

学校编码: 10384

分类号_____密级

学号: X2009230334

UDC_____

厦 门 大 学

工 程 硕 士 学 位 论 文

面向生产计划的数据清洗和挖掘系统
设计与实现

Design and Implementation of Data Cleaning and Mining
System Oriented Production Planning

林 峰

指导教师姓名: 曾 文 华 教 授

专 业 名 称: 软 件 工 程

论文提交日期: 2014 年 6 月

论文答辩日期: 2014 年 7 月

学位授予日期: 2014 年 9 月

指 导 教 师: _____

答 辩 委 员 会 主 席: _____

2014 年 6 月

面向生产计划的数据清洗和挖掘系统设计与实现

林 峰

指导教师 曾文华 教授

厦门大学

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

在企业信息化大潮流的今天，信息是现代企业的重要资源，是企业运用科学管理、决策分析的基础。所以在生产计划中构建订单生产状态跟踪的数据仓库，并通过数据挖掘技术，分析得到的订单延误的知识模型对于贸易企业是十分重要的财富，也可以为评估各个供应商产能等综合信息提供数据支持。

目前国内对鞋子的生产流程跟踪还未能做到全自动的数据采集，本课题将致力于打造可以让客户可视化的查看订单备料、裁剪、针车、成型等生产状态相关信息的系统平台。

本文将以前电信行业中对通信网关的实时数据清洗引擎方案为基础，应用在鞋类制造行业中对产品生产周期状态数据的实时采集中。并对数据内部业务关联进行分析，得到对订单潜在延误的提前预警，以及提供各个不同工厂生产产能压力的汇总报告。本课题的主要研究内容有：

- 1) 实现对不同数据源的原始数据进行数据采集转化；
- 2) 实现对采集数据源的字段、格式、数据类型进行灵活配置；
- 3) 实现对数据内容有效的自动验证，避免采集脏数据；
- 4) 实现对采集数据与数据之间的业务逻辑有效性的灵活配置；
- 5) 实现自动异常纠错机制，提高可靠性；
- 6) 根据业务分析进行创建数据仓库，并进行信息知识挖掘。

本文以软件工程的思想为主线，从业务需求分析和定义、系统体系架构设计、系统流程设计、系统模块设计以及数据库详细设计等方面介绍了系统的设计和实现过程。最终成功的将电信网关数据采集方案迁移到本系统中，并达到了用户的预期目的和效果。

关 键 字： 数 据 清 洗 ； 数 据 挖 掘 ； 生 产 计 划

Abstract

In today's trend of enterprise information, information is the important resources in the modern enterprise, is the foundation of the enterprise with scientific management and decision analysis. In production plan to build the PO status tracking of data warehouse, and through the analysis of the data mining technology to get the knowledge of the order delay model is very important for trade enterprise's wealth. It can also provide the comprehensive information to assess the supplier capacity, etc data support.

The current domestic production of shoes process state data also failed to achieve automatic data cleansing, this topic is committed to building can provided customers visual view order material preparation, cutting, stitching, lasting and other production status information system platform.

This article will be in the telecommunications industry to communication gateway based on real-time data cleaning engines, production cycle of a product used in the footwear manufacturing industry in the state of real-time data acquisition. And analyze the internal business data correlation, we get the order early warning of potential delay, and to provide various factory production capacity pressure summary report. This topic research content includes:

- Implementation of different sources of data collection into the original data;
- To realize the collection of data fields and format, data types, flexible configuration;
- Implementation of the data content and effective automatic validation to avoid dirty data;
- To realize the collection of data and the data between the business logic of the effectiveness of flexible configuration;

- Automatic exception handling mechanism to improve reliability;
- Create a data warehouse according to the business analysis, and knowledge of information mining.

Based on the ideas of software engineering as the main line, from the requirement research and analysis, architecture design, function module design and program flow design, database detailed design introduces the realization process of the system. Finally successful moving telecommunication gateway data acquisition scheme to the system, and achieve the intended purpose and effect of the user.

Keywords: Data Cleansing; Data Mining; Production Plan

目 录

第一章	引言	1
1.1	研究背景和意义	1
1.2	研究现状和问题	2
1.3	研究内容和特色	2
1.4	本文的组织结构	3
第二章	相关技术介绍	4
2.1	数据清洗概述	4
2.2	数据仓库简介	5
2.2.1	数据仓库的定义	5
2.2.2	数据仓库的特点	6
2.2.3	数据仓库体系结构	7
2.2.4	实时数据仓库	7
2.3	数据挖掘概念与技术	8
2.4	相关开源框架	9
2.4.1	Struts2.x 框架	9
2.4.2	Spring 框架	10
2.4.3	Hibernate 框架	12
2.4.4	开源工具库介绍	13
2.5	本章小结	14
第三章	系统需求分析	15
3.1	业务需求分析	15
3.2	功能需求分析	15
3.3	定制报表分析	21

3.3.1	订单生产状态可视化	21
3.3.2	订单潜在延误分析报表	21
3.3.3	供应商产能分析报表	22
3.4	非功能性需求分析	23
3.5	本章小结	24
第四章	系统总体设计	25
4.1	系统架构设计	25
4.2	系统功能模块设计	28
4.3	系统流程设计	29
4.4	数据库设计	32
4.4.1	数据库选取	32
4.4.2	数据库设计工具	32
4.4.3	主要表结构	33
4.4.4	平台数据库设计	40
4.4.5	业务数据库设计	41
4.4.6	数据仓库设计	42
4.5	本章小结	47
第五章	系统详细设计与实现	48
5.1	系统开发和部署环境	48
5.2	关键技术实现	49
5.2.1	数据清洗	49
5.2.2	系统平台	53
5.2.3	数据分析	55
5.3	功能界面展现	56
5.3.1	数据清洗	56
5.3.2	系统平台	58

5.3.3	数据分析	59
5.4	系统测试	61
5.4.1	测试环境	61
5.4.2	测试用例	62
5.5	本章小结	63
第六章	总结与展望	64
6.1	总结	64
6.2	展望	64
参考文献	66
致 谢	68

厦门大学博硕士论文摘要库

Contents

Chapter 1 Introduction	1
1.1 Background and Significance	1
1.2 Research Status and Problems	2
1.3 Main Content and Characteristic	2
1.4 Organizational Structure	3
Chapter 2 Related Technology Introduction	4
2.1 Data Cleansing	4
2.2 Data Warehouse	5
2.2.1 Data Warehouse Definition	5
2.2.2 Data Warehouse Characteristic.....	6
2.2.3 Data Warehouse Architecture	7
2.2.4 Realtime Data Warehouse	7
2.3 Data Mining Concepts And Techniques	8
2.4 Research Open Source Framework	9
2.4.1 Struts2.x Framework	9
2.4.2 Spring Framework	10
2.4.3 Hibernate Framework	12
2.4.4 Open Source Tools Library Introduction.....	13
2.5 Summary	14
Chapter 3 System Requirements Analysis	15
3.1 Business Requirements Analysis	15
3.2 Functional Requirements Analysis	15
3.3 Costom Reports Analysis	21

3.3.1	Production Status Visibility.....	21
3.3.2	Potential Delay Reports Analysis	21
3.3.3	Supplier Capacity Reports Analysis	22
3.4	Performance Requirements.....	23
3.5	Summary	24
 Chapter 4 System Design.		 25
4.1	System Architecture Design	25
4.2	System Functional Modules Design	28
4.3	System Functional Flow Chart Design	29
4.4	Database Design.....	32
4.4.1	Select The Database.....	32
4.4.2	Database Design Tool	32
4.4.3	Main Tables.....	33
4.4.4	Platform Database Design	40
4.4.5	Business Database Design	41
4.4.6	Data Warehouse Design.....	42
4.5	Summary	47
 Chapter 5 Detailed Design and Implementation		 48
5.1	Development Environment and Tools.....	48
5.2	Implement of Key Technology	49
5.2.1	Data Cleansing.....	49
5.2.2	System Platfrom	53
5.2.3	Data Mining.....	55
5.3	Show of Function Interface	56
5.3.1	Data Cleansing	56
5.3.2	System Platfrom	58

5.3.3 Data Mining.....	59
5.4 System Test.....	61
5.4.1 Testing Environment.....	61
5.4.2 Test Case.....	62
5.5 Summary.....	63
Chapter 6 Conclusions And Outlook	64
6.1 Conclusions	64
6.2 Outlook.....	64
References	66
Acknowledgements	68

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