

生产质量控制数据仓库模型设计与实现

王晓鹏, 武 彤

(贵州大学 计算机科学与信息学院, 贵州 贵阳 550025)

摘 要: 为了充分利用已建立的电视机生产线质量控制信息采集系统所采集的大量基层数据, 给高层管理人员提供更多的决策建议, 提出了电视机生产线质量控制信息分析系统的解决方案。首先针对电视机生产线质量控制管理领域的特点, 结合目前主要的数据仓库的设计理论, 提出了一种电视机生产线质量控制领域数据仓库模型, 然后详细阐述了该模型的设计、实现过程。通过实现电视机生产质量控制数据仓库, 为进一步实现电视机生产线质量控制信息分析系统打下了一定的基础。

关键词: 生产质量控制; 数据仓库; 决策支持; 联机分析处理; 数据挖掘

中图分类号: TP311

文献标识码: A

文章编号: 1673-629X(2015)06-0181-04

doi: 10.3969/j.issn.1673-629X.2015.06.040

Design and Realization of Data Warehouse Model on Production Quality Control

WANG Xiao-peng, WU Tong

(School of Computer Science and Information, Guizhou University, Guiyang 550025, China)

Abstract: In order to make full use of a large number of basic data have been established in TV production line quality control information collection system, providing more decision-making suggestions to senior management, a TV production line quality control information analysis system is presented. Firstly, according to the characteristics of quality control management in the field of TV production line, combined with the current design theory of data warehouse mainly, put forward a kind of TV production line quality control field data warehouse model. Then, the model processing of design and implementation is described in detail. Finally, by the establishment of data warehouse on TV production quality control system, laid a foundation for the further implementation of the TV production line quality control information analysis system.

Key words: production quality control; data warehouse; decision-making support; OLAP; data mining

0 引言

随着市场竞争的日趋激烈, 信息对于企业的生存和发展发挥着越来越重要的作用。由于计算机技术的普遍应用, 承载信息的数据随着时间的推移而不断增长, 并且分布在不同的系统平台上, 具有多种存储形式, 这使得企业难以从巨量的数据中找到真正有用的决策信息^[1]。因此企业迫切需要新的技术和方法对原有的信息进行提炼和加工, 需要为企业领导提供集成化和历史化的数据, 需要为企业全局的战略决策和长期趋势分析提供更有力的支持。

目前某电视机厂已初步建立基于 B/S 结构的电

视机生产线各生产环节质量控制信息的采集系统。在该系统中, 生产现场的工人通过各质量检查及控制信息采集点将产品质量控制信息录入到数据库中, 企业管理人员可以对实时采集的现场数据进行各种处理, 如查询、统计、汇总等。生产线质量控制信息采集系统代替了传统的基于手工操作的电视机生产线质量监控系统, 实现了“无纸化管理”。但是该采集系统是一个 OLTP 环境, 只能对采集到的信息进行简单的统计和汇总, 其有关生产线质量信息的分析和生产决策支持方面的功能比较薄弱。为了能够支持管理层的科学决策, 在现有采集系统基础上建立电视机生产线质

收稿日期: 2014-07-16

修回日期: 2014-10-20

网络出版时间: 2015-05-06

基金项目: 贵州省 2010 年工业攻关项目 (黔科合 GY 字 [2010] 3061); 2011 年贵州大学校立项目 (X102007)

作者简介: 王晓鹏 (1978-), 男, 硕士, 讲师, 研究方向为数据仓库技术、信息安全技术; 武 彤, 硕士, 副教授, 研究方向为数据仓库技术、OLAP、数据挖掘。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20150506.1627.020.html>

量控制信息数据仓库系统有着非常重要的意义。该系统的结构图如图 1 所示。数据仓库模型设计是建立数

据仓库中的关键^[2],文中将对数据仓库模型的设计进行阐述。

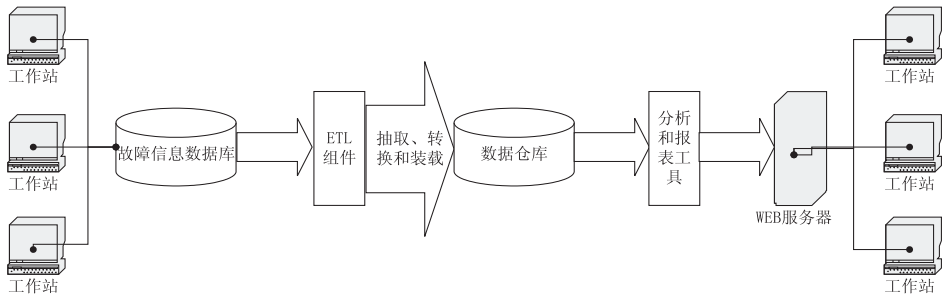


图 1 电视机生产线质量控制信息数据仓库系统结构

1 数据仓库模型的设计

数据仓库模型设计是建立数据仓库的基础,模型的优劣决定着数据仓库性能。由于之前建立的电视机生产线各生产环节质量控制信息的采集系统是基于 B/S 结构的,所以在电视机生产线质量控制信息数据仓库的设计上也考虑使用 Web 技术,利用 Web 实现分析数据的获取和展现。通过 Web 技术使信息分布和处理更加高效、更加经济^[3]。基于 Web 的数据仓库技术是一种信息集成技术^[4],就是从现有的质量信息数据库中获取原始数据,经过加工整理后,将数据存储在数据仓库的内部数据库中,最后通过 ASP.NET、Silverlight 等 Web 技术向终端用户提供分析数据,以支持管理层的决策。数据仓库模型的设计大体上分为概念模型设计、逻辑模型设计和物理模型设计三个过程。通过三个过程逐层细化,得到精炼的数据仓库模型^[5]。

1.1 概念模型设计

1.1.1 确定系统边界

概念模型的设计首先要界定系统边界^[6],界定系统边界的工作就是根据数据仓库设计的需求分析划定一个大致系统边界。电视机生产线质量分析数据仓库系统主要是实现产品直通率的有效提高及产品返修率的大幅降低,同时辅助高层决策人员根据科学的依据制定管理决策。

(1)对机型故障现象进行分析。根据时间、机型、故障现象和产品类型等对故障数目进行分析。

(2)对机型直通率进行分析。根据时间、机型和产品类型等对直通率进行分析。

(3)对生产线进行分析。根据时间和生产线等对故障数目进行分析。

1.1.2 确定主题域

在确定系统边界后需要确定主题域。由前期的用户需求以及此系统的总体功能要求可知:选取机型直通率分析、生产线分析和机型故障现象分析三个主题来考虑进行 OLAP 分析及挖掘。根据确定的主题给出

面向主题的信息包图,其中机型直通率主题的分析信息包图如图 2 所示。

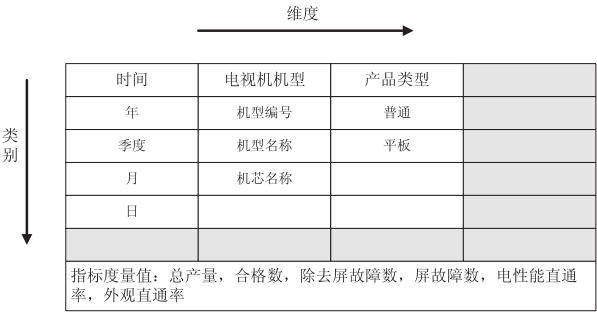


图 2 机型直通率分析信息包图

1.2 逻辑模型设计

1.2.1 数据组织结构的形式

逻辑模型主要描述与业务有关的重要实体及相互之间的关系。数据仓库的基本实体主要由事实表和维表构成。事实表是数据组织结构中最重要的表,它直接反映了数据仓库应用的主题,也包含了数据仓库中最基本、最主要的信息。维表用来描述事实表,事实表中大多数属性都与维表相关^[7]。

针对机型直通率主题,设定机型直通率分析事实表包括采集时间、机型、产品类型、总产量、合格数、非屏坏的故障数、屏故障数、电性能直通率和外观直通率,其中采集时间、机型和产品类型为指向维表的外键。设定采集时间维、机型维和产品类型维 3 个维度表。

目前,数据组织结构的形式主要有星型模型、雪花模型和事实星座模型等^[8]。星型模型以数据仓库用户最为关心的事实为中心,将数据仓库连接组合为一个整体^[9],能清晰地反映概念模型中各种实体间的逻辑关系,并且设计方法简单,易于被用户理解,并且星型模型适于查询^[10]。因此采用星型模型结构来设计完善的数据仓库逻辑模型。事实表和维度表的星型模型具体如图 3 所示。

针对机型直通率分析主题中的维度表,相应的粒度级别划分如下:时间维分为年、季度、月和日;电视机

机型维分为机型编号、机型名称和机芯名称;产品类型维分为产品类型编号和产品类型名称。

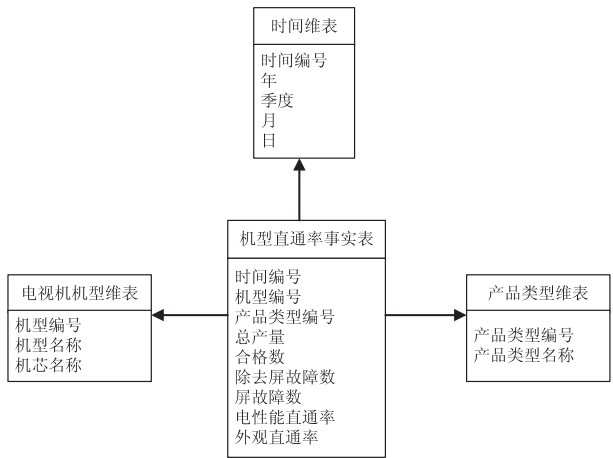


图3 机型直通率分析主题的星型模型

1.2.2 数据粒度的选择

根据系统数据量的估算值、要接受的分析类型、可接受的最低粒度和能存储数据的存储容量等因素,选取多重粒度策略。

在数据挖掘模型中,如果将复杂的算法直接在数据仓库中存放的海量数据上运行,则计算时间和空间复杂度将太高,以至于系统难以承受。如果根据一定的采集率从数据仓库中抽取一个子集,在这个子集上进行数据挖掘,则既能达到目的,又能减小问题规模,提高效率。按照变化缓慢的数据可以采用低采样率,变化快的数据采用高采样率的原则,将粒度进行划分。

1.2.3 数据分割策略

由于数据仓库中的数据量极大,数据量大使用起来会遇到很多问题,因此,对数据进行合理的分割进而进行处理是必不可少的工作。选择适当的数据分割标准,一般要考虑以下几个方面的因素:数据量的大小,数据分析处理的实际情况,简单易行,与粒度的划分策略相统一,数据的稳定性。

在生产线分析主题中,企业强调了对各个生产线上采集到的数据能进行实时监控。基于客户需求,对数据仓库的数据按生产线的不同进行划分,以便于数据分析。再根据不同时间对数据进行轻度综合,也就是按时间单位以日、月、季、年进行划分。

1.3 物理模型设计

数据仓库的物理模型就是数据仓库逻辑模型在物理系统中的实现^[11]。其中包括了逻辑模型中的各种实体表的具体化。为确定数据仓库的物理模型,必须做以下几方面工作:首先要全面了解所选用的数据库管理系统,特别是存储结构和存取方式;其次了解数据环境、数据的使用频率、使用方式、数据规模及响应时间要求等,这些都是对时间和空间效率进行平衡和优

化的重要依据;最后还需要了解外部存储设备的特征。只有这样才能在数据的存储需求与外部存储设备条件两者之间获得平衡。例如表的数据结构类型、索引策略、数据存放位置和数据存储分配等。

为数据仓库选择合适的索引策略是物理设计中的一个重要内容。因为数据仓库的数据量很大,所以需要 对数据的存取路径进行仔细的设计和选择。常用的索引策略有:B-Tree 索引、BitMAp 索引、BitWise 索引、连接索引、广义索引。

根据具体的需求要求,选择连接索引。连接索引,就是将事实表和维表中的索引项进行连接运算,然后将结果作为索引保留。由于索引项比整个记录条目要小,连接索引也比直接对事实表和维表进行连接的结果集合要小得多,因此,连接索引能够起到较好的索引作用。索引表如表1所示。

表1 索引表

索引主键	说明
timeID	时间维度主键
ProductTypeID	产品类型分析维度表主键
ProductLineID	生产线分析维度表主键
TvTypeName	机型分析维度表主键
FaultappearancesID	故障分析维度表主键

2 数据仓库的实现

2.1 数据预处理

目前大多数数据仓库都是建立在关系型数据库的基础上,最终数据的存储是由数据库系统进行管理的^[12]。在该数据仓库的设计中,选用 MS SQL Server 2008 及其组件分析服务器作为数据库和数据仓库的管理系统。利用 SQL Server 2008 中的 Business Intelligence Development Studio 中的 SQL Server Integration Services(SSIS)功能来实现 ETL,即将源 OLTP 数据库中的数据通过提取,转换,加载到新建的数据仓库中。

进行数据预处理的主要 SQL 代码如下:

```
SELECT
FaultStatistics. ProductTime ,
TVInfo. TVTypeName ,
FaultStatistics. TVName ,
FaultAppearances. Name as FaultAppearancesName ,
Faultcause. Name as faultcauseName ,
FaultStatistics. ProductLineName ,
FaultStatistics. ProductTypeName ,
FaultType. Name as FaultTypeName ,
TVType. ModuleName ,
FaultStatistics. ID
FROM
```

```

FaultStatistics INNER JOIN TVInfo ON FaultStatistics. TVName
=TVInfo . Name
INNER JOIN FaultAppearances ON FaultStatistics. FaultAp-
pearanceID =FaultAppearances. ID
INNER JOIN Faultcause ON FaultStatistics. FaultcauseID =
Faultcause. ID
INNER JOIN FaultType ON FaultAppearances. FaultTypeID =
FaultType. ID
INNER JOIN TVType ON TVInfo. TVTypeName = TVType.
Name

```

2.2 数据仓库的访问

使用 ADO MD 技术建立与数据仓库的接口。

创建接口的主要代码如下:

```

String con = " Datasource = localhost; Initial Catalog = BasicInfo;
UserName = sa; Password = 123456"

```

使用 MDX 语句对数据仓库进行各种 OLAP 操作。数据库的查询使用传统的 SQL 语句,但是用 SQL 语句完成数据仓库的操作将会非常困难和复杂。而利用 MDX 语句可以轻松实现对数据仓库的上钻、下钻、切片和旋转等操作^[13]。

如下是电视机生产线质量控制数据仓库模型的某个 MDX 语句:

```

SELECT {[ MEASURES]. [FaultAppearances]} ON
COLUMNS,
{[ TIME]. [2010]} ON ROWS
FROM [TVTypeRolled]
WHERE [TVType]. [ LED24K01 ]

```

2.3 前端展示

通过 SQL Server Reporting Services 功能将 OLAP 分析的结果显示为不同的图表和报表,最后的结果通过 B/S 方式在前端的浏览器中展示。为此,除了应用服务器外还需要安装、配置一台基于 IIS 的 Web 服务器。客户端以 HTTP 的方式向 Web 服务器提交请求,再由 Web 服务器将请求传递至应用服务器。应用服务器再将请求传递至数据仓库,数据仓库处理相应的请求后将结果传递给应用服务器,应用服务器再将结果传递至 Web 服务器,Web 服务器将结果以网页的形式返回给客户端,客户端通过浏览器查看最终的结果。实现方式如图 4 所示。

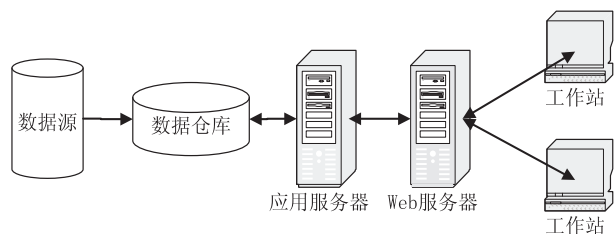


图 4 数据仓库的实现方式

在 Web 接口部分采用 ASP.NET 技术创建 Web 程序,用 C#语言作为后台的开发工具,前台使用 Silver-

light 富界面技术对数据进行更丰富的展示。

3 结束语

文中主要针对电视机生产线上的质量分析需求,基于已建立的生产线数据采集及 OLTP 分析数据库系统,提出了电视机生产线质量控制系统数据仓库的设计方案,并给出了具体的实现方法。该方案能为高层管理人员进行电视机生产质量的分析和控制提供有力的支持。下一步的具体工作将针对问题对模型进行不断的优化^[14]、建立电视机生产质量控制的数据挖掘体系,并对基于 Web 的访问体系所存在的网络安全性问题进行解决。

参考文献:

- [1] 周春成. 电信行业数据仓库的设计与实现[D]. 西安:西安电子科技大学,2009.
- [2] 侯建春,陈梦东. 干部管理领域数据仓库模型设计[J]. 微机发展(现更名:计算机技术与发展),2005,15(11):40-42.
- [3] Perez J M, Berlanga R, Aramburu M J, et al. Integrating data warehouse with web data; a survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2008, 20(7): 940-955.
- [4] Xue Hong, Liu Zaiwei, Meng Haiyang. Study and realization of supermarket BI system based on data warehouse and web technique[C]//Proc of international conference on computer science and software engineering. Wuhan: IEEE, 2008: 482-485.
- [5] Hammergran T. Data warehousing: building the corporate knowledge base[M]. North Carolina: Ventana Communication Group, 1997.
- [6] 孙水华,王加阳. 库存控制决策支持系统数据仓库模型设计[J]. 福建工程学院学报, 2007, 5(6): 596-599.
- [7] 岳彬. 大宗油料供应数据仓库模型设计与实现[J]. 微计算机信息, 2007, 24(5): 149-150.
- [8] 徐洁磐. 数据仓库与决策支持系统[M]. 北京: 科学出版社, 2005: 12-30.
- [9] 武彤. 高校教学质量分析与评估系统的数据仓库模型[J]. 贵州工业大学学报: 自然科学版, 2007, 36(5): 54-58.
- [10] 刘维晓,陈俊丽,王旺根,等. 基于 Web 的税务数据仓库的设计与实现[J]. 计算机工程与设计, 2010, 31(8): 1722-1725.
- [11] Inmon W H. Building the data warehouse[M]. [s. l.]: John Wiley & Sons Inc, 2005: 59-60.
- [12] 王小洁,朱程荣. 面向 CRM 系统的数据仓库的设计与实现[J]. 计算机工程与设计, 2007, 28(21): 5088-5090.
- [13] 蔡艳宁,叶雪梅,汪洪桥,等. 军事训练考核数据仓库模型设计与实现[J]. 计算机工程, 2006, 32(1): 276-278.
- [14] 程辉. 决策树算法在生产线质量控制决策系统的应用研究[D]. 贵阳: 贵州大学, 2013.

生产质量控制数据仓库模型设计与实现

作者：[王晓鹏](#)，[武彤](#)，[WANG Xiao-peng](#)，[WU Tong](#)
作者单位：[贵州大学 计算机科学与信息学院, 贵州 贵阳, 550025](#)
刊名：[计算机技术与发展](#)[ISTIC](#)
英文刊名：[Computer Technology and Development](#)
年，卷(期)：2015 (6)

引用本文格式：[王晓鹏](#). [武彤](#). [WANG Xiao-peng](#). [WU Tong](#) [生产质量控制数据仓库模型设计与实现](#) [期刊论文] - [计算机技术与发展](#) 2015 (6)