

文章编号 :1007 - 649X(2000)04 - 0070 - 04

数据仓库技术在税务征管业务中的应用

王 愚¹,邵后印²,陈勇孝³,郎 洪³,段银田³

(1. 郑州工业大学电气信息工程学院,河南 郑州 450002;2. 商丘市技工学校,河南 商丘 476000;3. 郑州工业大学数理力学系,河南 郑州 450002)

摘 要:分析了税务征管部门在计算机应用方面存在的问题,提出了在税务征收管理方面的数据仓库技术的解决方案.根据计算机在税务征管部门应用的现状,规划了数据仓库的框架,建立了数据仓库的逻辑模型并作了物理实现,同时对于数据仓库的数据迁移技术进行了一定的研究.在税务征管部门建立数据仓库系统,为征管部门的科学决策提供了强大的分析工具.

关键词:数据仓库;主题;维度表;事实表;星形模式

中图分类号:TP 317 文献标识码:A

0 引言

目前,税务征管部门采用的税务综合管理信息系统,实现了在税务登记、申报征收等方面的办公自动化.随着税务综合管理信息系统在税务征管部门的广泛应用,积累了大量的历史数据,如何能够充分利用大量的历史数据,提供对于税务征管数据的综合统计分析、辅助决策,帮助征管单位提高征管质量,已成为税务部门急需解决的问题.本文提出了把数据仓库(Data Warehouse)技术应用到税务征管业务的解决方案.

1 数据仓库的概念和特征

数据仓库是面向主题的、集成的、不易失的、时变的数据集合,用于管理决策^[1].数据仓库主要有以下特征.

1.1 面向主题

主题是一个在较高层次将数据归纳的标准,如纳税人、征管单位、发票等.主题组织的数据被划分为各自独立的领域,每个领域有自己的逻辑内涵,互不交叉.

1.2 集成

这是数据仓库建设中最关键、最复杂的一步.数据在从面向应用的操作环境中提取到数据仓库中时,都要经过集成化.首先要统一原始数据中的

所有矛盾之处,其次还要将原始数据结构做一个从面向应用到面向主题的转换.

1.3 非易失性(只读性)

数据仓库的数据主要供企业决策分析之用,所涉及的数据操作主要是数据查询,一般情况下并不进行修改操作.

1.4 时变性

数据仓库中的数据都是历史数据,数据时限远远长于操作型环境中的数据时限,数据仓库中的数据时限一般为 5~10 年,而后者只有 60~90 天.

2 税务部门构造数据仓库面临的主要问题

2.1 是逐步实施还是一步到位

企业级数据仓库的实现通常有两种途径:一种是从建造某个部门特定的数据市场(Data mart)开始,逐步扩充数据仓库所包含的主题和范围,最后形成一个能够完全反映企业全貌的企业级数据仓库;另外一种则是开始就从企业的整体来考虑数据仓库的主题和实施,类似于软件工程的“自顶向下”的方法,这种方法投资大,周期长.

从目前税务部门的实际情况来看,第一种方法较为可行.首先是技术上的限制.由于各部门采用的开发平台、数据库管理系统、前端开发工具各

收稿日期:2000-05-12;修订日期:2000-06-28

基金项目:河南省科技攻关项目(98520010)

作者简介:王 愚(1973-),男,河南省确山县人,郑州工业大学硕士研究生.

万方数据

不相同,因此在现阶段建立统一的数据仓库也就不太可能;第二是认识上的原因.如果没有领导层的支持和协调,数据仓库和决策支持系统是不能成功的.这也是最关键的原因.

2.2 建立数据仓库需要长远的规划

采用逐步积累的方式建立数据仓库,最大问题是已有的框架难以把新的业务集成进来,在设计阶段必须对此有充分的估计.需要加强不同部门的沟通和协调,系统设计应从企业整体的角度来考虑.

3 建立面向税务征收管理的数据仓库

根据以上对于数据仓库建立途径的分析,选择税务征收管理作为主题来建立数据仓库.

3.1 设计数据库^[2]

设计和创建关系型数据库是成功地创建数据仓库的一个关键步骤.星形模式是一项非常重要的技术,它通过使用一个包含主题的事实表和多个非正规化描述的维度表来执行典型的决策支持查询.数据仓库常常使用星形模式来存储数据作为分析的基础,以便尽可能快地响应复杂的查询.根据税务征管包含的业务范围和决策分析的需求设计数据库的逻辑模型,该数据库的逻辑模型如图 1 所示.

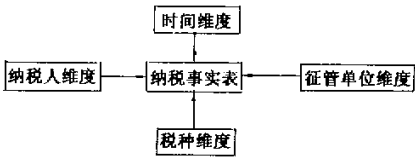


图 1 征收管理星形模式视图

3.2 维度表

维度表是用来为用户过滤数据或者聚合数据而提供路径的,可看作用户分析数据的窗口.在星形模式视图中,维度实体用星的角度来表示,如图 1 所示的时间维度、纳税人维度、征管单位维度和税种维度.

维度表是高度非规范化的,每个维度都有一个主键,用于区别每一维中的单个记录,同时包含分析数据所需的属性.分别设计时间维度、纳税人维度、征管单位维度和税种维度,如表 1 所示.

3.3 事实表

在星形模式视图中,事实表包含在星的中心,由两部分组成.一部分包含了由各维度表的主键构成的事实表的主键;另一部分包含了数据仓库的数值指标,称为事实或指标.这些指标是数值化的,具有可加性或半可加性.事实表是高度规范化的,通过各维度表的外键同数据分析过程中使用的指标连接起来.

表 1 维度表、事实表及其属性

时间维度	纳税人维度	征管单位维度	税种维度	纳税事实
时间键	税务登记号	征管单位键	税种键	时间键
日期	纳税人名称	征管单位名称	税种名称	税务登记号
月份	纳税人地址	地/市级税务局		征管单位键
季度	行业	县/区级税务局		税种键
年份	经济性质			应纳税额
	纳税期限			实纳税额

3.4 星形模式的物理实现^[3]

采用 SQL Server 7.0 来建立数据库.建立数据库可以使用 GUI 界面的 Enterprise Manager 组件或者使用 Query Analyzer 组件.星形模式由“taxcollecting”事实表和“time”、“taxkind”、“taxpayer”、“tax-office”4 个维度表组成.把 taxcollecting 表和索引放入一个文件组 factfg 中,把所有维度表和索引放入另一个文件组 dimensionfg 中.

3.4.1 创建事实表及主键

```
create table taxcollecting( Timekey int not null ,
Registerid int not null ,
Taxofficekey int not null ,
```

```
Taxkindkey int not null ,
Taxpaying money ,
Taxpayed money
Taxowing money ) on factfg
Go
Alter table taxcollecting add constraint taxcollect-
ingpk Primary key
( Timekey ,Registerid ,Taxofficekey ,Taxkindkey )
on factfg
```

3.4.2 创建维度表及其主键

```
create table time( Timekey int not null ,
Date dauctime
```

Month smallint
Quarter tinyint
Year smallint) on dimensionfg
Go

Alter table time add constraint timepk
Primary key (timekey) on dimensionfg

其他三个维度表及其索引的创建同 time 表相似。

3.4.4.3 参照完整性(RI)

参照完整性是在数据仓库中强制数据在不同表中关系的机制 ,因此在物理实现数据仓库时 ,一定要建立各种外键约束来定义不同表间的关系。

建立时间维度和纳税事实表之间的关系如下 :

Alter table taxcollecting add constraint timefk
Foreign key (timekey) references time

同理 ,可以建立其他维度表和纳税事实表之间的外键约束关系 ,以保证数据仓库的参照完整性。

4 提取和加载数据仓库

建立物理数据仓库后 ,下一步就是把税务征管系统中的数据提取出来加载到数据仓库中。如果数据源中的数据和将要出现在数据仓库中的数据是直接关联的 ,那么提取和加载数据仓库将比较简单 ;如果数据源的数据驻留在多个异构系统中 ,并且在加载数据之前需要大量的格式转换和修改 ,则将比较复杂。一般提取和加载数据的进程如图 2 所示。

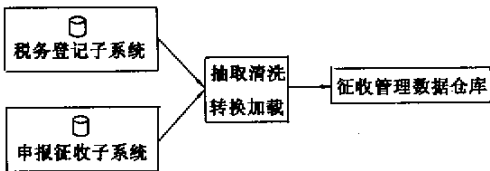


图 2 数据仓库的数据迁移

4.1 数据抽取和清洗

我们国家的税制和税务征管模式是统一的 ,因此纳税人的税务登记证号、税目、税率、税款级次等编码是统一规范的 ,当 OLTP 系统中的数据加载到数据仓库时 ,一般不需要数据清洗(Scrubing)的处理。

在税务登记管理子系统的税务登记数据库中 包含关于纳税人的大量数据 ,诸如税务登记证号码、工商营业证号码、注册资金、经济性质等。从登记库中抽取纳税人维度表所需的数据 ,加载到数据仓库中。

在税务申报征收子系统的申报征收数据库中 ,包含有关纳税人纳税情况的大量数据 ,诸如纳税人税务登记证号码、纳税日期、税种、应纳税额等。

4.2 数据转换和加载^[4]

在数据迁移的过程中 ,通常要将可操作系统 (Operational system)中的数据转换成另外一种格式 ,以更加适用于数据仓库的设计。在大多数情况下 ,转换是将数据汇总 ,有时需要把数据分解 ,例如纳税人的纳税日期 Date 在数据加载时 ,Date 要拆分成年、季、月 3 个列值来构建时间维度 ,以便按照年、季、月进行数据汇总。

SQL Server 7.0 提供了数据转换和加载的工具 DTS(Data Transformation Services)。可以利用 DTS 的 Import and Export Wizards 交互式创建 DTS packages ,DTS packages 可通过 OLE – DB 和 ODBC 实现数据的导入、数据的导出、数据的有效性检查和数据的转换。

利用 DTS 工具 ,可以把税务登记管理子系统 中的数据和税务申报征收子系统 中的数据经过抽取、清洗、转换后 ,加载到数据仓库中。

5 结束语

建立面向税务征管业务的数据仓库 ,为税务征管单位提供了多角度、多层次查询分析数据和制作动态报表的功能。该数据仓库应用于税务征管业务管理中 ,必将进一步规范税收管理 ,理顺税务系统内部的业务关系 ,为税务征管部门科学决策提供强大的分析工具。

参考文献 :

[1] INMON W H. 数据仓库[M]. 王志海 ,译. 北京 :机械工业出版社 ,2000.
[2] TOM Hammergren. Data Warehousing :Building the Corporate Knowledge Base[M]. North Carolina :Ventana Communications Group ,1997.
[3] MICHAEL Corey ,MICHAEL Abbey. SQL Server 7 Data Warehousing[M]. New York :McGraw – Hill Book Company ,1999.
[4] 闪四清. 数据仓库分析及应用策略[J]. 个人电脑 , 1999(10) :160 – 166.

Applications of Data Warehouse to Taxation Management

WANG Yu¹ , SHAO Hou - yin² , CHEN Yong - xiao³ , LANG Hong³ , DUAN Yin - tian³

(1. College of Electrical & Information Engineering ,Zhengzhou University of Technology ,Zhengzhou 450002 ,China ; 2. Shangqiu Mechanic School ,Shangqiu 476000 ,China ; 3. Department of Mathematics ,Physics & Mechanics , Zhengzhou University of Technology , Zhengzhou 450002 ,China)

Abstract :The existing problems of management in the taxation collection bureau in the computer application are analyzed. A solution to building data warehouse for management of collection in the taxation bureau is proposed. The frame of data warehouse is planned on the basis of computer application in the taxation collection bureau at present. The logical model and physical database are built and techniques of transferring data are studied. Building data warehouse provides a powerful analytical tool of scientific decision for the taxation collection bureau.

Key words :数据仓库 ; subject ; dimension table ; fact table ; star schema