

数据仓库在口岸物流中的应用

吴先斌, 林国龙, 杨 斌, 王 洁

(上海海事大学, 上海 200135)

摘要:随着信息技术的发展,口岸物流活动中积累的信息量正以指数级增长,到目前已经积累了海量数据,但这些数据存在数据库中对决策支持作用很小,以数据分析支持口岸决策提高口岸工作效率已成为迫切需要。为满足上述要求,首次将数据仓库技术应用于口岸物流,并详细介绍了口岸物流数据仓库的体系结构,口岸数据仓库的维度建模方法,并用 Oracle 强大的前端展示工具实现 OLAP 分析、定制报表等,为用户提供决策支持。通过数据仓库在口岸物流中的实际应用证实了数据仓库能很好地应用于口岸物流,对于提高口岸物流效率,降低运营成本具有积极意义。

关键词:数据仓库;口岸物流;Oracle

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2009)12-0233-03

Data Warehouse Application in Port Logistics

WU Xian-bin, LIN Guo-long, YANG Bin, WANG Jie

(Shanghai Maritime University, Shanghai 200135, China)

Abstract: With the development of information technology, the accumulation of information in logistics activities is exponential growth, to the present it has accumulated huge amount of data. However the data in database can't make decision-making support. Increasing the efficiency of ports, supporting decision-making thought data analysis have become an urgent need. In order to meet these requirement, apply data warehouse technology in the port logistics for the first time. Introduce architecture of the data warehouse for port logistics, dimensions modeling, and apply powerful front-end display tools realize OLAP analysis, customized statements etc. to provide users with decision-making. Through practical application of data warehouse in port logistics, data warehouse confirmed can be used in port logistics and have positive significance of improve efficiency, reduce operating costs in port logistics.

Key words: data warehouse; port logistics; Oracle

0 引言

口岸物流是指利用口岸货物集散的优势,以先进的物流服务基础设施、设备为依托,以进出口贸易和转口贸易为支撑,以现代信息技术为手段,以优化物流资源整合为目标,强化口岸周边物流辐射功能的综合物流形态。随着数据库的成熟和信息技术应用的普及,口岸物流活动中积累的信息量正在以指数级增长,到目前为止已经积累了海量的数据,但是这些数据存在数据库中,对决策的支持作用微乎其微。将数据仓库技术运用到口岸实际业务中,以数据分析支持口岸决策,提高口岸工作效率已成为迫切的需要。

数据仓库技术在我国还处于起步阶段,大多处于理论研究阶段。虽然在金融、电信等领域有一些比较成功的案例,但有关口岸的数据仓库的设计方面还处于起步阶段,相关的研究成果还几乎没有。大连理工大学王宇、邢郁达对港口生产数据仓库进行了研究^[1],但是是建立在港口信息系统基础之上,与文中基于口岸物流流程、着重研究口岸通关效率在分析的主题上有很大的不同。因此首先文中的意义在于所研究的口岸数据仓库设计能为今后口岸及相关的数据库设计起到抛砖引玉的作用。其次在实际应用方面:通过设计数据仓库和 OLAP 分析可以定量、科学地评价口岸服务商(报关单位,货代)的服务绩效。由于口岸物流流程存在不确定性,因此容易造成成本黑洞,通过数据挖掘能够找到口岸业务链上的“成本黑洞”。通过口岸数据仓库环境下的数据挖掘和 OLAP 分析能够发现货主业务量的变化情况,预测业务量的变化趋势以及货量的分布情况。

文中在熟悉口岸业务流程的前提下,将口岸业务

收稿日期:2009-04-01;修回日期:2009-07-08

基金项目:上海市科委能力建设项目(071705107)

作者简介:吴先斌(1984-),男,硕士研究生,研究方向为数据仓库与数据挖掘;林国龙,教授,研究方向为数据仓库与数据挖掘;杨斌,副教授,研究方向为数据仓库与数据挖掘;王洁,讲师,高级工程师,研究方向为数据仓库与数据挖掘。

划分为海运进口、海运出口、空运进口、空运出口四个主题,建立数据仓库针对不同的口岸服务商感兴趣的模式进行多维数据分析和数据挖掘。口岸相关单位提高口岸服务效率、降低服务成本、提供决策支持将是一件非常有意义的工作。

1 口岸数据仓库

1.1 数据仓库概念

数据仓库目前最权威的定义是被称为数据仓库之父的 William H. Inmon 在《building the warehouse》一书中给出的,他认为数据仓库是一个面向主题的 (SubjectOriented)、集成的 (Integrated)、非易失的 (Non-volatile),且随时间变化 (Time - variant) 的数据集合,用来支持管理人员的决策分析^[2]。对于数据仓库的概念,可以从两个层次理解:首先,建设数据仓库系统的目的在于支持决策,面向分析型数据处理,它不同于企业现有的操作型数据库;其次,数据仓库是对多个异构的数据源有效集成,集成后按照主题进行重组,并包含历史数据,而且存放在数据仓库中的数据一般不再修改。根据上述数据仓库的概念可以总结出数据仓库具有以下特点:面向主题性、集成性、非易失性、时变性、集合性^[3]。

1.2 口岸数据仓库体系结构

数据仓库从多个信息源获取数据,经过加工、整理,存放在数据仓库的内部数据库中,通过数据仓库的访问工具,给用户提供一个集成的、能对数据进行综合分析的、提供决策信息的环境。数据仓库体系结构由数据源、数据预处理系统、数据仓库、数据分析工具四部分组成。口岸数据仓库体系结构如图 1 所示^[4]。

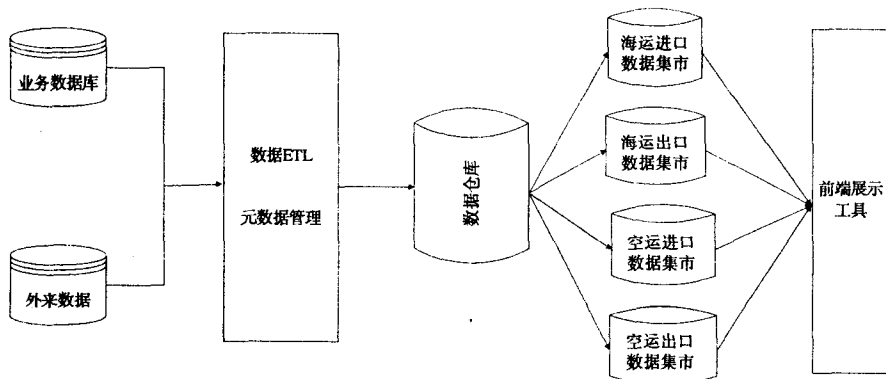


图 1 口岸物流数据仓库体系结构

2 口岸数据仓库的设计方法

2.1 口岸数据仓库实现方法

口岸数据仓库的实现方法一般有两种^[5]:

自顶向下:由 W. H. Inmon 提出的一种数据仓库

结构,该结构的开发策略是把原来分散存储在企业各处的 OLTP 数据库中的有用的数据通过提取、清理、转换聚集等处理建立一个全局数据仓库,提供给用户一致的数据格式,一致的软件环境。

自底向上:是从建立各个部门的数据集市开始,全局的数据仓库是建立在各个部门的数据集市之上。无需等待安置好更大业务范围的数据仓库设计,这并不意味着不会开发更大业务范围的数据仓库设计,随着数据仓库实现的扩展,将逐渐增加对它的构建。

口岸数据仓库的特点是:主题比较固定,包括海运进口、海运出口、空运进口、空运出口,而且四个主题的维度比较一致,采用自顶向下的模式可以共享元数据,实现数据和数据的抽取、清洗、转换的统一管理,因此口岸数据仓库比较适合采用自顶向下模式。

2.2 概念模型设计

概念模型通常可以称作是需求分析,也就是在与用户交流的过程中可以根据下面的问题确定系统的边界:(1)数据库中有什么样的数据;(2)决策者感兴趣的问题是什么;(3)解决这些问题需要什么样的信息;(4)这些信息对应于数据库中的哪些数据。

在口岸数据仓库概念模型设计时,以海运出口为例:通过对决策者的需求分析了解到决策者希望其通关的各个阶段所用的时间情况(从货主送货到预录入的时间,现场报关到现场放行、预录入到现场放行、现场放行到结关、从送货到现场放行的总时间,从送货结束到结关的总时间,预录入的总时间、查验时间、预录入到查验放行、从查验放行到现场放行、现场报关时间、预录入重审次数等),还需要了解其客户的分布情况、查验情况等。然后根据这些问题确定信息的来源,

以及这些信息对应数据仓库系统中的那些数据部分。

在确定系统边界之后,接下来就是确定系统的主题,并对每个主题做较明确的描述。通过对原有数据库系统的分析,考虑到决策者的需求在系统边界的基础上确定了口岸数据仓库的四个基本主题:海运进口、海运出口、空运进口、空运出口。

2.3 逻辑模型设计

数据仓库逻辑建模方法也有两种方法:实体关系建模和维度建模。实体关系建模技术是一种设法消除冗余的一种逻辑设计技术;维度建模技术是一种逻辑设计技术,该技术试图采用某种直观的标准技术框架

来表现数据,并允许进行高性能的存储^[6]。它必然会遵循维度方面的规范,并且坚持带有某些重要条件的关系模型规范,在模型的建立上每个模型都是由具有复合键的某个表(事实表)和一系列小型表(维度表)组成。维度建模拥有数据仓库方面的很多重要优点,而这些优点正是实体关系模型所缺少的。首先维度建模是可预测的标准框架。报告写作程序、查询工具以及用户界面为维度模型提供了强大的前提条件,使用户的理解性更强,处理过程更有效;维度建模的第二个优点是星型连接模式的可预测框架能够忍受不可预知的用户行为变化。每个维度都是相等的,所有维度都可以看作是对称相等的事实表入口点;维度建模的第三个优点是具有非常好的扩展性,以便容纳不可预知的新数据源和新的设计决策;第四个优点是:有很多标准的方法可以用来处理业务环境中的常见的建模情况。

文中遵循 Kimball 的方法采用维度建模,并在最低粒度的数据建模上建模,这样设计的模型可以满足各种未知需求的查询。维度建模有星型模式和雪花模式两种模式。

(1)星型模式:星型关系是一种多维的数据关系。有关维的信息存储在一个或若干个表中,这些表称为维表。维所修饰的数据存储在一个或者若干个表中,这些表称为事实表,事实表和维度表一起构成了星型模式,事实表处在这个星型模式的中间,而维表则分布在事实表的周围。星型模式包含两类表:事实表和维度表。事实表:事实表是维度建模中的基本表,用于存储事实的度量值以及各个为的键。维度表:用于存储维的键值和和维度的描述信息。

星型模式优点:结构清晰简单,利于用户的理解和信息技术人员管理,表间连接关系清晰。缺点是维表结构非规范化,造成数据冗余,可操作的灵活性低。

(2)雪花模式:雪花模式是对星型模式的扩展,使得维表进一步规范化、层次化。星型模式中的维表是非规范化的,当维度具有多层次时维度表具有冗余数据。为了节约空间星型模式用多张表来描述一个多层次的维,避免了数据冗余。雪花模式的优点是减少了数据冗余,对表做了规范化处理。但雪花模式增加了用户必须处理的数量,增加了查询的复杂度^[7]。

尽管规范的雪花模式中减少了冗余,但需要更多的连接,笔者在建立口岸数据仓库时为了方便用户查询和信息技术人员的管理,采用星型模式。图 2 为海运出口星型模式。

2.4 物理模型设计

根据逻辑模型设计阶段的星型模式能够方便地定义物理数据结构。一般将指标实体转化为物理数据库

表,称为事实表。事实表首先包括星型模式中心的指标量,其次应包括星型模式角上的维度实体中层次最低单位的主键。维度实体通常也转化为维数据库表,称为维表,它包括其每一层次的主键和对应的值。维表的关键字是该维度实体对应的详细类别实体的主键。维表和事实表通过维表关键字相关联。口岸数据仓库物理模型的设计中除了上述物理数据结构还包括定义数据标准、定义实体、确定实体特征、确定数据容量和更新频率、确定索引策略,提高数据仓库性能^[8]。

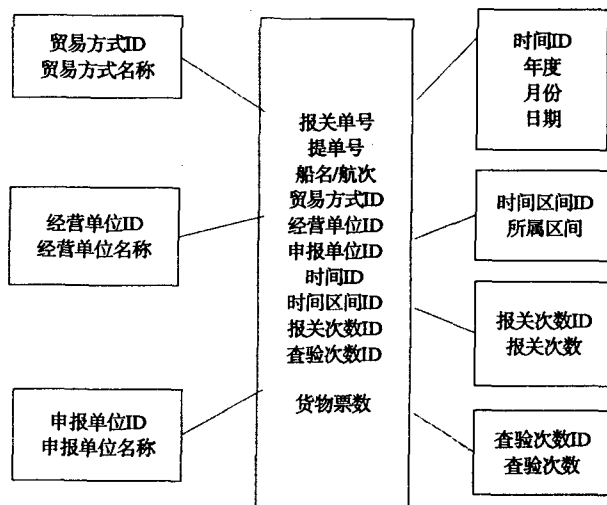


图 2 海运出口星型模式

3 口岸数据仓库前端展示

Oracle 数据仓库提供了数据仓库的全面支持,在数据仓库前端展示方面我们介绍一种 Oracle 集成工具—Oracle Discover。它是为数据仓库和联机事务处理系统而设计的即席查询和在线报表工具。Oracle Discover 是具有世界领先水平的简单易用用户界面,独特的查询时间预测功能、汇总管理等功能^[9]。

利用 Oracle Discover 进行前端展示的基本步骤:

- 1) 建立数据库用户并授权 CREATE USER DISCOVER IDENTIFIED BY DISCOVER DEFAULT TABLESPACE test TEMPORARY TABLESPACE TEMP GRANT CONNECT, RESOURCE TO DISCOVER;
- 2) 建立 EUL: 创建 EUL 有两种方式,一种是基于应用用户,一种是基于数据库用户;
- 3) 创建业务区;
- 4) 建立文件夹;
- 5) 给用户授权。

4 结束语

Oracle 提供了非常成熟的数据仓库解决的方案,

(下转第 239 页)

属中断方式,系统功能均在中断服务子程序中完成。主程序首先进行初始化,包括 I/O 口、定时器、中断系统的初始化^[8],然后判断功能键 SET 是否被按下以及按下的时间,确定系统工作模式,最后等待定时器中断。在定时器中断服务子程序中,完成检测温度,刷新显示温度,进行模糊控制并输出控制结果等功能,并对电路故障发出报警^[9]。

4 系统的测量结果

在此控制系统中,可以根据需要修改温度设定值,不同的温度设定值,其对应的温度测量值最后也会相应地稳定在某个范围内。

此系统取模糊控制周期为 6 秒,其温度测量值变化范围和温度设定值的关系如表 3 所示。在恒温阶段,控制精度达到±2℃。

表 3 系统测量结果表

温度设定值	温度测量值范围
50℃	49℃~51℃
100℃	99℃~101℃
150℃	148℃~151℃
200℃	198℃~202℃
250℃	248℃~252℃
300℃	299℃~302℃
350℃	348℃~352℃
400℃	398℃~402℃

(上接第 235 页)

文中介绍的 Oracle 集成工具能够即席查询和在线报表,对提高口岸的服务水平,提高口岸效率,降低口岸运营成本具有积极意义,这也使得数据仓库技术首次应用于口岸物流就取得了比较好的效果。但口岸物流流程十分复杂,而且流程具有不确定性,要计算整个流程中各个阶段的时间效率需要对时间做大量的处理,而且有关时间的处理也是一件十分复杂的事情,因此口岸数据仓库的设计需要经过较长的时间才能慢慢完善,而且要达到智能管理系统的程度还有很多工作需要去做。

参考文献:

[1] 邢郁达.港口生产数据仓库设计研究[D].大连:大连理工大学,2007.

[2] Inmon W H. Building the Data Warehouse[M]. [s. l.]: John Wiley & Sons, Inc, 2003.

5 结束语

文中介绍了用模糊控制的电阻炉温度控制系统,实践表明,模糊控制方法提高了控制的实时性、稳定性和精确度,并且实现了操作过程的简化。目前,该系统处于实际工业电阻炉温度控制的应用和试行阶段,取得了良好的效果。

参考文献:

[1] 刘喜梅,王波,于飞.工业电炉温度的模糊控制系统设计[J].微计算机信息,2007(25):73-75.

[2] 李红岩,侯媛彬,王秀.实现温度自动调节的模糊控制器设计[J].计算机技术与发展,2006,16(3):149-151.

[3] 余永权,曾碧.单片机模糊逻辑控制[M].北京:北京航空航天大学出版社,1995.

[4] WU Zhiqiao, Mizumotom. PID - type Fuzzy Controller and Parameters Adaptive Method[J]. Fuzzy Set and Systems, 1996,78(1):23-26.

[5] King P J, Mamdani E H. The application of fuzzy control systems to industrial process[J]. Automatic, 1997(3):235-242.

[6] 何立民.单片机应用系统设计[M].北京:北京航空航天大学出版社,1990.

[7] 吴金戎,沈庆阳,郭庭吉.8051 单片机实践与应用[M].北京:清华大学出版社,2002.

[8] Narendre K S, Muldaoppadhyay S. Adaptive control of non, near multivariable systems using neural networks[J]. Neural Networks, 1999,7(5):737-752.

[9] 赵世涛,马莉,王现君,等.一种自适应的模糊关联规则挖掘算法[J].计算机技术与发展,2008,18(5):64-66.

[3] 王珊.数据仓库技术与联机分析处理[M].北京:科学出版社,1998.

[4] Kimball R, Reeves L, Ross M. The Data Warehouse Lifecycle Toolkit: Export Methods for Designing, Developing, and Deploying Data Warehouse[M]. 北京:电子工业出版社,2003.

[5] 黄玉明,毛宇光.数据仓库中粒度划分的层次编码解决方案[J].计算机技术与发展,2008,18(10):1-4.

[6] 周蓝桢,周肆清,杨炼.数据仓库技术在医院病情诊疗分析中的应用研究[J].计算机技术与发展,2008,18(8):230-232.

[7] 陈京民.数据仓库与数据挖掘技术[M].北京:电子工业出版社,2000.

[8] 王秀娟,曹宝香.基于面向对象原型法的 N 层数据仓库设计[J].计算机技术与发展 2009,19(1):117-120.

[9] Oracle Database 11g for Data Warehousing and Business Intelligence[M]. Oracle white paper, 2007.