审计署计算机审计中级培训系列教材

2016版

审计数据采集与分析

吴笑凡 曹洪泽 编著

清华大学出版社

审计署计算机审计中级培训系列教材

审计数据采集与分析

吴笑凡 曹洪泽 编著

清华大学出版社
北京

内 容 简 介

本书是近年来审计数据采集与分析实践的经验总结和理论提升。全书分三篇共9章，重点介绍审计数据采集与分析的理论和实务。第一篇是基础篇，主要介绍实施审计数据采集与分析的必备知识；第二篇是审计数据采集篇，主要介绍审计数据采集以及采集之后的清理、转换、验证等数据维护工作的技术方法；第三篇是审计数据分析篇，主要介绍审计数据分析的技术方法，重点阐述查询型分析、验证型分析和发掘型分析在审计中的应用。

本书对于从事政府审计、内部审计和社会审计工作的审计人员，学习研究审计数据采集与分析理论及应用的高校和科研院所的相关人员，以及对审计数据采集与分析感兴趣的其他人员有参考价值。

随书附带光盘提供了“审计数据采集分析2012”正式版软件，可供读者在学习和工作中使用。

图书在版编目(CIP)数据

审计数据采集与分析 / 吴笑凡，曹洪泽编著. —北京：清华大学出版社，2017
（审计署计算机审计中级培训系列教材）
ISBN 978-7-302-45682-7

Ⅰ. ①审… Ⅱ. ①吴… ②曹… Ⅲ. ①计算机应用－审计－数据采集－技术培训－教材 Ⅳ. ①F239.1

中国版本图书馆CIP数据核字(2016)第285837号

责任编辑：王　青
封面设计：何凤霞
责任校对：宋玉莲
责任印制：何　芊

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦A座　　邮　　编：100084
社 总 机：010-62770175　　邮　　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：三河市春园印刷有限公司
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张**：22.75　　**字　　数**：497千字
（附光盘1张）
版　　次：2017年1月第1版　　**印　　次**：2017年1月第1次印刷
印　　数：1～4000
定　　价：49.00元

产品编号：070216-01

前　言

本书按审计署计算机中级培训大纲对“审计数据采集与分析”课程的要求编写而成。是继2001年出版的《计算机审计数据采集与分析技术》、2010年出版的《计算机数据审计》之后再次由清华大学出版社在全国范围内发行的“审计数据采集与分析”课程的配套教材。

审计署在“十三五”国家审计工作发展规划中就审计信息化发展目标提出“总体分析、发现疑点、分散核实、系统研究”的数字化审计方式，进一步要求审计人员能够充分利用内部电子数据和外部电子数据，通过数据分析发现问题，更好地发挥审计的“免疫系统”功能。为了促进这一目标的实现，本书以介绍计算机审计实践的流程化方法“七步流程法”为切入点，紧密结合被审计单位信息化发展的实际，紧紧围绕审计数据采集——审计数据清理、转换、验证——审计数据分析这一主线，系统研究了审计数据采集与分析的内容和方法，理论与实践相结合，对实施计算机数据审计工作具有很好的指导作用。

本书按照审计数据采集与分析实务操作的逻辑来安排篇章结构，介绍的知识与审计数据采集与分析实务紧密结合。第一篇是基础篇，包含第1～3章，主要介绍计算机审计实践的流程化方法“七步流程法”，被审计单位对电子数据的存储技术和管理技术，审计接口原理、数据库访问技术、SQL语言等电子数据采集与分析技术原理，力求使读者掌握实施审计数据采集与分析工作的必备的基础知识。第二篇是审计数据采集篇，包含第4～5章，主要介绍审计数据采集、清理、转换、验证的策略、原因及技术方法，力求使读者对于审计数据采集、清理、转换、验证的全过程从理论上得到提高，在实践中获得指导。第三篇是审计数据分析篇，包含第6～9章，主要介绍审计数据分析的技术方法，重点阐述查询型分析、验证型分析和发掘型分析在审计中的应用，通过大量的分析实例，力求使读者掌握审计数据分析的内容、模型和主要技术方法。

本书有两大特色：一是通过介绍新版审计软件“审计数据采集分析2012”在实现采集XML文件、汉化数据库表结构、容错采集电子数据等方面的新功能，使在审计软件的帮助下切实提高审计数据采集与分析效率成为可能；二是本书每章都提供拓展阅读部分，方便不同基础的读者选择学习，也给学有余力的读者提供了进一步拓宽视野、强化学习成果的途径。

本书内容翔实、实用性强，力求为审计一线业务人员提供一本实用的培训教材，使他们能够快速掌握审计数据采集与分析的基础知识和操作技能，用本书介绍的技术和方法有效开展工作。本书穿插介绍了许多实用的专题，包括数据采集的容错方法、数据库表结构的批量汉化方法、Oracle数据库DMP文件恢复法、SAP ERP数据采集方法等，同时还对审计实践中经常用到的T-SQL函数、PL-SQL函数给出了详细的使用说明，目的是

使审计人员在实际工作中遇到问题时，可将本书当作一本备查的工具书，易学、易懂、易用，能够快速上手入门。

本书由吴笑凡、曹洪泽编写。全书成稿后由吴笑凡统稿，曹洪泽进行了修改。

审计署计算机技术中心杨蕴毅主任审定了全书。

神舟通用数据技术有限公司孙磊工程师为本书提供了技术支持，审计署南京特派办万建国、姜小舜、马社亮、杨海荣、安景琦、孔伟宁、余向阳、谢彭峰、高瞻、嵇旭珍、汪姬、沈晔华等同志为本书的编写提供了许多有益帮助，谨向他们表示衷心感谢！

由于审计数据采集与分析的理论和实践都在不断发展，加上作者的水平和经验有限，书中有些问题的研究尚不透彻，有些内容还有待于在审计实践中不断丰富和完善，有些观点还值得商榷，真诚希望广大读者批评指正！

联系方式：wxf_cnaonj@163.com。

作　者

2016年9月于北京

目　录

第一篇　基　础　篇

第二篇　审计数据采集篇

第三篇 审计数据分析篇

第一篇 基 础 篇

本篇共包含3章，主要介绍审计数据采集与分析概述、被审计单位的电子数据、电子数据采集与分析技术原理。本篇给出了实施审计数据采集与分析所必需的知识储备。

第1章 审计数据采集与分析概述

本章要点

- ◆ 七步流程法
  - 🕮 调查阶段
  - 🕮 数据采集
  - 🕮 数据清理、转换、验证
  - 🕮 建立审计中间表
  - 🕮 把握总体，选择重点
  - 🕮 建模分析
  - 🕮 疑点延伸、落实、取证
- ◆ 拓展阅读
  - 🕮 信息系统审计的安全性、有效性、经济性
  - 🕮 信息系统审计和数据审计辨析

1.1 引言

审计实务和计算机技术的结合，在中国是随着20世纪80年代末90年代初开始的会计信息化而逐步加深的。20世纪90年代信息技术突飞猛进，金融、财政、海关、税务等部门，民航、铁道、电力、石化等关系国计民生的重要行业开始广泛运用计算机、数据库、网络等信息技术进行管理，国家机关、企事业单位会计电算化趋向普及。被审计单位管理存储财政、财务信息的手段发生了质的变化，报表、账簿、凭证等审计人员熟悉的纸质资料以电子数据的形式重新组织管理，审计人员面临“进不了门、打不开账”的尴尬局面。

审计对象的信息化，客观上要求审计机关具备运用计算机技术，全面检查被审计单位经济活动的能力，发挥审计监督的应有作用。20 世纪末，时任审计署审计长的李金华同志指出“审计人员如果不掌握计算机将失去审计的资格”，首次将计算机审计上升到审计人员必备的专业技能的高度。

为了应对信息化对于传统审计的挑战，计算机审计应运而生。在世纪之交那个风云际会的年代，计算机审计要解决的最急迫的问题是实现协助审计人员审查被审计单位的电子账这个技术问题。

1999 年，审计署组织开发了“审计数据采集分析”[①]软件。该软件吸取了加拿大审计软件 ACL 和 IDEA 系统化的优点，是一个基于 Microsoft SQL Server 数据库管理系统的通用审计软件。“审计数据采集分析”软件成功实现了对电子账簿数据的获取，并支持对电子数据的查询、分析、统计和计算。运用该软件，审计人员在对被审计单位财务和业务等电子数据的审计过程中逐渐掌握了主动。

在实现财务电子数据审查的同时，新的问题也随之出现：仅审查被审计单位的财务数据是否足够，业务电子数据怎么办？需不需要审查？如何审查？举个例子，当审计人员对中国铁路总公司（即原铁道部）进行审计时，当需要反映货运系统中罐车、敞车等车皮调度的真实性、合理性时，仅审查财务数据是不够的，对于货运系统中的业务电子数据的审查是必不可少的。因此，计算机审计的范畴从财务数据的审查拓宽到对财务数据和业务数据都要进行审查。这也就是计算机审计的第一层含义，对电子数据的审计，或者也可以称为计算机数据审计。

目前，计算机审计实务主要包括计算机数据审计和信息系统审计。

计算机数据审计是指运用计算机审计技术对信息系统中所存储和处理的电子数据进行的审计，这些信息系统不但包括被审计单位与财政收支、财务收支有关的计算机信息系统，还包括来自被审计单位之外的，与被审计单位有关的，从不同的角度体现被审计单位实际经营情况的方方面面的信息，如来自工商系统、国税系统、地税系统、电力系统、社保系统等信息系统记录的被审计单位的工商注册情况、纳税申报情况、生产用电情况、为职工缴纳社保情况等。计算机数据审计从技术上讲主要包括通过对来自被审计单位内部和外部相关的电子数据进行采集、转换、清理、验证和分析，帮助审计人员掌握总体情况，发现审计线索，收集审计证据，形成审计结论，实现审计目标。

信息系统审计是通过对被审计单位信息系统的调查和了解，对系统控制及系统功能的分析和测评，综合评价信息系统能否满足安全性、有效性和经济性的目标，促进被审计单位完善信息系统控制，防范业务风险；加强信息系统运行管理，规范业务行为；促进提高信息系统运行效率和投资效益。信息系统审计的更多知识请参看本章的拓展阅读部分 1.3.1 节的相关内容。

本书的重点在于阐述计算机数据审计的两个重要组成部分：审计数据采集和审计数据分析。

① “审计数据采集分析”历经 1.0、2.0、3.0 三个版本的升级，目前最新版是“审计数据采集分析 2012”。

审计数据采集，指的是宏观意义上的采集，它既包括将来自外部的电子数据采集到审计人员的计算机系统中，也包括电子数据获取之前对于需要获取的存储电子数据的信息系统及使用信息系统的单位的调查研究，还包括电子数据获取之后对于获取到的电子数据进行数据清理、数据转换、数据验证，以提高数据质量，并生成适合审计数据分析的审计中间表。

审计数据分析，指的也是宏观意义上的分析，它从已建立的审计中间表入手，根据审计中间表的内容，对审计目标进行总体分析，把握被审计单位的总体情况，找准薄弱环节，确定审计重点；对各审计重点建立分析模型，撰写分析语句，发现审计疑点；最后将发现的疑点交给审计核实组去审计现场完成对疑点的延伸、落实和取证工作。

从计算机数据审计的流程实现的角度上讲，审计数据采集和审计数据分析的具体内容可以概括为七个阶段的工作，简称“七步流程法”。这七个阶段的工作如下。

第一阶段：调查阶段；

第二阶段：数据采集；

第三阶段：数据清理、转换、验证；

第四阶段：建立审计中间表；

第五阶段：把握总体，选择重点；

第六阶段：建模分析；

第七阶段：疑点延伸、落实、取证。

1.2　七步流程法

“七步流程法”是计算机数据审计的一套完整的程序和流程。

1.2.1　调查阶段

调查阶段，审计人员应根据审计工作方案中审计目标的要求收集被审计单位相关业务系统的技术文档，以详细了解系统的数据库和数据情况，并在此基础上提出审计数据需求说明书交由被审计单位提供数据。

在调查阶段，应当了解被审计单位的组织结构，掌握计算机系统在组织内部的总体应用情况。根据审计目标，选择那些对实现审计目标有重要影响的计算机信息系统（即信息系统对被审计单位业务支持程度高，被审计单位业务对信息系统的依赖程度高）作为深入调查的子系统或功能，进行全面、详细的调查了解。调查了解的内容应包括软硬件情况、系统的开发情况和有关技术文档，系统的运行、维护、配置、管理情况，系统的功能、数据情况等。还可以包括业务系统的名称、版本、开发商、功能等内容，特别是业务系统自身是否有将业务数据导出的功能，如果可以导出，应了解导出文件的数据格式。

对计算机数据审计而言，数据情况调查是重点。

数据情况调查的主要任务是搞清楚被审计单位各业务系统存储和处理电子数据的基本情况。调查工作的一般思路是：审计目标→审计内容与重点→审计内容所涉及的信

息系统→与信息系统相关的电子数据。

数据情况调查是对信息系统产生的电子数据进行全面、深入认识和了解的过程，因此必须首先调查了解信息系统的业务流程。要详细了解被审计单位业务环节的具体操作方式和目的，必要时可以根据了解的情况绘制业务流程图，目的是使审计人员有一个初步的审计思路，更好地设计切实可行的审计方案，同时初步确定数据采集的范围。

在理解业务流程的基础上，还应进一步理解数据流程，从而了解系统应该生成哪些数据以及数据的生成过程、来源、去向等情况，为提出数据需求和下一步的数据分析工作打下基础。

在数据情况调查过程中，审计人员应将相关的技术文档尽量收集齐全，以便详细了解系统的数据库和数据情况。数据库及数据的说明信息都包含在这些技术文档中。审计人员应当首先阅读各种设计说明书，了解数据库总体结构，包括数据库总体布局、各级服务器上的数据库系统的名称、版本、内容、各数据库之间的关系等总体情况。在此基础上进一步了解数据库中数据表的具体情况等内容，以便根据审计需求确定从哪一层次采集数据、采集哪些数据库的数据，甚至可确定采集哪个数据库中的哪些表。通过对表间关系和表结构描述的了解，可以为后续的数据转换和创建审计中间表工作打好基础。

在上述工作的基础上，应提出审计数据需求说明书交由被审计单位提供数据。审计数据需求说明书中应指定数据采集的系统名称、数据库名称（必要时可以精确到需要采集的数据表的名称）、数据采集的具体方式、数据传输的格式、所需数据的时间段、数据交付的方式、数据交付的期限和其他注意事项等内容。

1.2.2 数据采集

数据采集，在调查阶段提出数据需求的基础上，按照审计目标，采用一定的工具和方法对被审计单位信息系统中的电子数据以及来自被审计单位之外的，与被审计单位有关的，从不同的角度体现被审计单位实际经营情况的电子数据进行采集。

数据采集可分两步实现：第一步是通过调查阶段掌握的信息来选择需要采集的电子数据；第二步是通过一定的技术和手段实现对目标数据的采集，及时获取全面、完整的电子数据。常用的数据采集策略有三种：一是通过数据接口采集；二是直接复制；三是通过备份文件恢复。

对被审计单位信息系统中的电子数据的采集一般需要在被审计单位技术人员的配合支持下完成。对来自被审计单位之外的，与被审计单位有关的，从不同的角度体现被审计单位实际经营情况的电子数据，如果审计单位的审计数据分中心已经积累了这部分的电子数据，可以由审计组向本单位提出申请，调出这部分电子数据；如果尚未掌握这部分电子数据，应通过正规途径尽可能协调相关单位，获取这部分电子数据。

1.2.3 数据清理、转换、验证

在实际工作中，经常需要将来源众多的被审计单位电子数据集成到一起进行分析和处理。由于被审计单位数据来源繁杂，数据格式不统一，信息表示代码化，数据在采集和

处理过程中可能失真，被审计单位可能有意更改、隐瞒数据真实情况等诸多影响因素，对采集到的电子数据必须进行清理、转换、验证，使得数据能为审计所用。

数据清理是指为提高数据质量而对缺失的、不准确的、不一致的有质量问题的电子数据进行处理。通常，通过重复记录的识别与清理、缺失数据的补充、空值处理等操作可以提高数据质量。

数据转换包括数据库格式的转换以及数据内容的转换。数据库格式的转换要求将采集到的来源于被审计单位不同类型数据库格式的电子数据统一存储为一种数据库格式。数据内容转换要求对采集到的原始数据的含义进行识别，明确地标识出每张表、每个字段的经济含义及其相互之间的关系。通常，通过数据类型转换、日期/时间格式的转换、代码转换、值域转换可以实现数据内容的转换。

数据验证贯穿数据采集、数据清理、数据转换的全过程。

在数据采集阶段，数据验证要检查被审计单位提供资料的完整性，保证数据采集工作准确有效地进行，同时对采集到的数据进行确认，排除遗漏和失误。

在数据清理阶段，数据验证要确认数据清理工作没有损害数据整体的完整性和正确性。

在数据转换阶段，审计人员会将原始电子数据中表名、字段名、记录值代码以及表表关联的经济含义明确标识出来，这需要进行大量的查询、替换修改、插入数据、更新数据、删除数据等操作，每一步转换工作都有可能影响数据的完整性和正确性，因此应在这一阶段进行数据验证，以确保数据转换工作未对电子数据引入新的错误。

1.2.4　建立审计中间表

建立审计中间表，对获得的电子数据按审计需求进行投影、连接等“再加工”，从电子数据中选择满足审计需要的精简的数据集合，生成审计中间表，为建立审计分析模型形成数据基础。

审计中间表是利用被审计单位数据库中的基础电子数据，按照审计人员的审计要求，由审计人员构建，可供审计人员进行数据分析的新型审计工具。

通常在设计数据库的时候要对数据进行范式分解。范式分解这种规范化过程会将描述一个业务对象的数据分解成关系数据库中的多个逻辑表，这些表之间存在一定的关联关系。要利用被审计单位数据库中的数据来实现审计分析，必须对清理、转换、验证后的电子数据按审计需求进行投影、连接等“再加工”，从电子数据中选择出满足审计需要的较为精简的数据集合，生成一系列中间数据表——审计中间表。审计中间表是审计人员建立审计分析模型的基础。

建立审计数据中间表是一个循序渐进的过程。在对数据进行清理、转换、验证后，就应考虑设计出初步的审计中间表，此时的主要工作是帮助审计人员选定审计所需的基础性数据，如去掉与审计无关的字段、建立表与表之间的基本连接等。这一阶段创立的中间表称为基础性中间表。在建立分析模型进行具体的数据分析时，还要建立分析性中间表，即按照审计分析模型，对基础性中间表再进行字段选择、连接等处理，以帮助审计人

员实现对数据的建模分析。

1.2.5 把握总体，选择重点

把握总体，选择重点是指根据审计目标进行总体分析，把握被审计单位的总体情况，找准薄弱环节，确定审计重点，建立审计分析模型以形成较具体的理论指导。

在总体分析时，可以从不同层次、不同角度对被审计单位的电子数据进行汇总、核对与分析，如进行账表核对、表表核对，也可以建立指标或指标体系进行分析，还可以使用多维分析工具从不同的层次和角度来观察被审计单位的电子数据。这些总体分析的技术和方法可以帮助审计人员把握被审计单位有关经济业务的总体情况，寻找薄弱环节，确定审计重点，避免审计工作的片面性和盲目性。

1.2.6 建模分析

建模分析，在总体分析的基础上，审计人员需要根据确定的审计重点，利用审计分析模型进行具体的数据分析。审计分析模型是审计人员用于数据分析的技术工具，它是按照审计事项应该具有的时间或空间状态（如趋势、结构、关系等），由审计人员通过设定判断和限制条件建立起来的一系列数学的或逻辑的表达式，并用于验证审计事项实际的时间或空间状态的技术方法。在实际工作中，要根据不同层次的审计需求建立不同层次的审计分析模型。一般而言，有总体分析模型、类别分析模型和个体分析模型。

目前常见的审计分析模型包括根据法律、法规和制度规定的状态和关系来建立；根据业务的逻辑关系来建立；根据不同类型数据之间的对应关系来建立；根据审计人员的符合客观实际的经验来建立；根据审计人员的合理的预测来建立等。

1.2.7 疑点延伸、落实、取证

根据审计分析模型的分析结果，派出重要问题核实组、重大违法违纪违规问题突破组等到被审计单位对疑点进行延伸，对发现的问题线索进一步核查、落实并取证。

如果数据分析的结果能直接发现与核实问题，审计人员可以利用有关电子数据直接取证。这时审计人员应将被审计单位提供的原始数据、分析处理产生的中间表数据以及数据分析和处理的过程的语句代码妥善保存，以便作为审计证据。如果数据分析的结果仅能揭示问题的线索，不能直接发现与核实问题，则应根据线索进行延伸审计，获取审计证据。在编制审计工作底稿时，应对数据分析的过程、方法、使用的数据等情况进行详细记录，同时还应记录审计人员对数据分析结果的判断。

1.3 拓展阅读

本章的拓展阅读包括两个部分：一是对信息系统审计的介绍；二是应用“七步流程法”开展计算机数据审计的示例。

1.3.1　信息系统审计

信息系统审计是通过对被审计单位信息系统的调查和了解，对系统控制及系统功能的分析和测评，综合评价信息系统能否满足安全性、有效性和经济性的目标，促进被审计单位完善信息系统控制，防范业务风险；加强信息系统运行管理，规范业务行为；促进提高信息系统运行效率和投资效益。

以下介绍对信息系统安全性、有效性、经济性检查的建议和常用方法。

1. 检查信息系统的安全性

检查的对象包括：局域网、广域网、因特网接入网等网络；服务器、存储、网络及安全、计算机终端等设备；操作系统、中间件、数据库、应用等软件；机房、电源、空调、安防等条件环境。针对这些系统要素，检查被审计单位是否制定了安全管理制度并得到贯彻执行；信息系统设计是否符合国家等级保护、分级保护要求；对关键的数据信息资源是否采取了适当的保护措施；数据信息物理访问、逻辑访问安全措施是否适当有效；系统安全员、安全审计员履行职责是否到位；防病毒木马、防网络攻击等安全措施是否落实；防火、防雷、防盗等安全设施是否有效。

检查信息系统的安全性要突出重点，讲求实效，根据系统的规模大小、被审计单位业务对信息系统的依赖程度，评估其采取的安全措施，评价安全缺陷可能造成的损害，实事求是地提出审计整改意见。

2. 检查信息系统的有效性

检查的对象包括：业务流程还原或者再造、关键节点控制等规范管理的功能；避免重复操作、自动批量处理等提高效率的功能；数据同源，消除空间局限、实现远程处理等共享信息的功能；接口及逻辑处理正确、真实记录经济活动等数据准确的功能。针对这些功能要素，检查被审计单位系统软件、关键业务应用软件功能是否能够满足业务开展的需要；信息系统日常操作管理、配置管理、问题管理等是否有效；数据、参数的生成、存储、传输、处理等是否进行适当有效的控制；不同系统之间是否建立有效控制的数据接口；检查是否存在基础信息混乱、信息无法共享等问题；检查信息系统灾难恢复计划是否能够保证关键业务的持续运行，电子数据备份是否符合相关规定，是否能够保证数据及时、准确地恢复。

检查信息系统的有效性要紧密结合审计目标，突出重点，注意对业务流程和关键控制节点进行分析，从中查找信息系统有效性可能存在的缺陷，防患于未然，发挥“免疫系统”功能。在数据审计中发现异常时，要注意从信息系统功能中排查原因，揭示信息系统舞弊，打击利用信息系统漏洞的违法犯罪活动。对于信息系统功能存在缺陷，有碍被审计单位管理制度落实，甚至危及信息安全、经济安全的，要提出整改要求。

3. 检查信息系统的经济性

检查的对象包括：项目建议书、可行性研究报告、初步设计等项目建设文档；实施、验收过程；建设投入；使用状况；满意度等。针对这些项目要素，检查项目建设文档确定的建设目标与被审计单位业务的符合程度，建设目标能否带来合理的回报或推动业务发