库。世界上最早的数据仓库是 Teradata 公司为全美、也是全世界最大的连锁超市集团 Wal-Mart 在 1981 年建立的，而最早将数据仓库提升到理论高度进行分析并提出数据仓库这个概念的则是著名学者 Bill Inmon，他对数据仓库所下的定义是<sup>[3]</sup>：数据仓库是面向主题的、集成的、稳定的、随时间变化的数据集合，用以支持管理决策的过程。由此可见，数据仓库是一个综合的解决方案，主要用来帮助企业有关主管部门和业务人员做出更符合业务发展规律的决策<sup>[4]</sup>。因此，在很多场合，决策支持系统也成了数据仓库的代名词<sup>[5]</sup>。

### 2.2 数据仓库相关概念

主要阐述了在数据仓库发展过程中几个重要的概念，及数据仓库的传统体系结构。

#### 2.2.1 OLAP 和 OLTP

数据仓库是在数据库基础上发展而来的，通常包含：数据仓库、联机分析处理(OLAP, On-Line Analytical Processing)、数据挖掘(Data Mining)<sup>[6]</sup>。

OLAP (联机分析处理 On-Line Analytical Processing)是基于数据仓库的信息分析处理过程，是数据仓库的用户接口部分。OLAP 系统是跨部门、面向主题的<sup>[10,11]</sup>，其基本特点是：

- 本身不产生数据，其基础数据来源于生产系统中的操作数据(Operational Data)<sup>[13]</sup>
- 基于查询的分析系统

- 复杂查询，经常使用多表连结、全表扫描等，牵涉的数据量往往十分庞大
- 响应时间与具体查询有很大关系
- 用户数量相对较小，其用户主要是业务人员与管理人员
- 由于业务问题的不固定，数据库的各种操作不能完全基于索引进行

早期的数据库系统主要应用于日常事务的操作性处理，即 OLTP(联机事务处理 On-Line Transaction Processing)，重点在于完成事务处理，在短时间内给予用户响应，服务于操作型业务需求，可以称之为“操作型数据库”<sup>[14]</sup>。

OLTP 系统也称为生产系统，它是事件驱动、面向应用的<sup>[14]</sup>，比如银行的储蓄系统就是一个典型的 OLTP 系统。OLTP 的基本特点是：

- 数据在系统中产生
- 基于交易的处理系统 (Transaction-Based)
- 每次交易牵涉的数据量很小
- 对响应时间要求非常高
- 用户数量非常庞大，主要是操作人员
- 数据库的各种操作主要基于索引进行

从数据库结构来看，如果用银行作为一个例子，两者的区别可以用图 2-1 表示：

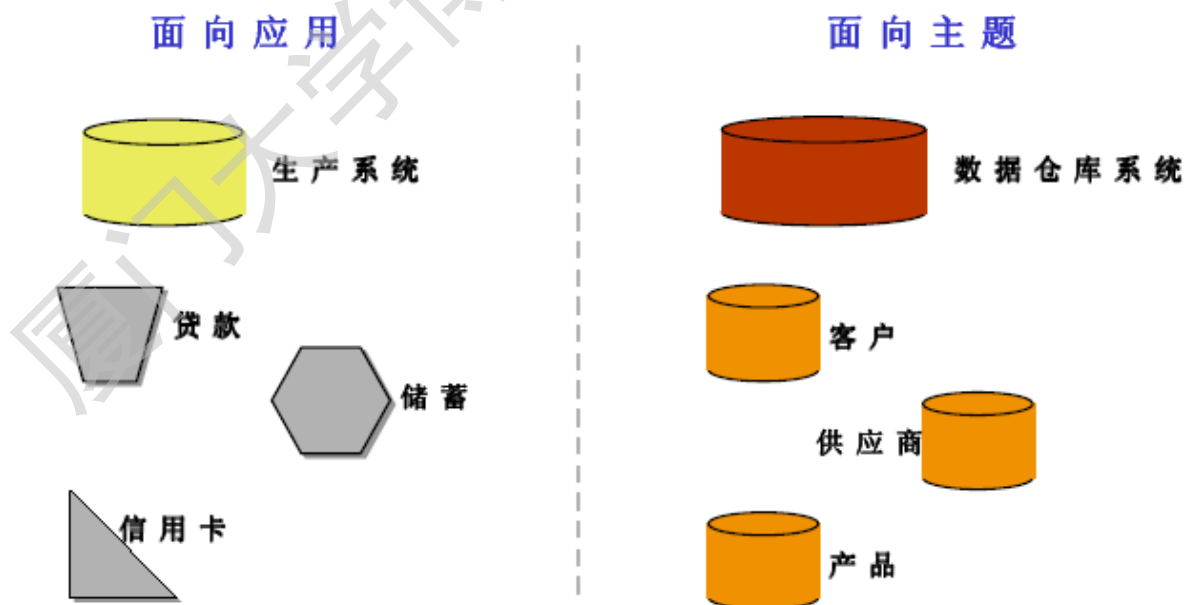


图 2-1 面向应用的 OLTP 系统与面向主题的 OLAP 系统

对银行而言，一般都针对其各项业务开发了相应的 OLTP 系统，用于解决银行日常的生

产自动化问题，如储蓄系统、信用卡系统等。这些系统都是用来处理银行的日常业务的。而图 2-1 中右边显示的数据仓库系统则是面向主题的。到底哪些内容是主题呢？这与具体要解决的业务问题有关系。对银行来说，客户是一个永恒的主题，要研究客户的消费行为和心理，就要对客户进行各种分类，以此解决哪些客户的信用好、哪些客户的风险较大，等等诸如此类的问题。另外，银行的产品(如贷款、储蓄等业务项目)也是一个值得研究的主题，据此可以分析这些产品的市场定位情况。

把 OLAP 与 OLTP 两者的分离<sup>[1]</sup>，具体来说主要有以下五个原因：

- 业务应用的响应时间不同

在操作型（事务处理）计算环境中，其业务应用的处理特点是数据的存取操作频率高，而每次操作处理的时间短。因此，用户的业务请求往往在很短时间内即可完成，使得系统在多用户并发访问的情况下，也可以保持较高的实时响应速度。

在分析型（决策支持）计算环境中，其决策分析问题的解决，往往需要遍历数据库中的大部分数据，不仅需要消耗大量的系统资源，还可能导致长达数小时的运行。因此，用户的业务请求并不要求瞬间响应，只要在允许的时间里给予响应即可。

- 数据需求的集成要求不同

在操作型数据库中，数据往往是分散而非集成的。主要原因有操作型（事务处理）应用分散的问题、数据不一致问题、外部数据和非结构化数据等问题。

在分析型数据库中，首先需要解决的就是数据集成的问题。只有全面、广泛、准确、相关的数据收集和有效集成，才有可能得到可信的分析结果。

- 历史数据的使用价值不同

操作型（事务处理）系统一般只处理当前数据，在数据库中也只是短期存储数据，且不同数据的保存期限也不一样，即使有一些历史数据保存下来了，无特殊需求一般不再访问。

在分析型（决策支持）系统中，历史数据是相当重要的，因为联机分析处理与数据挖掘都需要大量的、详细的、且有一定时间、空间跨度的历史数据为基础，才可能发现潜在发展规律和趋势走向。

- 服务支持的用户对象不同

操作型（事务处理）系统直接面向业务操作人员，提供数据处理支持。

分析型（决策支持）系统面向使用本系统的业务者和中高层管理人员，提供决策分析支持。

- 数据的综合问题

在事务处理系统中积累了大量的细节数据，一般而言，决策支持系统并不对这些细节数据进行分析。在分析前，往往需要对细节数据进行不同程度的综合。而事务处理系统不具备这种综合能力，根据数据库的规范化理论，这种综合还往往因为是一种数据冗余而加以限制。这也是二者相矛盾的地方。

综上所述，要提高分析和决策的有效性，必须把分析型数据从操作型（事务处理）环境中分离出来，按照分析型（决策支持）系统处理的需要进行重新组织，建立相应的分析处理环境。数据仓库正是为了构建这种新的分析处理环境而出现的一种数据存储、组织和处理技术。

### 2.2.2 数据仓库查询特点

数据仓库和生产系统不同，它保存的是历史交易数据，因此用户针对数据仓库的交易主要是查询。数据仓库的查询和生产系统的查询有很大的区别。

针对生产系统的查询都比较简单，不常使用表的连接操作(Table Join)，每次操作所牵涉的数据量很小。这类问题的特点是“知道自己要找什么”，例如根据银行帐号查余额。这类操作往往是基于索引进行的，如帐号可以作为储蓄系统的索引。由于这些特点，数据量大小对系统性能影响不大。

而针对数据仓库系统的查询大都非常复杂，主要有两种：一种以报表为主，从数据库中产生各种形式的业务报表。这种查询是预先规划好的(Pre-Defined Query)；另一种则是随机的、动态的查询(Ad-Hoc Query)，对查询的结果也是不能预料的。例如世界上最大的连锁超市集团 Wal-Mart 曾利用查询和数据挖掘工具发现周末时啤酒与一次性尿布顾客一块购买的机会最大，结果出人意料。经过分析，原来是美国的太太们在周末时常叮嘱她们的丈夫在下班后为小孩买尿布，而丈夫们在买尿布时又顺手捎带些啤酒。发现这一规律后，Wal-Mart 就在它的各个连锁商店里把这两种商品放在一起，方便父亲们购买，结果这两种商品的销售量都增加了很多。这个故事已经成了数据仓库成功应用的经典案例。

数据仓库中的查询由于其复杂性，会经常使用多表的联接、累计、分类、排序等操作，这些大都要对整个表进行搜索(Full Table Scan)，特别消耗系统资源，而且每次查询返回的数据量一般都很大。对于 Ad-Hoc 查询而言，经常需要根据上次查询的结果进行进一步的搜索和分析，以便找到蕴含在数据中的潜在规律。

根据这些特点，数据库大小对数据仓库系统的性能影响很大。当数据仓库投入使用后，各业务部门的要求会越来越多，使得数据仓库中数据量的增长速度很快，因此设计数据仓库



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