

陈 鍊 编著

数据采集方法

SHUJUCAIJIFANGFA



中国审计出版社

图书在版编目 (CIP) 数据

数据采集方法/陈鍊编著, —北京: 中国审计出版社, 2001.10

ISBN 7-80169-089-3

I.通… II.陈… III.数据采集 IV.TP274

中国版本图书馆 CIP 数据核字 (2001) 第 069610 号

数据采集方法

陈鍊 编著

出 版	中国审计出版社
地 址	北京东城区东四十条 24 号青蓝大厦
邮政编码	100007
电 话	(010) 88361317 64066019
传 真	(010) 64066026
发行经销	新华书店总店北京发行所发行 各地新华书店经销
印 刷	北京市白帆印刷厂
开 本	787×1092 1/16
版 次	2001 年 10 月北京第 1 版
印 次	2001 年 10 月第 1 次印刷
印 张	19
字 数	430 千字
印 数	1~5000 册
定 价	29.00 元
书 号	ISBN 7-80169-089-3/TP·002

版权所有 侵权必究

序

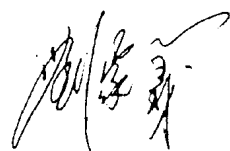
古人云：工欲善其事，必先利其器。器者，科学技术也，抑或通常谓之理论研究、发明创造、器械设备、技巧方法、技术手段等等之总称也。

在信息科学和技术飞速发展，我国正向信息化社会急速奔进的今天，国民经济各方面、社会再生产各环节、经济运行各要素和社会经济生活以及管理社会经济生活的方式、手段等正在发生深刻变化。作为监督国家财政收支、国有金融机构和企业事业组织财务收支真实、合法、效益的国家审计机关，欲善其事——依法履行职责：维护财经秩序与信息公允、促进经济发展与廉政建设、推进法制建设与规范管理，必利其器——实施有效监督：运用先进设备与手段、改进审计方式与方法、提高审计技巧与技术。惟利其器，才能提高审计效率与质量，降低审计风险与成本，大善与尽善其事。

为利其器，各级审计机关、广大审计理论研究人员和实务工作者，孜孜求索，几年奋斗，终拨云雾，初露曙光：审计准则的制定、审计基本技术与方法的研究、专业审计操作指南的研制和计算机网络的构建、软件的开发、设备的配置等审计信息化建设的快速推进……，真乃雨后春笋，百花绽放。《数据采集方法》便是其中之一。

事虽纷繁，纲领乃一，器虽众多，关键者寡。关键之器击纲领之事，方解“瓶颈”之束。随着信息化的快速发展，采集获取、查询分析各种电子数据即如取得各种纸质帐本、凭证、资料一样，必然成为审计监督的首要步骤，否则，审计将无物可审。然在信息化快速发展过程中，统一平台的建立、技术标准的规范等尚未解决，如何采集获取电子数据，采集获取后如何查询分析即如取得各种纸质帐本、凭证、资料后如何查阅、

审核，必然成为审计监督的关键环节，否则，审计将无从下手。当前计算机审计，大多遇此“瓶颈”之困。为解此困，众多志士不懈努力，探求了多种方案与技巧。南京审计学院的陈鍊老师将教学、科研与总结实际工作之经验紧密结合，融为一体，经过几年探索，撰写了《数据采集方法》一书。虽未细读，初阅介绍，亦很欣喜。嘱予作序，实难当之，聊书几言，谨表贺意。



二〇〇一年十月十一日

前言

计算机在各行各业的应用，给审计、财政、税务、工商行政管理等经济监督部门提出一个崭新的课题：如何提取分析各种存储格式不同数据库里的数据。不解决这个难题，审查和监督就无从谈起。审计署审计长就此告诫过审计人员：“审计人员不掌握计算机应用技术，将失去审计的资格。”推而广之，从事经济监督的人员不掌握计算机应用技术，也将失去监督的资格。

解决这个难题十分紧迫。当前，以计算机技术、网络技术为代表的信息科学和技术，对世界经济的发展已经并将继续产生深刻影响，突出表现在国民经济中信息产业比重上升和经济管理的信息化，我国的经济管理信息基本上都是电子数据了，这种特征越往管理高层越明显。“十五”期间，我国的信息化还将有一个长足的发展，必将导致各类管理数据尤其是经济管理数据的进一步电子化。

任何事物的发展都要有个过程，自然是先有信息化的经济管理，再有信息化的审查监督方式。在以往对数据库应用技术的研究，尤其是推广数据库应用书籍的内容中，主要侧重于数据库的建立、维护操作和数据控制方面，缺乏系统有效的数据采集方法，可视化使用的具体操作介绍就更难见到，因而在急需数据采集技术的审计、财政、税务、工商管理等行业形成上述“瓶颈”般计算机应用难题。《数据采集方法》就是为寻求一个较好的解决方案而萌发，并在为审计人员的计算机应用技能培训工作中逐步充实完善。

运用系统工程的思想方法，本书首次系统完整地归纳出在计算机应用系统不同层次采集数据的通用方法。首先，列举了

在操作系统层复制数据文件副本的方法，并着重叙述局域网登录联网实用技术；然后，重点介绍在数据库管理系统层采集数据既简便又快捷的通用方法：ODBC+SQL SELECT，详细阐述各种常见环境使用 ODBC 的技术，以大量示例说明 SQL SELECT 命令的每个细节；最后向读者推荐一个运用 ODBC+SQL SELECT 方法的通用数据采集小工具：MS Query。

“拿来主义”是本书的显著特点，就是完全用人所皆知、随处可得的通用软件，开发出在各种环境下采集分析电子数据的一整套方法，而且尽可能使用可视化的操作方法。本书介绍的方法不仅使管理监督部门的非计算机专业用户“拿来”就能用，而且在用户的灵活运用下能够解决各种复杂多变的数据采集问题，尤其与使用某些专用软件转换数据相比，“版本升级换代”的后顾之忧可能就少多了。

本书的编著得到审计署刘家义副审计长、审计署计算技术中心王智玉主任、南京审计学院易仁萍院长等领导同志的热情关心和支持；东南大学计算机科学与工程系的博士生导师孙志挥教授审阅了全书，提出了许多指导性意见。中国审计出版社宋灵恩社长和总编室王永平主任大力支持，使本书以最短的出版周期与读者见面。在此，对所有为本书的编著、审阅和出版给予支持和帮助的同志，包括来自全国各地审计一线曾经提出各种有关数据采集问题的学员们，一并表示衷心的感谢。

由于作者水平有限，加之时间匆促，书中错误和不妥之处在所难免，诚望读者批评指正。但愿这本小册子，能为计算机在审计、财政、税务、工商行政管理等经济监督部门的应用和推广，为在我国营造更好的经济发展环境做点贡献。

作 者

2001 年夏 于南京

目 录

第 1 章 电子数据采集分析方法概论	1
1.1 数据采集困难的原因	2
1.1.1 计算机硬件系统的结构类型	2
1.1.2 操作系统的结构类型	2
1.1.3 数据库管理系统的结构类型	3
1.1.4 数据文件的存储格式	3
1.2 不同层次的数据采集方法	4
1.2.1 操作系统 (OS) 层的采集方法	4
1.2.2 数据库管理系统 (DBMS) 层采集方法	5
1.2.3 应用程序接口 (API) 层的采集方法	6
1.2.4 电算化会计软件系统 (AIS) 层采集方法	7
1.3 不同采集方法的思考	8
第 2 章 操作系统层数据文件复制方法	11
2.1 复制什么文件	12
2.1.1 复制文件的文件类型	12
2.1.2 复制文件的文件名	12
2.2 单机环境的文件复制方法	13
2.2.1 Windows 环境的复制方法	13
2.2.2 MS DOS 环境的复制方法	32
2.3 网络环境复制文件方法	36
2.3.1 网络环境文件复制基本方法	36
2.3.2 登录局域网	37
2.3.3 从网上邻居复制文件	45
2.3.4 从网络驱动器复制文件	49

第 3 章 ODBC 原理及其应用——桌面数据库环境采集方法	51
3.1 ODBC 概述	52
3.1.1 什么是 ODBC	52
3.1.2 创建 ODBC 原因	52
3.1.3 使用 ODBC 的好处	53
3.1.4 使用 ODBC 的缺陷	53
3.2 ODBC 体系结构	54
3.2.1 ODBC 数据库应用程序	54
3.2.2 ODBC 驱动程序管理器	55
3.2.3 数据库驱动程序	55
3.2.4 数据源	55
3.3 ODBC 驱动程序类型及数据库访问模型	56
3.3.1 单层 ODBC 驱动程序与本地数据库访问模型	56
3.3.2 多层 ODBC 驱动程序与客户/服务器数据库访问模型	57
3.4 基于 Windows 平台的 ODBC 使用	59
3.4.1 ODBC 驱动程序安装	59
3.4.2 ODBC 驱动程序管理器	59
3.4.3 数据源管理	63
3.4.4 ODBC 驱动程序参数	68
3.4.5 数据源例	75
3.5 桌面数据库环境数据采集方法概述	76
3.5.1 内置驱动程序及两类方法比较	76
3.5.2 桌面 DBMS 数据采集方法	77
3.6 Access 环境采集数据	78
3.6.1 MS Access 的内置驱动程序	78
3.6.2 MS Access 内置驱动程序安装	79
3.6.3 从 MS Access 导出数据	80
3.6.4 MS Access 导入数据	84
3.7 Visual FoxPro 环境采集数据	89
3.7.1 从 Visual FoxPro 导出数据	89
3.7.2 Visual FoxPro 导入数据	99
3.7.3 用 Visual FoxPro 连接采集数据	112
3.8 MS Excel 环境采集数据	132
3.8.1 EXCEL 的内置转换功能	132
3.8.2 打开异构数据文件	133
3.8.3 工作表另存为异构数据文件	135
3.8.4 ODBC 转换工具	136

第 4 章 SQL 原理及其应用——大型数据库环境采集方法	137
4.1 SQL 概述	138
4.1.1 什么是 SQL	138
4.1.2 SQL 语言特点	139
4.1.3 SQL 语言的基本概念	141
4.2 MS SQL Server 概述	142
4.2.1 单进程、多线程数据库引擎	143
4.2.2 客户机/服务器结构	143
4.2.3 多种访问数据库方法	144
4.2.4 支持分布式数据库结构	146
4.3 MS SQL Server 界面工具及其使用	147
4.3.1 SQL Server 界面工具简介	147
4.3.2 SQL Server 的安全机制	148
4.3.3 Enterprise Manager 使用	150
4.3.4 Query Analyzer 使用	157
4.4 数据查询准备	161
4.4.1 掌握被审数据结构	161
4.4.2 查询结果的保存与共享	162
4.4.3 T-SQL 数据类型简介	163
4.5 T-SQL SELECT 命令	165
4.5.1 SELECT 命令概述	165
4.5.2 选取数据列 < data >	167
4.5.3 指定数据源表 < source >	170
4.5.4 指定结果去向 < into >	171
4.5.5 结果排序 < order >	172
4.5.6 筛选条件 < filter >	174
4.5.7 连接条件 < join >	187
4.5.8 列统计分组 < group >	200
4.5.9 行统计 < compute >	206
4.5.10 查询联合 < union >	208
4.5.11 子查询 (subquery)	212
第 5 章 MicroSoft Query 及其应用——通用工具	225
5.1 MicroSoft Query 简介	226
5.1.1 MS Query 功能用途	226
5.1.2 MS Query 安装	226
5.1.3 MS Query 的启动	227

5.2 建立数据源	228
5.2.1 指定检索数据的来源	228
5.2.2 指定数据源的注意事项	231
5.3 用向导建立查询及查询一般使用	231
5.3.1 创建新查询的准则	232
5.3.2 向导建立查询	232
5.3.3 查询的一般使用	235
5.4 用 MS Query 查询设计器建立查询	237
5.4.1 查询设计器界面窗口	237
5.4.2 设计简单查询	240
5.4.3 设计条件查询	243
5.5 用 SQL 建立查询	250
5.5.1 查看 MS Query 的 SQL	250
5.5.2 用 SQL 建立查询	250
5.6 查询及其数据结果集的编辑修改	252
5.6.1 查询数据浏览框的编辑	252
5.6.2 查询的修改	254
5.6.3 撤消/恢复编辑修改	256
5.6.4 更新数据源表	256
5.7 MS Query 查询排序	257
5.7.1 单列排序	258
5.7.2 多列排序	258
5.7.3 删除排序	259
5.8 MS Query 查询的分组和统计	259
5.8.1 可在计算字段中使用的内置函数	260
5.8.2 列分组	260
5.8.3 浏览框列标题栏设置计算列	261
5.8.4 使用选单创建计算字段	261
5.8.5 分组统计的工具按钮操作	262
5.9 MS Query 的多表连接查询	263
5.9.1 MS Query 的连接类型	263
5.9.2 用设计器建立表的连接	264
5.9.3 用设计器修改查询表的连接类型	265
5.9.4 用 SQL 建立编辑连接	269
5.9.5 删除连接	270
5.10 组合查询	271
5.11 建立参数查询	272
5.11.1 用 MS Query 建立参数查询	273

5.11.2 用 SQL 建立参数查询	274
5.11.3 参数的编辑设置	274
5.11.4 执行参数查询	275
5.12 在 Excel 使用查询数据	276
5.12.1 查询结果返回 Excel	276
5.12.2 在 Excel 编辑查询	277
5.12.3 在 Excel 共享查询数据	283
5.13 附录	286
5.13.1 用快捷键在 Microsoft Query 的窗格中移动	286
5.13.2 用快捷键选定数据和条件	287
5.13.3 用快捷键编辑数据和条件	289
5.13.4 Microsoft Query 的功能键	289
5.13.5 用于"Office 助手"的按键	290

第1章

电子数据采集分析方法概论

对于计算机审计来说，一个重要的内容是对电子数据账的实质性审查，这也是当前我国审计行业最重要的一项实务。用 IT (Information Technicality) 技术对电子数据账审计，从方法论上说，主要解决两个问题，一是审计人员如何采集到电子数据副本；二是如何审查分析获得的电子数据。为了能使用审计人员所熟悉的软件技术审查电子账，审计人员首先要获得被审单位电子账的数据，并能以可读形式打开，即首先要能采集到指定格式的被审电子数据。从审计实务来说，当前这一重要的实质性审计第一步工作遇到了来自两个方面的阻力：一是人为的阻碍，由于众所周知的原因，被审单位不仅不予配合，还可能提供不完整的或虚假的数据，甚至以种种借口不提供任何数据；二是技术性的障碍，面对被审单位五花八门的电算化会计软件，我们审计一线的同志感到棘手，因而电子数据的采集成为我国当前电子账审计的瓶颈。我们希望通过技术方法上找到采集分析各种电子数据账的有效途径，并且还要是简单适用，所花费的时间、精力和经费最少的通用方法。

1.1

数据采集困难的原因

电算化会计系统是一个计算机（软件）系统，而计算机（软件）系统是一个层次性的结构，从底层到用户界面一般依次为：

1. 计算机硬件系统
2. 计算机操作系统
3. 数据库管理系统
4. 电算化会计系统

其中，电算化会计系统可以直接运行在操作系统之上。由于不同的层次，可以有不同的结构类型，因而导致我们面对的电算化会计系统不仅在界面上，而且在结构上均呈多样性。

1.1.1

计算机硬件系统的结构类型

1. 单机结构，电算化会计软件运行在一台个人机（PC）上，所有的数据、操作也都集中在这台 PC 机上。
2. 网络结构，由若干台 PC 机连成网络，其中 PC 机称做工作站，数据可以集中在一台 PC 服务器上，或分散在连网的各个单机上，操作分散在网络上各单机。
3. 多用户结构，由一台主机和若干台终端组成，终端一般只有一个显示器一个键盘，也可以是 PC 机。数据集中在主机，分散在各终端的操作最后仍然作用到主机上。

1.1.2

操作系统的结构类型

1. 单任务操作系统，如 MSDOS、CP/M 等，计算机只能运行一个应用程序，在完成一个作业之前，不能交付另一个作业给计算机，以字符命令作为用户界面。一般作为单机的操作系统，也可作为网络上工作站的操作系统。
2. 多任务、多窗口操作系统，如 Windows 3.X, Windows 9X, 计算机可以同时运行多个应用程序，并可方便地在各应用程序之间传递信息，以图形、窗口作为用户界面，可

以作单机或网络工作站的操作系统。

3. 网络操作系统, 如 Novell 的 NetWare, MicroSoft 的 Windows NT 等, 管理调度计算机网络内的所有资源, 提供网络用户的使用环境, 早期的 NetWare 是字符界面, 与 MSDOS 兼容, 而 Windows NT 则与 Windows 9X 有完全相同的用户界面, 作为网络的操作系统。
4. 多用户操作系统, 如 Unix, Lunix 等, 管理调度多用户系统主机软、硬件资源, 为多用户系统的所有终端用户服务, 仅作为多用户的操作系统。

1.1.3

数据库管理系统的结构类型

1. 桌面 (Desktop) 系统, 安装在单机或小型网络系统上的小型数据库系统, 数据和操作都集中在安装数据库管理系统的 PC 机上。一般数据处理量较小, 管理较简单, 安全性较差。种类很多, 常见的有 DBase, FoxBase, FoxPro, Access, Excel 等。
2. 客户/服务器 (Client/Server) 结构系统, 数据库系统的核心软件 (称做数据库服务器) 安装在网络服务器或多用户主机上, PC 机上只能安装数据库系统的使用界面软件 (称做数据库客户), 这种大型数据库系统, 数据集中在服务器上, 操作通过分散的工作站或终端最后作用在服务器上, 数据处理能力强, 安全性、完整性好, 运行在网络或多用户环境, 如 Sybase, MS SQL Server, DB2, Oracl 等。

1.1.4

数据文件的存储格式

电算化会计软件一般是用数据库管理系统开发的, 而同一结构层次的数据库管理系统其数据文件的存储格式还有很大差异, 即使在简单的桌面系统 DBMS 中, 不同的系统其数据文件的存储格式也不一样, 更不用说复杂的大型数据库系统, 甚至是同一系列的 XBase (dBase, FoxBase, FoxPro, VFP) 数据库系统, 其数据文件的格式也是在不断变化的, 只能保持向下兼容。如果是使用通用编程语言 (如 C, COBOL 等) 自行设计开发的电算化信息系统, 则其数据文件存储格式的差别就会更大。

上面我们分析了组成电算化会计系统的不同层次结构上的差异, 这种差异理所当然会导致出现结构不同的电算化会计系统。

1.2

不同层次的数据采集方法

众所周知，我们使用计算机一般不会直接使用硬件系统，而是在操作系统，或者数据库管理系统，或者电算化会计系统等不同层次的软件环境里进行不同层次的操作，也就是说在软件的不同层次有不同的使用方法，因此也就会有不同的数据采集方法。即使对于同一个被审系统，在不同的软件层次也可以有不同的数据采集方法，更何况被审系统还存在不同层次上的结构差异。

1.2.1

操作系统（OS）层的采集方法

目前常见的操作系统（Operating System）有单用户、网络和多用户等几种。

1.2.1.1

被审系统是单机

1. 操作系统是单任务（如 DOS）系统，在被审系统单机上使用拷贝、备份命令或软件工具(如 PCTools)复制数据文件。
2. 操作系统是多任务（如 Windows）系统，在被审系统单机上，使用操作系统复制数据共享数据的方法(如 Windows 的复制、粘贴)，或启动备份应用程序复制数据文件。

1.2.1.2

被审系统是网络系统

1. 使用被审系统的一台工作站，按上述单机系统的办法复制数据文件副本。
2. 审计人员自带便携机（需安装有网卡）登录被审系统网络，从被审机器的共享（若不然，要求被审机器暂时将有关目录文件夹设置成共享属性）文件夹下复制数据文件副本到自带便携机硬盘。
3. 审计人员自带便携机登录联网后，映射被审系统有关盘成自带便携机的一个新逻辑盘，然后用拖拽、复制、粘贴或发送的方法复制数据文件副本。

1.2.1.3

被审系统是多用户系统

1. 将数据文件输出至指定软盘或磁带成为数据副本。
2. 将有关数据文件复制副本集中到一个目录，然后将此目录卸载至软盘或磁带。

1.2.2

数据库管理系统（DBMS）层采集方法

被审系统是在某个数据库管理系统环境下运行，或被审系统的机器上安装有可打开被审系统数据文件的数据库管理系统，或审计人员自带便携机上装有某个数据库管理系统，若审计人员可进入数据库管理系统，则可根据不同的系统类型采用不同的数据采集方法。本书将在第五章推荐微软的一个通用工具，可采集查询各种大小数据库数据。

1.2.2.1

数据库管理系统是小型桌面系统

1. 用被审系统 DBMS 提供的拷贝功能复制相同文件格式的数据副本。
2. 用被审系统 DBMS 提供的导出或另存为功能从被审系统内将数据文件格式转换为审计人员要求格式的数据副本。
3. 审计人员自带便携机与被审系统联网后，用自带便携机上数据库管理系统提供的数据导入功能，将被审数据文件的数据，由数据库管理系统自动转换并输入重建成本系统格式的数据副本。
4. 审计人员自带便携机与被审系统建立起 ODBC 连接后，将被审数据文件的数据在审计人员自己的数据库管理系统内转换数据文件格式并重建数据副本。

1.2.2.2

数据库管理系统是客户/服务器结构的大型数据库系统

1. 利用被审系统的工作站或自带便携机登录被审大型数据库系统，并利用 SQL 语言查询被审计系统数据库服务器上的数据。
2. 审计人员自带便携机登录被审大型数据库系统建立本机桌面 DBMS 系统与大型数据库的 ODBC 连接，在本系统内转换重建被审数据文件的数据副本。

1.2.3

应用程序接口 (API) 层的采集方法

应用程序接口 (Application Program Interface 简称 API) 层的采集方法就是俗称的编程方法, 象微软的 MS Query 和审计署驻南京特派员办事处开发的数据采集分析软件就属于此种方法。由于不同的数据库管理系统有不同的数据存储结构, 因而需要有各自相应的数据库引擎来完成用户的同一种服务。为了满足用户开发通用的数据库应用程序的需求, 微软公司提供了开放式数据库互连 (Open DataBase Connectivity 简称 ODBC) 和 OLE DB (Object Linking and Embedding Data Binding) 一些程序调用级的 API, 通过 Visual Basic (VB)、Visual C++ (VC) 等应用程序开发工具, 可使用 DAO, 数据控件, ADO 等 API 的简化封装对象模型, 请求各种异构数据库管理系统的各种服务, 访问各种异构数据库系统的数据文件, 包括复制其数据副本。在 Delphi、PowerBuilder 等其它开发工具中也提供了类似的方法。

1.2.3.1

开放式数据库互连 ODBC API

开放式数据库互连 (Open DataBase Connectivity 简称 ODBC) 应用程序接口 (API), 由一些程序调用级接口 (Call Level Interface 缩写 CLI) 函数组成, 被应用程序调用的 ODBC API 函数, 调用 ODBC 驱动程序 (DLL), 访问不同格式存储的 ODBC 数据源和实现对各种数据源请求的服务, 包括对源数据的查询和复制副本。

1.2.3.2

数据访问对象 DAO

数据访问对象 (Data Access Object 简称 DAO) 是定义在 ODBC 上的一组简化封装对象模型, 主要通过微软的 Jet 数据库引擎与 ODBC 接口, 实现 SQL 语法分析和数据操作转换。程序员只能通过对象的属性、事件、方法, 控制连接的数据源及其访问过程, 不及 ODBC API 那样灵活通用, 由于提供了对象的简单、直观、一致的界面, 显著地简化了直接调用 ODBC API 函数的编程。