

失业保险数据仓库的构建

Construction of an Insurance data warehouse for unemployment

汪 斌 (重庆三峡学院经贸系 404000)

摘要:本文分析研究了数据仓库技术,介绍了运用这种技术构建失业保险数据仓库,提出了具体的解决方案。

关键词:数据仓库 失业保险 DTS OLAP

1 引言

随着社会劳动保障制度的改革和发展,各地加强失业保险管理的力度,先后使用了失业保险管理软件,解决了人工管理的难度和复杂性,解决管理部门缴费单位台帐记录和失业保险金发放等问题,提高效率。同时随着决策信息需要,研究失业保险数据仓库的构建已很有必要。失业保险数据库管理系统的广泛应用,积累了丰富的基础数据,这些数据真实地记录了单位失业保险情况和运行过程,隐含着大量有价值地信息,对其抽取,能使我们发现参保单位和失业人员的情况,有效的支持单位领导决策,并从中得到有价值的信息,为决策层提供及时、准确的信息。

1999年7月失业保险数据库管理系统由作者开发并推广使用至今,该软件已具备基本模块有:初始数据、失业保险管理、失业人员管理、报表统计、失业保险基本情况和系统管理,在此基础上构建数据仓库,并提供决策依据和数据是很有价值的。

2 数据仓库技术

数据仓库是一个用以更好地支持企业或组织决策分析处理的、面向主题的、集成的、不可更新的、随时间不断增加的数据集合。数据仓库是决策支持系统等分析型处理系统的基础,用来更好的支持分析型处理,它存储的是分析型数据。它从多个同构或异构的传统数据库中获取原始数据,先按辅助决策的主题要求形成当前基本数据层,再按综合决策的要求形成综合数据层,主题是在较高层次上将信息系统中的数据综合、归类并进行分析利用的抽象,它对应单位中某一宏观分析领域所涉及的分析对象。数据仓库存放历史数据,这些数据通过数据析取程序从联机事务处理(OLTP)系统的数据库抽取。数据析取可对 OLTP 系统数据库中数据拷贝、过滤,也可对 OLTP 数据库数据作变化、统计。一个典型的数据仓库系统一般由三层组成:

- (1) 最底下的一层是数据仓库,即数据层;
- (2) 中间的一层是中间件,即逻辑层;

(3) 顶上的那层是决策支持工具,即表示层,如联机分析处理(OLAP)、数据挖掘等。如图1。

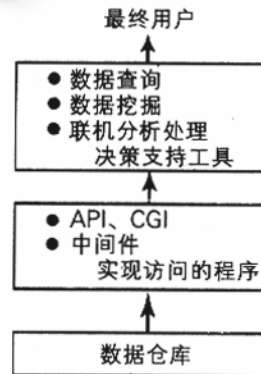


图1 数据仓库系统的一般结构

3 数据仓库的设计

数据仓库中的数据是面向主题进行组成,它对应单位中某一宏观分析领域所涉及的分析对象。数据仓库的最终目标是把范围内的所有数据集成在一个大仓库中,让用户能运行查询、产生报告、执行分析。经过分析,失业保险数据仓库的主题域主要有参保单位缴费台帐、失业保险花名册、失业保险金。根据这三个主题域设计的数据仓库结构具体如下:

(1) 参保单位缴费台帐(单位名称、单位编码、时间、投保人数、单位缴费工资、个人缴费工资、单位应交、个人应交、单位实交、个人实交、单位欠交、个人欠交)。

(2) 失业保险花名册(单位名称、日期、单位编码、序号、姓名、性别、出生年月、参加工作时间、工种、身份证号码、家庭住址、备注、是否已不在本单位)。

(3) 失业保险金(失业证编号、身份证号码、姓名、性别、出生年月、原工作单位、参加工作时间、失业时间、失业登记时间、失业原因、投保年限、月数、起始时间、到期时间、已领月数、未领月数、领取金额、年份、月份、经办人)。

4 数据仓库的构建与 OLAP 的实现

国际各大 IT 公司均提出自己的数据仓库解决方案,如:IBM、Oracle、SAS、Microsoft 等。这里系统采用 Microsoft 数据仓库技术,C/S 模式,数据库管理系统采用 Microsoft SQL Server 7.0,前端采用 Visual Basic 开发。SQL Server 7.0 提供数据仓库解决方案,通过数据转换服

务(Data Transformation Service,简称 DTS)和联机分析处理服务(OLAP Service)两个服务来支持数据仓库技术。

4.1 数据析取的实现

由于数据源的多样性和异构性,建立一个数据采集系统按数据仓库的设计要求从 OLTP 应用数据库中提取数据,重整后存放在数据仓库。各业务数据库的数据类型一般是不一样的,因此,必须确保数据的一致性和可用性,进行必要的数据库转换,根据业务的需求,抽取必要的数据库,进行综合。从不同的数据表和视图中提取数据并加载到数据仓库中,去掉一些环境数据,添加时间元素,在适当的位置增加由汇总或计算产生的数据,根据多维分析的需要人为增加数据间的关系。

失业保险数据仓库的建立采用 DTS 实现数据析取,完成数据的初试装入和追加。通过 DTS 可在多个使用以 OLE DB 为基础的体系结构的异质数据源间导入和导出数据,可以变换联机事务处理系统中的数据以建立数据仓库。

DTS 主要功能有导入和导出数据、变换数据和传送数据库对象。DTS 允许在一个过程中导入、导出和变换数据。这个过程的定义可以保存在 DTS 包(Package)中。DTS 包含三种类型对象:连接(Connection)对象、任务(Task)对象和步骤(Step)对象。连接对象定义每一个是数据的源或目标的 OLE DB 数据源;任务对象定义工作项目,例如执行一段 SQL 语句、拷贝一个表的内容或执行一段 ActiveX 脚本;步骤对象定义任务中心的次序。

定义包方式有三种:使用 DTS 设计器(DTS Designer)、DTS 导入和导出向导、DTS 编程接口。失业保险系统采用 DTS 设计器定义包。包可以三种方式保存:DTS 基于 COM 的文件(.dts)保存、保存在 SQL Server 的 msdb 数据库中、保存在 Microsoft Repository 中。定义好的包可以通过 SQL Server 代理(SQL Server Agent)服务让 SQL Server 自动、定期执行。

4.2 OLAP 的实现

联机分析处理服务(OLAP)是对大量多维数据的动态综合、分析以及归纳。OLAP 中的一个主要操作是“多维分析”,使用户能从多角度、多侧面观察数据库中的数据。OLAP 工具提供多维视图帮助分析人员观察数据,找出规律,因此它主要是一种数据分析工具,是对原来数据库上的查询工具的一种扩充。

当数据抽取完成后,就可利用 SQL Server 提供的 OLAP 工具进行分析,也可利用可视的、面向对象的图形化开发环境创建和维护定制的 OLAP 应用,利用查询和报表工具,返回最终分析结果。应用数据挖掘及 OLAP 技术还可作一些更为复杂和综合的查询分析。比如可分析某行业在某时期内失业人员与单位缴费的关系,从而对该区域失业保险

金发放起指导作用,同时为再就业工程提供依据。上级部门数据仓库的建立还可解决实际应用中多数据源问题、数据不一致性、历史数据充分利用、分析效率和随机查询分析等问题。

在 OLAP 实现中举例如下:需要提取某年各个部门(或行业、片区等有不同分类方法)的汇总缴费台帐,其中提取数据的程序用 SQL 语句来表达如下:

```
SELECT sbdw.所属部门, (Year([时间])) AS 年份,
Sum([缴费台帐查询].应交单位 1) AS 应交单位 Of Sum,
Sum([缴费台帐查询].应交个人 1) AS 应交个人 Of Sum,
Sum([缴费台帐查询].实交单位) AS 实交单位 Of Sum,
Sum([缴费台帐查询].实交个人) AS 实交个人 Of Sum,
Sum([缴费台帐查询].欠交单位) AS 欠交单位 Of Sum,
Sum([缴费台帐查询].欠交个人) AS 欠交个人 Of Sum
```

```
FROM [缴费台帐查询] LEFT JOIN sbdw ON [缴费台帐查询].单位编码 = sbdw.单位编码
```

```
GROUP BY sbdw.所属部门, (Year([时间]))
```

```
HAVING (((Year([时间])) = [Forms]![失保 CT1]![Text12]));
```

这样就很快得到上级需要的有价值的数据库,即各种分类台帐,如果是县城可按乡统计,是区级按行业统计,是市级可按区县统计,准确、及时。而在以前是很难做到这点的。

5 结束语

本文分析数据仓库技术及其在失业保险业务上的应用,提出具体应用解决方案。该数据仓库是建立在成功运行失业保险数据库管理系统基础之上的,由于系统已在十家单位成功使用,有了 6 年的基础数据,因此提升了失业保险数据库管理系统数据的价值。对于失业保险业务科室具有解决问题和指导实践的作用,使之工作规范、统一、高效、准确,以便更好的管理和决策。失业保险数据仓库的构建为解决地方社会保障和再就业工程决策依据提供有效途径,为提高社会保障事业的电子政务应用,促进社会保障信息化的发展,将带来明显的社会效益,可为进一步研究它的知识发现和数据挖掘奠定基础,是我国“金保工程”务实的基础工作之一。

参考文献

- 汪斌,失业保险管理系统的设计与开发[J],计算机应用与软件,2003.7.
- 郭朝珍、杨俊杰,基于数据仓库技术和 OLAP 的决策支持系统[J],计算机系统应用,2000.3.
- 曹重英等,应用数据仓库技术实现决策支持系统[J],计算机系统应用,2000.1.
- 李英杰,基于数据仓库技术的决策支持系统的应用研究[J],微型电脑应用,2000.11.