

人事劳资系统数据库设计

王 平 (航空航天部三院 304 所)

摘要:本文结合全国政协人事劳资管理信息系统数据库设计的过程,提出了对人事劳资系统数据库设计的方法和步骤,并叙述了每阶段实现的目标和可采用的技术及工具。

一、现行系统分析与设计

对人事劳资部门现行业务的分析,我们采用了结构化分析(SA)的方法,按照两个步骤进行。

(一)调查和收集数据

我们从数据处理的角度来理解和分解用户的各科业务活动,从而查清人事和劳资部门现行业务处理流程,需要调查和收集的数据有:

1.系统未来的用户

人事处和劳资处的报表统计人员、人事劳资管理人员。

2.用户的信息要求

输入要求和输出要求的一览表,所要求的报表形式。

3.系统的使用方式

交互式、批处理方式等。

4.系统存储的数据

职工档案、工资档案等信息。

5.系统的完备、保密和审查要求

把干部信息与工人信息分成两个子系统,分配不同的用户名/口令。

6.数据的来源、流向、性质、取值范围、数值使用频率和数据间的关系。

(二)分析数据及其加工要求

在对人事劳资信息系统分析中,收集到描述系统的业务处理内容,涉及的数据及其加工要求是多种多样的,必须对其进行认真的分析,若不透彻地分析这些数据,可能会忽略或误解了系统的要求和信息系统起作用的重要方面。

用户活动不一定必需由计算机来处理,也不一定都能由计算机来处理,因此,就要划清计算机管理信息系统的边界,即确定哪些活动由计算机去完成,哪些活动仍由手工去完成。我们通过对人事劳资部门的应用进行分析,最后划分出人事劳资部门系统边界。同时对各类用户的应用方式(交互式、批处理),应用类型(维护、查询、统计分析、图表生成、模型参数、决策因子)使用的数据进行分析,找出最基本的信息和处理过程,汇总各类数据查询及更新频率,确定响应时间和平均事务处理时间,确定用户存取权限。分析数据项和数据项之间的联系,产生出人事劳资部门的数据字典。通过分析可以得出:人事劳资部门的最基本信息是职工档案中干部和工人的基本信息,人事劳资部门的主要信息处理是人事劳资部门干部和工人花名册的报表统计。

二、逻辑数据库设计

逻辑数据库设计是把一个企业管理中所需要的信息用实体、属性以及用实体之间的关系来进行描述,从而导出逻辑数据库格式,它是一种抽象的数据格式,是以企业对数据的需求以及数据之间的联系的角度来分析问题的。逻辑模式设计是设计出一个既独立于计算机硬件,也独立于特定的 DBMS 的逻辑数据库结构,如何设计出人事劳资部门的逻辑数据库,使之能正确地反映人事部门实际的业务活动,是逻辑数据库设计阶段所要解决的主要问题。

逻辑模式的设计方法有许多种,但大致上分为两类:一类是面向信息实体的;另一类则是面向应用的。第一类方法主要有 Peter chen 的实体关系方法(Entry-Relation Model), John Smith 抽象方法;第二类

方法主要有 Michael Mitoma 的自动化数据库格式设计和最优化方法,以及 S.B.Yao 的关系数据库格式设计工具。根据全国政协的实际情况,结合数据库设计方面的经验,选 Peter Chen 的实体关系方法。E-R 方法的基本思想就是把一个企业中各部门关心的信息抽象出其中的基本信息要素,然后用实体、属性、关系进行描述。鉴于以上考虑,采用 E-R 方法作为数据库设计的主要工具,对逻辑模式设计采用如下设计步骤:

(一)实体、属性和关系的提取

从各子系统所关心的信息中抽象出实体、属性和关系。“实体”是一个企业感兴趣的最基本的东西,实体可以是人、地方、物体、事件或概念,实体并不一定在该环境中物理地存在,“属性”是和某个特定实体有关的描述值或特性,“关系”是两个实体或多个实体之间的联系。实体的提取方法有两种:其一是实体和报表一一对应;其二是根据数据的逻辑关系来提取实体。第一种方法的缺点是:如果报表格式经常变动的的话,就会导致数据库结构和程序频繁变动,两者比较,决定采用第二种实体提取的方法。人事劳资部门比较关心的实体是职工基本信息(职工花名册)、简历、培训、奖惩、配偶子女等。其中干部的属性是:编号、姓名、性别、婚否、工作时间、调入时间、间断工龄、民族等。

(二)设计系统的 E-R 图

通过与管理人员进行不断反复讨论,逐步求精地设计出系统的 E-R 图。E-R 图中,实体用矩形表示,关系用菱形表示,属性用圆圈表示,初步设计出系统的 E-R 图可能存在冗余的属性,冗余的实体和实体之间的关系,经多次协调,消除不同视图之间的冲突或冗余,同时根据规范化理论,对全局的 E-R 图进行优化,得出满足第三范式的 E-R 图。

(三)逻辑数据库结构的构成和优化

逻辑数据库结构的构成是把 E-R 图转换成逻辑数据结构,其目标是产生一个满足信息需求的逻辑数据格式,在数据处理系统中,实体可以用记录型来表示,属性则典型地由字段或数据项来表示。表示关系可以用不同的方法,这取决于 DBMS 软件的数据描述能力。例如,可以利用相互联系实体的公共字段的办法表示关系。对于 ORACLE 数据库管理系统,最简单的逻辑映象是有关系的两个表设有公共属性的字段。根据以下规则,可

以把一个实体关系图转换成一组逻辑数据库格式:

1.将 E-R 图中每一个带属性的实体转换为一个“关系”,标识属性作为“关系”的主关键字。

2.将 E-R 图中每个带属性的关系也转换为一个“关系”,关系中涉及到的标识属性作为“关系”的主关键字。

3.弱实体的主关键字应包括其父实体的标识属性,也即父实体的标识属性是该弱实体主关键字的一部分。

对初步设计的逻辑数据库结构需要进行优化,优化过程就是要找出一个性能比较好的数据模式原则和改造过程,经过优化处理,就可得出比较好的逻辑数据库结构。

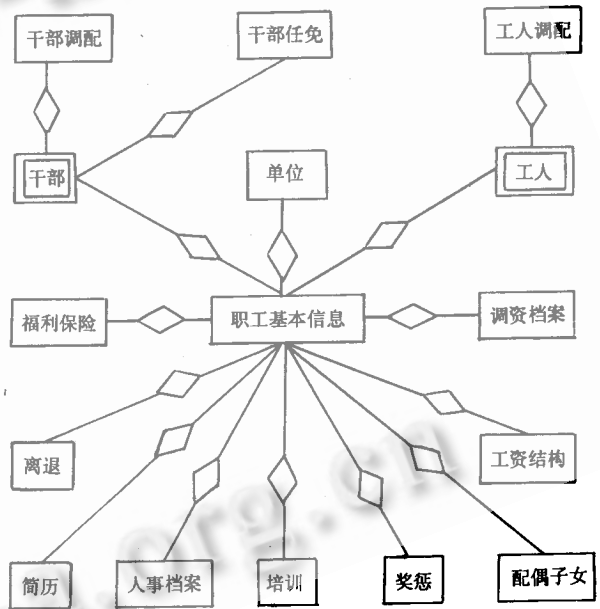


图 1 人事劳资部门 E-R 图

三、物理数据库设计

物理数据库设计是在逻辑数据库结构的基础上,通过对信息量估计,结合用户各种应用的处理频率、轻重缓急和 DBMS 特点,对时间和空间效率等进行平衡,对设计出的逻辑数据库结构进行转换,从而设计出一个较适合应用环境的物理数据库结构。全国政协人事劳资系统的物理模式设计,是在 AST/386 的 DOS、ORACLE 数据库管理系统的环境下进行的,采用如下设计步骤:

(一)数据信息量的统计与分析

各种数据的信息量大小,直接影响将来数据库的设计,为了便于分析和统计,将人事劳资数据分为两类:

1.静态数据

某一类数据的信息量是在一定时间内是相对稳定的,这类数据属于静态数据。如职工档案信息、职工简历信息、职工案友信息、职工工资结构档案、职工调资档案、人事劳资代码信息等。对全国政协机关干部和工人的档案等各方面静态的信息进行分析,从而得出全国政协人事劳资的静态信息为1.5MB。

2.动态数据

另一类数据是日常工作中随时产生,数据量随时间增长而增大,甚至每天、每月的数据量也不一样,这类数据属于动态数据。如干部工人花名册、干部工人奖惩信息、干部工人培训信息、人事劳资统计报表等。通过抽样调查,进行统计分析,便可获得这类数据的信息量。全国政协人事劳资的动态信息为IMB/年。

数据信息量的统计工作完成以后,应及时对搜集到的信息进行分类。将信息分为基本信息、统计信息、辅助信息和接口信息四类,按类进行数据库设计,减少数据库设计的复杂性。

(二)系统间接口设计

信息接口多是指人事子系统与劳资子系统之间、劳资子系统与经费子系统之间,以及与上级机关子系统之间,信息交换和处理频率,譬如劳资部门要共享干部花名册的基本信息。

全国政协管理信息系统各子系统的数据库交换很多,设计好数据之间的接口是实现数据库数据共享,保证管理系统协调工作的基础,各子系统之间数据交换是通过建立接口库来实现的。各子系统之间的接口格式,是由各子系统数据库设计人员经过反复讨论确定的。由于接口结构明确,接口控制关系清晰,为各子系统联调提供了基础。

(三)确定数据的物理表示

充分地了解用户对系统的要求,针对每一个程序的功能,采用“分块设计”的原则,掌握信息处理的现状及特殊的处理方式,确定所有数据项的类型及字段长度。

ORACLE 数据库表示的数据类型有三种:字符型、数字型、日期型,对逻辑数据结构中的所有数据项,规定数据类型、字段长度、包括在可能的条件下规定数据项是

否为主键、是否非空、是否代码,从而得出满足应用的物理数据结构(如表1所示)。

表1 人事劳资管理系统数据库结构

中文表名		西文表名				来源		
用途		频度				记录数		
序号	中文项名	西文项名	类型	长度	小数	代码	主键	非空
..
..
..

(四)代码设计和代码编制

为使数据查询、汇总和统计程序简便,方便信息交换和处理。对党派、民族、省市、职务、职称、专业、干部类别、工种、家庭关系、单位等需要编码的字段进行了代码设计。另一方面,采用代码值的存储,是一种重要数据压缩技术,它不仅在于节省空间,而且提高了数据的传输速度。代码设计是一项十分严肃的工作,要对编码的字段规定编码的方式,定义每位代码的意义,得出编码规则。

在完成代码设计以后,接着进行代码编制工作,形成全国政协人事劳资系统代码表。代码编制工作是很重要的。编制原则是有国标的尽量采用国标,没有国标则尽量采用中组部和人事部标准,没有国标又无部标的特殊代码则需自行编制统一代码。譬如党派、民族、省市、家庭关系、职称等采用国标,根据全国政协的具体情况,对单位、职务、工种等则自行编制代码(如表2所示)。

表2 代码选用标准汇总表

字段名	选用标准
党派	国标
民族	国标
省市	国标
文化	自定义
职务	自定义
职务级别	自定义
职称	自定义
职称级别	自定义
专业	国标
干部类别	自定义
家庭关系	中组部
单位	自定义
工种	自定义

(五)利用 ORACLE 特点

在设计中充分地考虑到 ORACLE 的特点,结合一般物理模式设计的原则,采用了以下五种设计技术:

1.冗余技术(denormalization)

为了减少 join 操作,对数据记录数多的表,将本来已规范化的第三范式进行修正,增加某些冗余字段,变成不规范形式,减少 join 操作,提高存取速度。

2.建立索引(index)

查找数据库时,存取速度与数据库中记录的多少有关,当记录数很大时,顺序查找速度是很慢的,为此利用了 ORACLE 数据库提供的索引功能,对一些记录数多的表加上索引,当数据库记录数达到 10000 以上时,加索引以后存取速度可以提高十倍。

3.表空间定义(define space)

由 ORACLE 系统资料可知,当一个表的记录很多时,系统分配的空间不连续,查询、修改的时间就增加,因而需对某些表进行空间定义,例如对货仓资料表进行空间定义,一次就分一百块初始的逻辑块,使录入的资料能

物理地连续存放在一起。

4.数据库的完备性、安全性和转存

ORACLE 管理系统有口令和授权等保护措施,对于本系统数据库管理人员给予较高的操作权限;对于非本子系统的用户,根据不同应用授予 SELECT、UPDATE 等权限。

经过以上几个步骤的设计,就可得出初步满足软件开发应用的物理数据结构。此时,并不表明数据库设计工作已经完成,随着软件开发工作的开发,需要在原设计步骤过程中不断的迭代完善和修改。

参考文献:

[1]R.S.Pressman, *Software Engineering—A Practioner's Aproach*

[2]Lawrence J. Peters, *Software Design and Methods & Techniques*

[3]吴鹤龄,数据库原理与设计,国防工业出版社,1987(12)

[4]姚卿达,数据库设计,高等教育出版社,1987(4)