

基于数据挖掘的商业智能平台的设计与实现

李金泽

指导教师 姚俊峰 教授

厦门大学

厦门大学博硕士学位论文摘要库

学校编码: 10384

分类号_____密级_____

学号: X2013230289

UDC _____

厦门大学

工 程 硕 士 学 位 论 文

基于数据挖掘的商业智能平台
设计与实现

Design and Implementation of Business Intelligence
Platform Based on Data Mining

李金泽

指导教师:	姚俊峰 教授
专业名称:	软 件 工 程
论文提交日期:	2015年4月
论文答辩日期:	2015年5月
学位授予日期:	2015年6月

指导教师: _____

答辩委员会主席: _____

2015年4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）：

2015 年 月 日

摘 要

商业智能是一个系统工程，它囊括了许多的概念和方法，应用在基于事实的支持系统来辅助制定商业决策。商业智能技术指的是使得企业能够对历史数据进行迅速分析的技术，包含了对数据收集、处理、分析，并且将数据结果转变成包含规则、对决策者有用的信息。商业智能目的是使企业的各级决策者能够从以往数据中获取知识以及规律，辅助决策者做出更为快捷、正确的决策

本文主要是研究商业智能产生、发展趋势、在企业中的应用，分析目前商业智能软件的研究现状与发展趋势以及不足之处；学习和研究商业智能平台的理论、技术、架构、流程等；然后针对企业在商业智能平台的需求，将企业关注的、真正需要的功能作为重点进行分析；并尝试运用所学的理论和技术，结合现有的开发环境、语言、工具插件，开发出一个完整、易用、高效的商业智能平台原型系统。

本文采用面向对象、模块化、面向服务（软件即服务，SaaS）以及提出的软件总线接口的设计思想，将商业智能平台架构分为若干个功能模块，针对当前信息化应用的特点，形式上采用 B/S 架构模式，功能上采用多层次、模块化的软件功能架构，技术上采用基于 MVC3.0 基础的体系架构，以 C# 为主要编程语言，利用 XML 作为配置文件以及 ADO.Net, Entity Framework, ExtJS 等相关技术实现了商业智能流程中的系统管理和挖掘工作调度、执行等功能。提出建立挖掘模型评价推荐系统，推荐挖掘模型在此基础上实现不断完善和功能扩展，为商业企业提供高效、共享的挖掘模型。在平台控制和维护方面，建立灵活管理的控制机制，实现平台系统的维护和调整，提供基于总线思想的挖掘工作模块接口，达到平台可维护性、可恢复性和可更新性的设计目的。作为平台实际检验，结合商业智能挖掘需要，设计实现了多个商业挖掘算法，进一步验证了平台系统的实用性和可用性，为提高企业决策水平发挥一定作用和影响。

本论文以优化商业智能挖掘平台开发作为设计目标，从宏观到微观、从定性到定量、从广度到深度、从点到面真正透彻、全面、深入地挖掘企业经营状况和内在规律，把握市场变化趋势，为企业提供重要的决策指引等多方面入手，为商业智能应用探索一条有效途径。

关键词：商业智能；数据挖掘；B/S 架构

Abstract

Business intelligence is a set of systems, including a series of concepts and methods used in fact-based support system to aid in making business decisions. Business intelligence technology is a techniques and method what can make enterprises analyze the historical data quickly, which includes data collection, management, analysis, and data conversion to regular and useful information for decision-makers. The purpose of business intelligence is to prompt all levels of decision-makers in the enterprise to make more rapid, correct decisions, during knowledge or rules from historical data,

In this thesis, the source, development trends, application in the enterprise of business intelligence are studied. The current status quo and development trend of business intelligence software, as well as inadequacies are analyzed. The theory, technology, architecture, processes, etc. of business intelligence platform are studied. Then for the needs of business, the function what business concern and really need is focused on analysis. Applying the learned theories and techniques, combined with existing development environments, languages, plug-in tools, a complete, feasible and efficient prototype system of business intelligence platform is developed.

In this thesis, with the design ideas of object-oriented, modular, service-oriented (Software as a Service, SaaS) as well as the software bus interface, business intelligence platform architecture is divided into several functional modules. According to the characteristics of the current information technology application, the B / S structure mode is applied in the form; multi-level, modular software architecture is applied in the function; technically based on MVC3.0 architecture, with the C # programming language, XML as the configuration file and the related technologies such as ADO.Net, the Entity Framework, Ext JS, these functions of system management, excavation work scheduling, and implementation in the business intelligence process are realized. Based the mining model to evaluate the recommendation system, the recommended mining model can make continuous improvement and extensions, to provide efficient, shared mining model for these commercial enterprises. In the term of platform control and maintenance, a flexible control mechanism is established to realize maintenance and adjustment of the platform, to provide the excavation module interface based on bus thinking, which realizes the design purpose of platform maintainability, recoverability and

updatibility. As actual platform inspection, combined with the need of business intelligence mining, a number of commercial mining algorithms are designed and implemented to further verify the practicality and usability of the platform, which plays some role and influence to raise the level of corporate decision-making.

In this thesis, to optimize business intelligence mining platform developed as a design goal, from macro to micro, from qualitative to quantitative, from breadth to depth, from point to plane, business conditions and internal rules are real thoroughly, comprehensively, deep mined. The changes in the market trends are grasped. These provide important decision-making guidelines for enterprises and explore an effective approach to business intelligence applications.

Key word: Business Intelligence; Data Mining; B/S Architecture

目 录

目 录	IV
第一章 绪论	1
1.1 商业智能概念及发展	1
1.1.1 商业智能概念	1
1.1.2 商业智能软件研究现状	2
1.1.3 商业智能领域发展趋势	3
1.2 论文研究的背景与意义	4
1.2.1 论文研究的背景	4
1.2.2 论文承担的主要工作	4
1.2.3 论文工作的主要特色	5
1.3 本文研究内容	5
1.3.1 学习和掌握商业智能的主要理论和技术	6
1.3.2 分析现有商业智能体系结构存在的问题	6
1.3.3 分析和设计商业智能体系结构与软件平台模型	6
1.3.4 分析和解决技术难点	6
1.4 论文组织结构	7
第二章 相关技术介绍	8
2.1 商业智能的基本概念和体系结构	8
2.2 数据仓库 (DW)	8
2.3 数据预处理 (ETL)	10
2.4 联机分析处理 (OLAP)	10
2.5 数据挖掘 (DM)	11
2.6 数据查询与数据可视化	11
2.7 本章小结	12
第三章 系统需求	13
3.1 可行性分析	13
3.2 业务流程分析	13
3.3 用户角色分析	15
3.4 功能性需求分析	15
3.5 非功能性需求	17

3.6 安全性分析.....	17
3.7 本章小结.....	17
第四章 系统设计	19
4.1 总体设计.....	19
4.2 技术路线.....	21
4.2.1 系统流程.....	23
4.2.2 系统数据库关系.....	25
4.2.3 运行环境.....	27
4.3 数据库设计.....	27
4.3.1 数据仓库和多维数据集配置文件.....	28
4.3.2 挖掘工作配置文件.....	28
4.3.3 挖掘结果配置文件.....	29
4.4 本章小结.....	30
第五章 系统实现	31
5.1 开发环境与工具.....	31
5.2 元数据管理模块.....	31
5.3 算法接口与加载模块.....	33
5.4 挖掘工作模块.....	36
5.5 工作执行与代理模块.....	38
5.6 海量数据处理.....	40
5.7 挖掘模型评价推荐模块.....	43
5.8 可视化模块.....	44
5.9 平台管理模块.....	46
5.10 本章小结.....	48
第六章 系统测试	49
6.1 测试用例和工具.....	49
6.2 流程测试.....	49
6.3 执行挖掘工作测试.....	51
6.3.1 挖掘工作执行器测试.....	51
6.3.2 挖掘结果展示测试.....	52
6.4 服务器压力测试.....	53
6.4.1 服务器空载.....	53

6.4.2 平台挖掘工作执行代理运转.....	55
6.5 测试总结.....	56
6.6 本章小结.....	57
第七章 总结和展望.....	58
7.1 总结.....	58
7.2 展望.....	58
参考文献.....	60
致 谢.....	63

厦门大学博硕士学位论文摘要库

Contents

Chapter1 Introduction	1
1.1 Business Intelligence Concept and Development	1
1.1.1 Business Intelligence Concept.....	1
1.1.2 Research of Intelligent Software Business.....	2
1.1.3 The Development Trend of The Field of Business Intelligence.....	3
1.2 The Background and Significance of The Paper Research	4
1.2.1 The Background to The Study.....	4
1.2.2 The Main work of This Thesis is to Bear.....	4
1.2.3 The Main Features of The Work.....	5
1.3 The Research Contents of This Paper	5
1.3.1 The Main Theory and Technology to Learn and Master The Business Intelligence.....	6
1.3.2 Analysis of The Existing Problems of Business Intelligence Architecture.....	6
1.3.3 Analysis and Design of Business Intelligence System Structure and Software Platform of Model.....	6
1.3.4 Analyze and Solve The Technical Difficulties.....	7
1.4 Structure	7
Chapter2 Technology Introduction	8
2.1 The Basic Concept and System Structure of Business Intelligence	8
2.2 Data Warehouse(DW)	8
2.3 Data Pre Processing(ETL)	10
2.4 On Line Analytical Processing(OLAP)	10
2.5 Data Mining(DM)	11
2.6 Data Query and Data Visualization	12
2.7 Summary	12
Chapter3 The System Requirements	13
3.1 Feasibility Analysis	13
3.2 Business Process Analysis	13
3.3 User Role Analysis	15
3.4 The Functional Requirements Analysis	15
3.5 Non Functional Requirements	17
3.6 Safety Analysis	17
3.7 Summary	17

Chapter4 Analysis of Demand	19
4.1 System Requirements Analysis	19
4.2 Technical Route	21
4.2.1 System Flow.....	23
4.2.2 Relational Database System.....	25
4.2.3 Operation Environment.....	27
4.3 The Mining Model to Evaluate The Recommended Requirements	27
4.3.1 The Requirement of Visualization.....	28
4.3.2 Safety Requirements.....	28
4.3.3 The Summary of This Chapter.....	29
4.4 Summary	30
Chapter5 The Realization of The System	31
5.1 The Development Environment and Tools	31
5.2 The Metadata Management Module	31
5.3 Algorithm Interface and Loading Module	33
5.4 Mining Module	36
5.5 The Execution of The Work and The Agent Module	38
5.6 Massive Data Processing	40
5.7 The Mining Model to Evaluate The Recommendation Module	43
5.8 The Visualization Module	45
5.9 Platform Management Module	46
5.10 Summary	48
Chapter6 System Test	49
6.1 Test Case and Test Tools	49
6.2 Process Test	49
6.3 The Implementation of Mining Work Test	51
6.3.1 Mining Actuator Test.....	51
6.3.2 The Results Show The Test Mining.....	52
6.4 Server Stress Test	53
6.4.1 The Server Load.....	53
6.4.2 Mining Executive Agent Operation Platform.....	53
6.5 Test Summary	56
6.6 Summary	57
Chapter7 Summary and Prospect	58
7.1 Summary	56

7.2 Prospect.....	58
References.....	60
Acknowledgements.....	63

厦门大学博硕士学位论文摘要库

第一章 绪论

1.1 商业智能概念及发展

1.1.1 商业智能概念

随着信息技术和通讯技术的迅速发展，现代企业在实施信息化的过程中累计了非常大的业务数据。这些数据隐含着大量的商业信息，对指导企业运作与发展有着巨大的潜在价值。然而传统的管理信息系统（MIS）只对数据进行简单的查询、统计、存储，不涉及其潜在信息的发掘和提取，以致数据没有得到充分利用^[1-3]。

时至今日，在企业经营活动产生的海量数据面前，单纯依靠人力从中快速分析出有用的信息和知识已经变得不切实际；而现今高速发展变化的商业环境，要求企业能快速的把握时机，并做出正确的反应。企业如何准确、快速地对各种事件做出反应成为商业数据分析研究的重点和难点。

商业智能（Business Intelligence，简称 BI）是对商业信息进行搜集、管理和分析的过程^[4]。商业智能概念最早是由 Gartner Group 公司的 Howard Dresner 于 1989 年提出，他描述了一系列的概念和方法，通过应用基于事实的支撑系统辅助商业决策的制定。

商业智能是一种多系统协作技术，它是数据挖掘、数据仓库、OLAP 的有机结合^[1]。商业智能系统以企业中的历史数据为基础，通过数据仓库、联机分析处理、数据挖掘等技术手段，加上决策和分析人员的专业知识，得出信息和知识，从而辅助企业快速做出正确决策。商业智能系统的出现，极大地方便了企业决策者们从海量数据中获取有用的信息。

商业智能也能帮助操作层次人员对数据进行分析处理并且进行数据挖掘，例如对某些数据的发展进行统计和预测、客户和商品的分类和配对、制定促销方案等。

对商业智能进行较为深入的研究，是具有重大意义的：一、推进企业各个运转数据系统之间的协同和融合；二、提供多种统计分析手段帮助企业分析人员及时了解和观察企业业务发展情况；三、提供信息与知识支撑决策活动，帮助决策者制定企业的各个生命周期中的发展计划，对企业决策所带来的商业价值进行及时评估。

Robert Grossman 认为，商业智能软件经历了或即将经历的时代有四个^[6-9]，如表 1-1 所示：

表 1-1 商业智能软件经历的 4 个时代

代	特征	挖掘算法	集成	模式	数据模型
1	独立的应用	一个或多个算法	独立的系统	单独机器	向量数据
2	集成数据库、数据仓库	多种算法，海量数据挖掘	数据管理系统	同质的小范围计算机群	部分系统支持媒体数据库
3	集成语言模型系统	多种算法，海量数据挖掘	数据管理系统	Intranet/extranet 组成计算网络	支持半结构化和 Web 数据
4	联合多种设备的数据	多种算法，海量数据挖掘	数据管理系统	移动、电子等多种设备协作	普遍存在的数据模型

1.1.2 商业智能软件研究现状

围绕商业智能，国内外有不少提供不同类型产品的企业。专注于为 BI 系统提供基础支撑平台（数据库、数据仓库、OLAP 服务器等），如 Microsoft, IBM, Oracle, NCR, Hyperion 等；专注于提供 BI 应用组件及决策支持平台，如 Business Objects, Cognos（已被 IBM 收购），Brio（已被 Hyperion 收购），Microstrategy 等^[6]。目前国产商业智能软件平台比较有代表性的有：用友华表，浪潮的 BI 产品 myGS-BI V3.0（主要用户：中国兵器、天士力、中国铁工），博易智软（北京）技术有限公司 BI-Pilot（基于 JAVA 技术研发的企业查询、报表和在线分析（OLAP）的工具包），清华同方公司的 ezBI 2.0（同方易众中间件的一部分，有地税、商业银行、审计等方面的应用）^[10]。

虽然已众多企业从事 BI 系统开发，开发出多种 BI 系统，但是 BI 系统依旧存在一些问题^[11-12]：

- (1) BI 平台使用复杂。目前的 BI 系统都过于专业化，操作步骤复杂，操

作人员需要较强的专业知识。中小企业很难找到专业人员使用 BI 系统分析数据。

(2) 追求通用性而降低了功能和效率。国外的 BI 系统大而全，并没有针对某一具体行业做相应的优化解决方案，从而导致整个 BI 系统运转效率不高。这也是很少有中小企业使用 BI 系统原因之一。

(3) 不能跨平台。各 BI 厂商提供的商业智能解决方案都是基于自己的数据库、中间件等，不能与其他系统兼容。而很多公司在多年经营下，积累的数据不一定可以迁移到其他 BI 厂商的数据系统中去。

(4) 移动办公、安全性不高，BI 系统不可避免的涉及到公司内部数据查询等操作，容易造成泄密。现今系统的操作都是基于本地程序，为了安全性不支持移动办公。

因此，如何针对某一个行业，设计出一个满足实际需要、功能全面、简单易用的 BI 系统依然是 BI 系统研究工作中的重中之重。

1.1.3 商业智能领域发展趋势

(1) 标准化

标准化是 Oracle、IBM、微软等主流 BI 公司力推的概念。现代企业信息化水平较高，同一个企业内部就可能存在多种信息化系统，这些系统之间又大多兼容性不高。因此，为了解决 BI 平台软件的相互不兼容的问题，主流 BI 软件公司或在标准组织的协调下制定 BI 技术的标准化。

(2) 智能化

商业智能发展到如今，用户对数据可视化的要求从传统的数据图表上升到面向特定分析型主题的数据可视化。所以，建立基于数据挖掘、神经网络等技术之上的 BI 可视化技术将是未来发展的方向^[9]。

(3) 集成化

随着企业信息门户功能的不断增强，要求对服务集成的要求不断提升，将商业智能和企业信息门户进行集成，实现对数据的统一管理，这已经成为商业智能可视化展示的重要策略之一^[13-15]。强调商业智能向业务决策分析转变，预置分析应用包。如 2004 年左右国内出现的智泽华财务报表分析工具及基于商业智能技术的毕盛财务分析软件、毕盛业务分析系统等都提供了面向中小企业的决策支持软件。

(4) 采取 SaaS 商业模式

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.