《基于非关系数据库的电子档案存储规范研究及系统实现》

研究报告

国家档案局档案科学技术研究所

2016年12月

目　录

1 概述

1.1 非结构化数据存储研究的必要性

1.1.1 非结构化数据存储研究是大数据时代的必然要求

云计算、物联网、社交网络等新兴服务促使人类社会的数据种类和规模正以前所未有的速度增长，大数据时代正式到来。大数据时代的特征之一就是数据种类繁多，数以千计，这些数据包含结构化、半结构化以及非结构化的数据，并且非结构化数据所占份额越来越大，据统计，其中 85%是非结构化数据。非结构化数据就是不能用数字或者统一的结构表示或没有固定结构，且不能用平面行列表格结构存放的数据，如不同格式的办公文档、文本、图片、HTML、各类报表、图像和音视频信息等。

2013 年，联合国以名为“Global Pulse”的倡议项目发布了《大数据发展：挑战和机遇》（Big data for development: Challenges and Opportunities）的报告，该报告主要阐述了大数据时代各国特别是发展中国家遭遇数据洪流时所面临的机遇与挑战。

如今，大数据的来源非常广，按产生的场景分，大数据基本上分为社交型和内务型两类。社交型大数据指的是社会化的，如互联网、物联网、移动互联网、社交网等网络平台与网民交互活动所产生的数据。内务型大数据是指部门、行业内部业务过程中产生的数据，主要发生在机关、团体、企事业单位和其他组织。

2015 年，国务院发布了《国务院办公厅关于运用大数据加强对市场主体服务和监管的若干意见》（国办发[2015]51 号），文件指出了加强大数据运用对维护国家统一、提升国家综合治理能力、提高经济社会运行效率的重大意义，要求全国各级政府部门充分运用大数据先进理念、技术和资源，加强对市场主体的服务和监管，推进简政放权和政府职能转变，提高政府治理能力。

可见，大数据正式上升为与历史上的互联网、超级计算同等重要的国家战略，因此，对非结构化数据存储研究已不容回避。

1.1.2 非结构化数据存储研究是档案行业的内在要求

2012 年 8 月 29 日国家电子文件管理部际联席会议第二次会议和国家档案局局务会议审议通过的《电子档案移交与接收办法》附件 2 规定了电子档案存储结构要求，从要求可以看出该办法并没有采用关系数据库接口做档案数据迁移方式，

而是采用了将电子档案数据按文件夹的方式进行移交和接收。另外，全国副省级市以上综合档案馆已数字化的档案占馆藏总量的比例越来越高，数据量急剧增加，仅中国第一历史档案馆的数字化量就已经达到 3800TB。可见，在档案馆的数字资源中，非结构化数据既有来自办公自动化环境下直接生成的电子文件，也有档案数字化形成的数据随着数字档案馆建设的不断推进，档案馆数字资源中的非结构化数据将会出现爆炸性增长。如何有效存储、管理、利用非结构化数据是档案工作者必须面对的课题。

1.1.3 非结构化数据存储研究是改善电子档案存储现状的迫切要求

档案系统目前对非结构化信息资源的存储，基本上采用如下四种方式，一是直接存储。通过文件系统直接存储在文件服务器上。由于传统的文件系统管理数据有很多缺陷，如数据共享性差、冗余度增加、数据和应用程序过分依赖，因此这种方式缺乏对数据的统一管理和控制。也有以 FTP 上传的方式直接存储在文件服务器或文件服务器外挂的磁盘阵列中，这种方式扩容比较难，管理效率低下。二是数据挂接数据库。一般挂接的数据是电子档案原文，数据库是关系数据库。挂接的方法是将数据的存储路径和名称存入数据库条目的某字段中，将挂接的数据存储在该存储路径下，使用时根据数据的存储路径和名称去查找和打开该数据。由于数据库和数据的关系松散，当数据库或数据的存储位置发生迁移时，会造成挂接关系断裂，从而造成存在的数据找不到。三是离线存址。一般离线存址的数据是电子档案原文，数据库是关系数据库。离线存址的方法是将数据存储在移动存储介质上，对数据所在的移动存储介质进行编号，将数据的离线存址存入数据库条目的离线存址字段中，将数据的名称存入数据库条目的数据名称的字段中。使用时根据数据名称和离线存址查找数据所在的移动存储介质的编号，再根据移动存储介质的编号找到数据所在的移动存储介质。四是将需要挂接的数据嵌入数据库。一般使用的数据库是关系数据库。该方法是将数据直接存储在数据库条目的某字段中，找到条目可立即打开和读取该数据。该方法不会造成挂接关系断裂，但是由于大量数据的嵌入，会使关系数据库的容量急剧膨胀，从而造成数据库运行速度的骤降，甚至宕机（即“死机”）。这四种方式都有明显的不足，迫切需要研究针对非结构化电子档案的更高效、安全的存储方式。

1.1.4 绿色节能存储是海量电子档案存储的现实要求

绿色存储是世界存储产业的重要发展方向。目前磁存储仍是存储数据的主流技术，其特点是存储系统容量大，存储和读取速度快，但耗能高，碳排放量大、易受磁冲击，安全性差。据日本权威企业调查，磁存储的空调耗电量占总耗电量40%以上。仅耗电量这一项支出就使档案馆不堪负重。光盘在节能减排方面具有天然的优势，但在非结构化数据处理、管理、检索和查找以及提取方面存在很大差距。综合利用磁光电存储技术，优势互补，在"安全存储、绿色存储、长寿命存储"的概念内，寻找与地球资源和环境的可承受力相适应的海量数据存储解决方案是档案馆迫切的现实要求。

1.2 非结构化数据存储研究的可行性

1.2.1 非关系数据库技术的发展应用提供了技术支撑

传统的关系数据库能够很好地管理结构化数据，但在管理非结构化数据时暴露出越来越明显的局限性，特别是在非结构化数据迅速膨胀到 PB 级后表现出了性能与可靠性问题。为了解决非结构化数据增长带来的问题与挑战，许多研究者在非关系数据库模型的设计上投入了大量的精力，促使非关系数据库成为了今后数据库技术的研究热点以及发展方向。现有的 NoSQL 已经超过 150 种，主流的 NoSQL 数据库有 BigTable 类数据库（HBase，Cassandra）、SimpleDB、MongoDB、CouchDB、Redis 以及 TRIP。TRIP 系统是一个非常成熟的非关系数据库，同时也是采用 Hash 表的全文搜索引擎，是 NoSQL 与搜索引擎两者的无缝集成系统，能够满足大规模电子档案存储、检索、利用的基本要求。

1.2.2 光盘技术的发展和光盘库容量的升级提供了硬件基础

光盘技术不断发展，单片光盘容量不断增加，寿命不断延长。光盘存储技术是 20 世纪 70 年代初发展起来的一项高新技术，从早期 CD 光盘容量 640MB, DVD 光盘容量 4.7GB，发展到 BD 蓝光光盘 25GB，其容量已经是 DVD 的 5 倍，一张单层蓝光光盘可以存储 25GB 的文档文件（例如，约 75 万页 word 文档），双层蓝光光盘 BD 容量达到 50GB，3 层 100GB 蓝光光盘 BD 已经量产。光盘记录信息的记录层由 DVD 的有机材料转变为 BD 的无机材料，实验表明，高质量的蓝光光盘 BD 保存数据的寿命达到 50 年以上。

光盘库容量不断升级。光盘库是一种带有自动换盘器的光盘网络共享设备。提高光存储系统的容量有三种方式，一是增加系统中光盘库的数量，二是增加光盘库容纳光盘的数量，三是增加单张光盘的存储容量。随着大型光盘库的不断问世，单个光盘库中放置光盘的数量达到 1000-10000 张，加之光盘容量的增加，光盘库的容量体密度逐渐达到与磁带库相同的水平，如放置 100GB 蓝光光盘的单个光盘库容量可达到 100TB-1PB。

1.3 研究目标与内容

1.3.1 研究目标

a）将非关系数据库技术运用于电子档案存储，通过将电子档案存储在非关系数据库，改变电子档案挂接在数据库上的方式。

b）将非关系数据库存储电子档案的技术与大容量光存储技术相融合，通过在大容量光盘上建立存储电子档案的非关系数据库，从根本上改变光盘存储电子档案和检索的方式，实现光盘上电子档案的全文检索。

c）将关系数据库技术与非关系数据库技术相结合，实现现有电子档案的关系数据库系统与非关系数据库的连接跳转。

1.3.2 研究内容

a）非关系数据库存储电子档案的研究及实现

研究非关系数据库存储电子档案的方法，将电子档案及其元数据全部存储在非关系数据库，不采用存储路径的方式将电子档案挂接在数据库上的方式。

b）非关系数据库的功能研究及实现

研究非关系数据库拆分和合并的技术，利用非关系数据库对电子档案进行分类和汇聚整合。

c）非关系数据库建立在光盘上的研究及实现

研究将非关系数据库建立在光盘上的技术，利用数据库拆分技术将大容量数据库建立在光盘上。

d）磁盘和光盘上数据库管理和访问的研究及实现

研究统一管理大规模存储在磁盘和光盘上数据库的技术，研究磁盘和光盘上数据库的检索机制，包括全文检索、二次检索、同构多库和异构多库检索、多（光）

盘多库检索、磁（盘）光（盘）多库检索、异构类型数据库检索等。

e) 异构类型数据库存储电子档案的研究及实现

研究关系数据库与非关系数据库相结合的技术，解决电子档案中的结构化数据与非结构化数据通过单一类型数据库进行数据处理时导致数据处理性能下降的技术问题，解决电子档案中的不同结构类型数据无法结合数据库类型进行数据存储的技术问题。

f）基于非关系数据库的电子档案存储规范的研究

研究采用非关系数据库在磁盘和蓝光光盘上存储电子档案和档案数字副本的方法，主要解决非结构化档案数据的存储问题。

g）非关系数据库存储电子档案的系统实现

开发数据库管理、数据管理、数据库拆分与合并、电子档案汇聚整合、磁盘和光盘上数据库检索的应用软件。

2 自主研发的关键技术

2.1 基于非关系数据库的电子档案存储技术

本课题研发的关键技术是将非关系数据库技术运用于电子档案存储。非关系数据库采用子字段、多值字段和变长字段机制，创建不同类型的非结构化的或任意格式的字段，以多维结构的处理方式突破关系数据库二维表结构，实现从数据属性管理转化为内容管理的数据库，能够存储各种格式的结构化数据、半结构化数据、非结构化数据。

电子档案一般包括著录信息(结构化数据和非结构化数据)和电子档案原文(非结构化数据）。采用非关系数据库能够存储著录信息和各种格式的电子档案原文，既不采用存储路径的方式将电子档案挂接在数据库上的方式，也不采用将需要挂接的电子档案嵌入数据库的方式，这将改变电子档案挂接或嵌入数据库的传统模式，保证电子档案的安全性和完整性。

2.1.1 非关系数据库的数据库主文件结构

非关系数据库的数据库主文件包括控制块、记录号索引、空位表和数据块。控制块中存放着记录号、记录号对应的字段，经过索引的最高记录号、上次索引后至今未做过修改或新增记录的记录号，以及记录号索引和空位表的地址等。记

录号索引由许多标准存储单元组成，每标准存储单元包含着指针，指向每个记录的最新版和该版生成/修改的日期及时间。空位表保存每个标准存储单元中还有多少空间可以使用。数据块由标准存储单元组成，是存放各记录字段数据本身的地方，每块连续地存储着整个记录，或记录的一部分。数据库主文件的物理结构示意图见图 2.1。

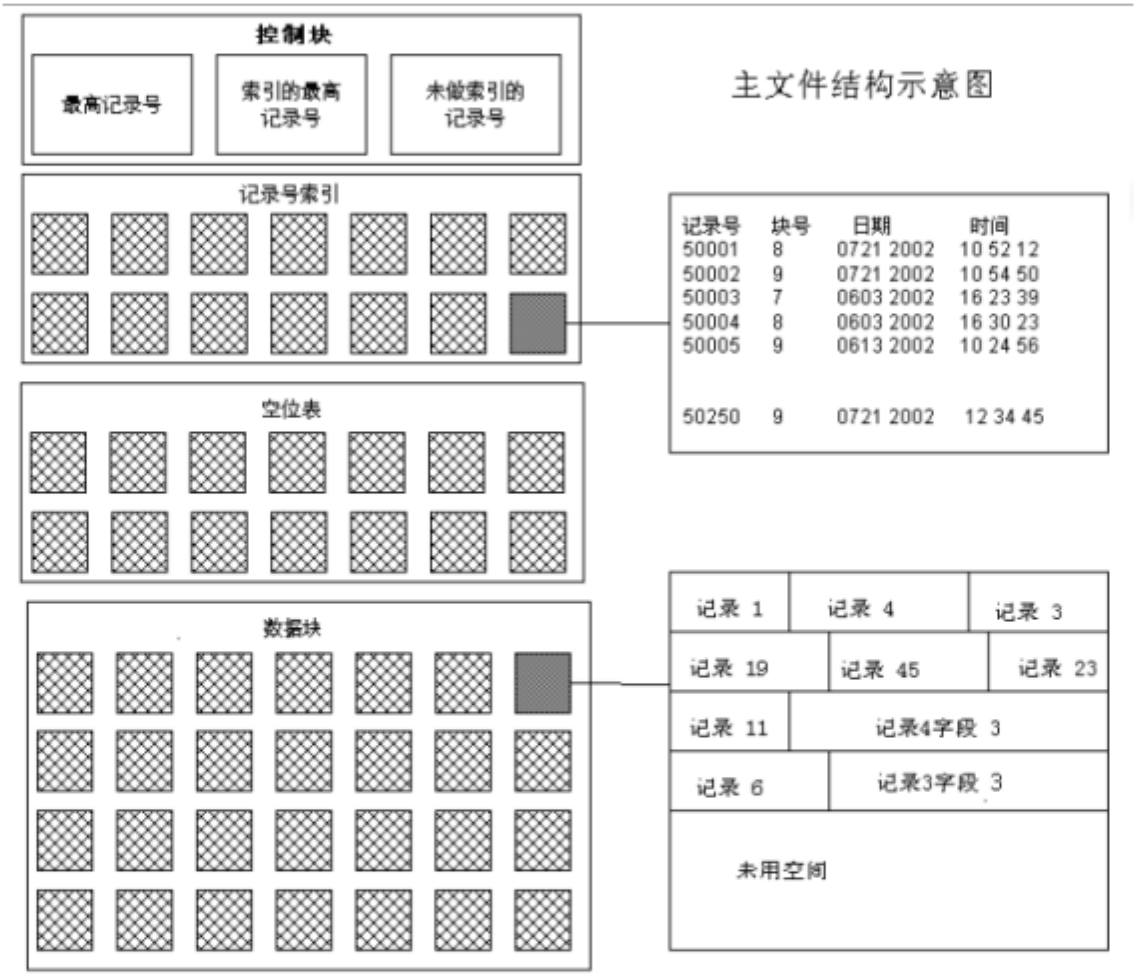

图 2.1 数据库主文件的物理结构示意图

2.1.2 字段的设计与定义

在非关系数据库结构描述中，有 7 种字段的数据类型：

a）词组型字段以固定长度的空间存储字符数不超过 256 个文字型数据表示。

b）文本型字段以可变长度的空间存储文字型数据表示，按段落、句子、词格式编写的自由文体，字符数没有限制。文本型字段可用于存储文本数据、从电子文件中抽取出来的字符。

c）整数型字段用于存储整数，整数范围为：-2,147,483,647 至 2,147,483,647。

d）数值型字段用于存储整数和实数，数值范围为：-1.7E+37 至 1.7E+37。

e）日期型字段用于存储公元纪年。

f）时间型字段用于存储时间。

g）二进制型字段用于存储非结构化数据，包括字处理文件、图像文件、图形文件、音视频文件、多媒体文件、数据库文件、计算机程序、数据文件、超文本文件等。

词组、整数、数值、日期、时间型字段由子字段组成，子字段的数量无限制。文本型字段下分段落，段落下分句子，句子下分词，段落的数量、句子的数量、词的数量无限制。

词组型、整数型、数值型、日期型、时间型字段用于存储结构化数据，文本型、二进制型字段用于存储非结构化数据。

电子档案存储在数据库一般包括三种数据类型的字段：词组型、文本型、二进制型字段。词组型字段用于存储电子档案名，文本型字段用于存储从电子档案中抽取出来的字符，二进制型字段用于存储电子档案原文。

2.1.3 电子档案存储方式

非关系数据库的主文件和索引文件是两个独立的数据库文件，可以分别存储在不同的存储介质和不同的位置。数据库主文件存储在磁盘（见图 2.2）或光盘上（见图 2.3 和图 2.4）；数据库索引文件存储在磁盘上（见图 2.2 和图 2.3），或存储在光盘上（见图 2.4）。由于单份超大计算机文件（电子档案原文）无法装入非关系数据库，因此，本课题研发了数据库双核存储系统，以解决此问题（见 2.2 节 数据库双核存储系统）。本课题是运用非关系数据库技术和数据库双核存储技术存储电子档案。

数据库主文件存储电子档案和从电子档案原文中抽取出的字符。每一条电子档案记录由著录信息、电子档案原文名、从电子档案原文中抽取出的字符、电子档案原文组成（见图 2.5 和图 2.6）。索引文件存储电子档案记录全部词条的索引信息。

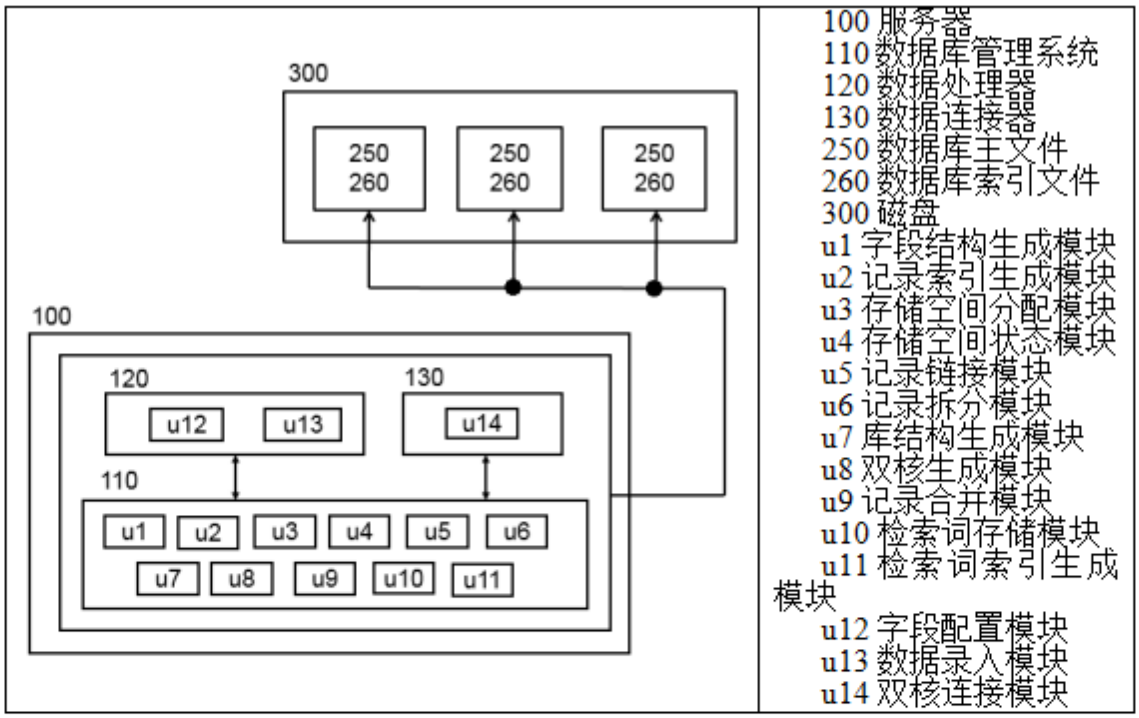

图 2.2　基于磁盘的数据库存储系统结构示意图

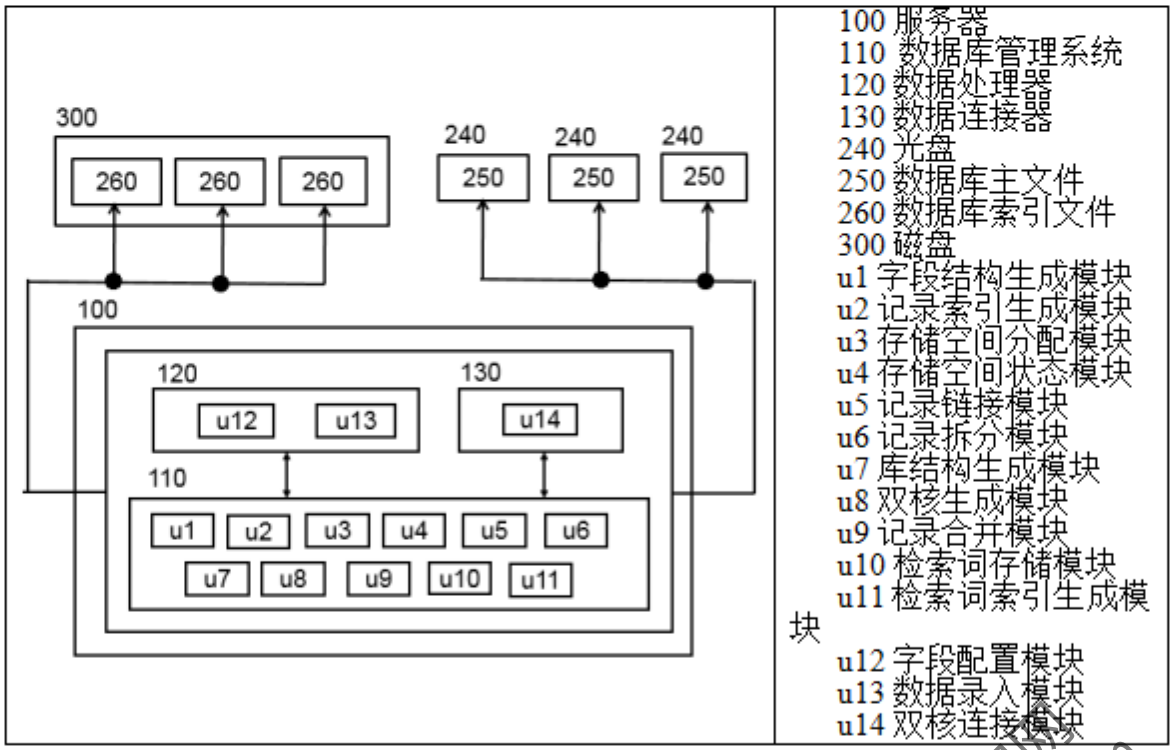

图 2.3　基于光盘的数据库存储系统结构示意图（索引文件在磁盘上）

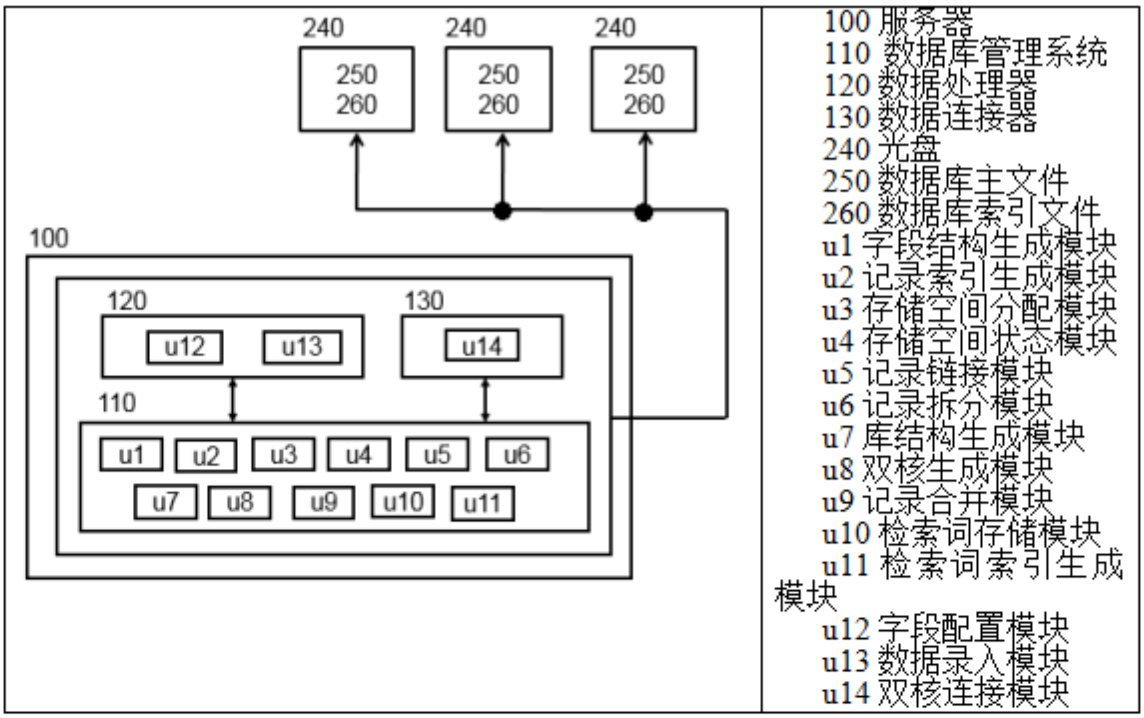

图 2.4 基于光盘的数据库存储系统结构示意图（索引文件在光盘上）

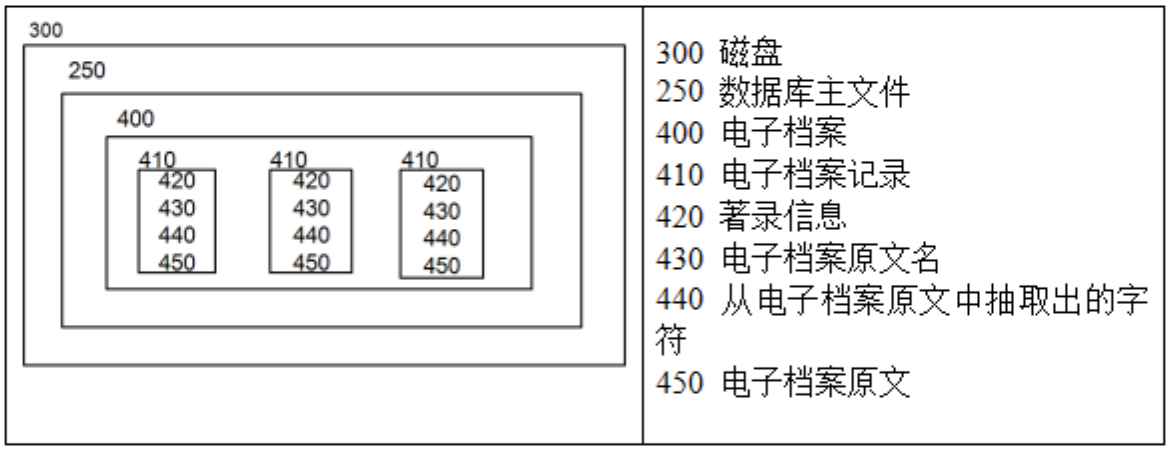

图 2.5 基于磁盘的数据库主文件存储电子档案结构示意图

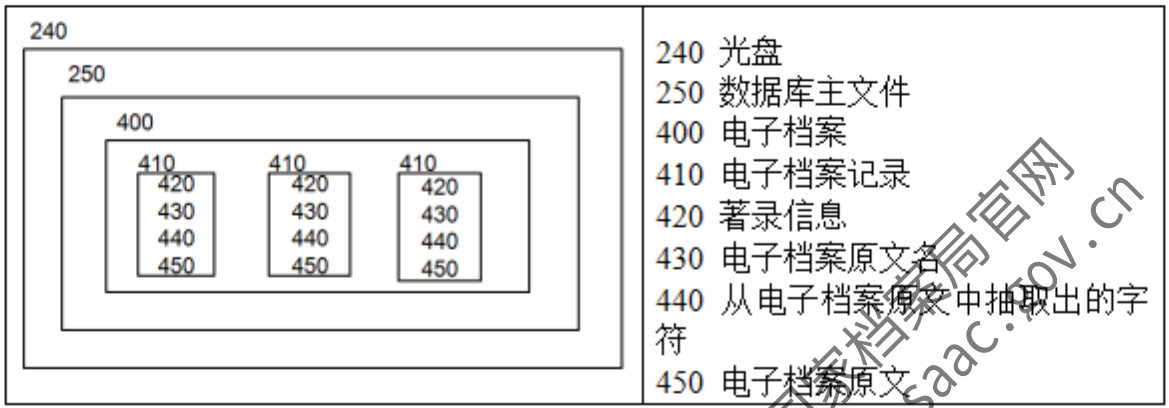

图 2.6 基于光盘的数据库主文件存储电子档案结构示意图

2.1.4 非关系数据库管理系统

非关系数据库存储系统包括服务器、磁盘和光盘。服务器与磁盘间建立数据连接，服务器与光盘间建立数据连接。数据库管理系统安装在服务器上。非关系数据库管理系统包括字段结构生成模块、记录索引生成模块、存储空间分配模块、存储空间状态模块、记录链接模块、记录拆分模块、光盘建库与录入模块、库结构生成模块、记录合并模块、检索词存储模块、检索词索引生成模块，其中：

字段结构生成模块，读取数据库中每一条记录的字段结构信息，包括记录中各字段的数据类型、长度，并写入数据库管理系统中。

记录索引生成模块，记录数据库中每一条记录的索引信息，包括记录中各字段的修改时间、修改内容，并写入数据库管理系统。

存储空间分配模块，记录为每一条记录所分配的标准存储单元在数据库文件中的位置信息，并写入数据库管理系统。

存储空间状态模块，记录数据库文件中已分配标准存储单元中未使用的空间信息，并写入数据库管理系统。

记录链接模块，将数据库中各记录的字段结构信息、索引信息、标准存储单元的位置信息和空间信息合并，形成数据库特征数据，并写入数据库管理系统。

记录拆分模块，根据数据库管理系统指令进行数据库中记录的拆分，数据拆分以记录为单位进行，读取数据库特征数据，确定记录中每个字段的数据位置和数据量，标记出符合指令参数的记录，并将标记信息写入数据库管理系统。

光盘建库与录入模块，用于将磁盘数据库中与光盘容量匹配的记录写入数据库管理系统在光盘上建立的子数据库文件，或在光盘上直接建立数据库文件，并将数据直接录入到光盘上的数据库。

库结构生成模块将数据库的库结构形成独立数据文件，数据库管理系统根据独立数据文件在磁盘建立相同库结构的数据库，或在光盘上建立相同库结构的子数据库。

记录合并模块根据数据库管理系统指令，将光盘上子数据库中的记录合并到磁盘存储装置上的磁盘数据库中，并通过数据库管理系统生成对应记录的字段结构信息、索引信息，以及记录标准存储单元位置信息和空间信息，形成磁盘数据库的数据库特征数据。

检索词存储模块，用于存储包含语义信息的检索词字库，检索词至少包括字、词和数字。

检索词索引生成模块，根据数据库中每一条记录的索引信息，建立与标记信息对应的记录的检索词索引数据，包括检索词出现的频率和在每一条记录中的位置，并写入数据库管理系统。

2.2 数据库双核存储技术

单份超大计算机文件（电子档案原文）无法装入非关系数据库，因此，本课题研发了数据库双核存储系统，以解决此问题。

2.2.1 数据库双核存储系统的结构

数据库双核存储系统是在数据库中设置数据库基本核和数据库扩展核，根据记录中各字段的数据类型和长度，配置数据库基本核的字段和数据库扩展核的字段。将记录的字段分为两部分，一部分字段在数据库基本核，另一部分字段在数据库扩展核，数据库基本核的字段组成基本核子记录，数据库扩展核的字段组成扩展核子记录，基本核子记录和相应的扩展核子记录形成完整的记录。由于数据库基本核和数据库扩展核是不可分割的两部分，基本核子记录和扩展核子记录也是不可分割的两部分。数据库双核存储系统的数据库文件的物理结构见图 2.7。

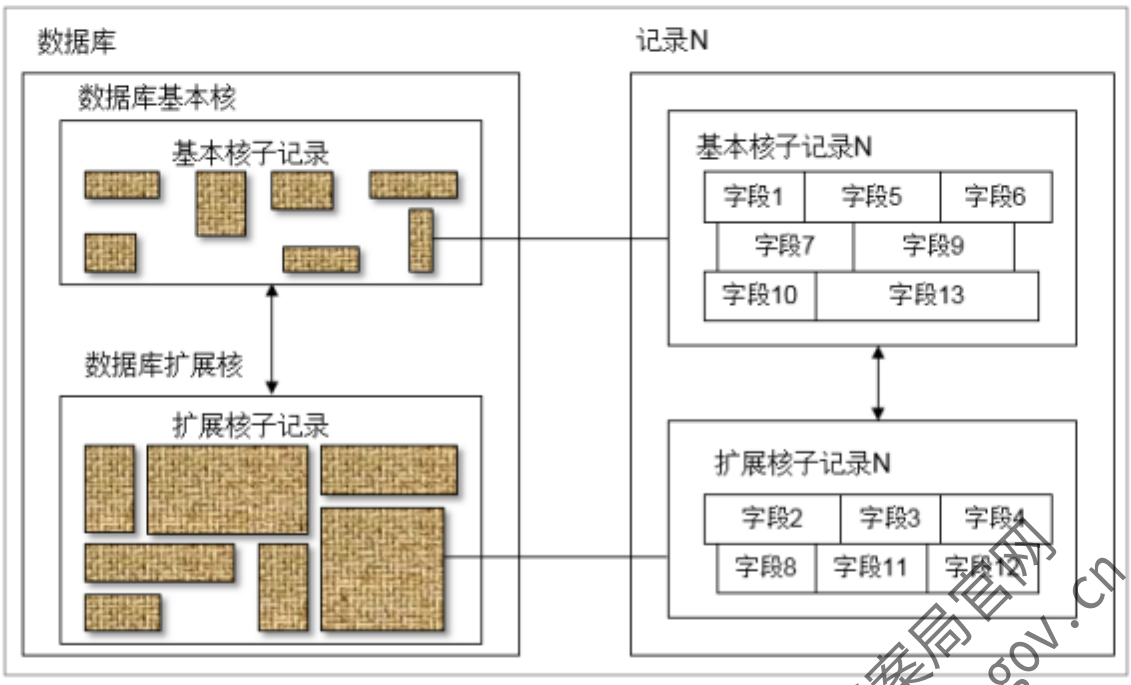

图 2.7 数据库双核存储系统的数据库文件的物理结构

2.2.2 数据库双核存储管理系统

数据库双核存储系统包括服务器，与服务器数据连接的磁盘存储装置和光盘存储装置，服务器上设置数据库管理系统，数据处理器，数据连接器，其中：

数据库管理系统用于响应数据请求，完成磁盘存储装置和光盘存储装置中数据库管理和数据管理；

数据处理器用于响应数据请求，配置数据库基本核的字段和数据库扩展核的字段，并将相应字段的数据分别写入数据库基本核和数据库扩展核；

数据连接器用于响应数据请求，建立数据库基本核和数据库扩展核的数据连接。

所述数据库管理系统包括字段结构生成模块，记录索引生成模块，存储空间分配模块，存储空间状态模块，记录链接模块，记录拆分模块，库结构生成模块，双核生成模块，其中：

字段结构生成模块，读取数据库中每一条记录的字段结构信息，包括记录中各字段的数据类型、长度，并写入数据库文件或数据库管理系统中；

记录索引生成模块，记录数据库中每一条记录的索引信息，包括记录中各字段的修改时间、修改内容，并写入数据库文件或数据库管理系统中；

存储空间分配模块，记录为每一条记录所分配的标准存储单元在数据库文件中的位置信息，并写入数据库文件或数据库管理系统中；

存储空间状态模块，记录数据库文件中已分配标准存储单元中未使用的空间信息，并写入数据库文件或数据库管理系统中；

记录链接模块，将数据库中各记录的字段结构信息、索引信息、标准存储单元的位置信息和空间信息合并，形成数据库特征数据，并写入数据库文件或数据库管理系统中；

记录拆分模块，根据数据库管理系统指令进行数据库中记录拆分，数据拆分以记录为单位进行，读取数据库特征数据，确定记录中每个字段的数据位置和数据量，标记出符合指令参数的记录，并将标记信息写入数据库文件或数据库管理系统中；

库结构生成模块，将数据库的库结构形成独立数据文件，数据库管理系统根据独立数据文件在磁盘存储装置上建立相同库结构的数据库，或在光盘上建立相同库结构的数据库；

双核生成模块，通过数据库管理系统在磁盘上的数据库中设置数据库基本核

或数据库基本核的数据库文件，设置数据库扩展核或数据库扩展核的数据库文件；

或在光盘上的数据库中设置数据库基本核或数据库基本核的数据库文件，设置数据库扩展核或数据库扩展核的数据库文件；

并将设置信息写入数据库文件或数据库管理系统中。

所述数据处理器包括字段配置模块，数据录入模块，其中：

字段配置模块，根据记录中各字段的数据类型和长度，配置数据库基本核的字段和数据库扩展核的字段，形成基本核子记录和扩展核子记录，并将字段配置信息写入数据库文件或数据库管理系统中；

数据录入模块，根据字段配置模块配置的数据库基本核的字段和数据库扩展核的字段，按照数据库管理系统指令，将相应字段的数据分别写入磁盘上的数据库基本核和数据库扩展核，或光盘上的数据库基本核和数据库扩展核。

所述数据连接器包括双核连接模块，按照数据库管理系统的指令，连接数据库的基本核子记录和相应的扩展核子记录，形成完整的记录。

2.2.3 数据存储的流程

在磁盘存储装置和光盘存储装置中完成数据在数据库基本核和数据库扩展核进行存储的步骤如图 2.8 所示，包括：

数据前向转移时：

数据库管理系统向数据处理器发出在数据库中配置数据库基本核的字段和数据库扩展核的字段的请求，数据处理器根据记录中各字段的数据类型和长度，配置数据库基本核的字段和数据库扩展核的字段；

数据库管理系统向数据处理器发出将相应字段的数据分别写入数据库基本核和数据库扩展核的请求，数据处理器将相应字段的数据分别写入数据库基本核和数据库扩展核。

数据后向转移时：

数据库管理系统向数据处理器发出在数据库中配置数据库基本核的字段和数据库扩展核的字段的请求，数据处理器返回数据库基本核配置的字段和数据库扩展核配置的字段的信息，数据库管理系统获得数据库基本核配置的字段和数据库扩展核配置的字段的信息；

数据库管理系统向数据处理器发出将相应字段的数据分别写入数据库基本核和数据库扩展核的请求，数据处理器返回写入数据库基本核和数据库扩展核的数据的信息，数据库管理系统获得写入数据库基本核和数据库扩展核的数据的信息；

数据连接器将写入数据库基本核和数据库扩展核的数据的信息形成完整的记录的信息。

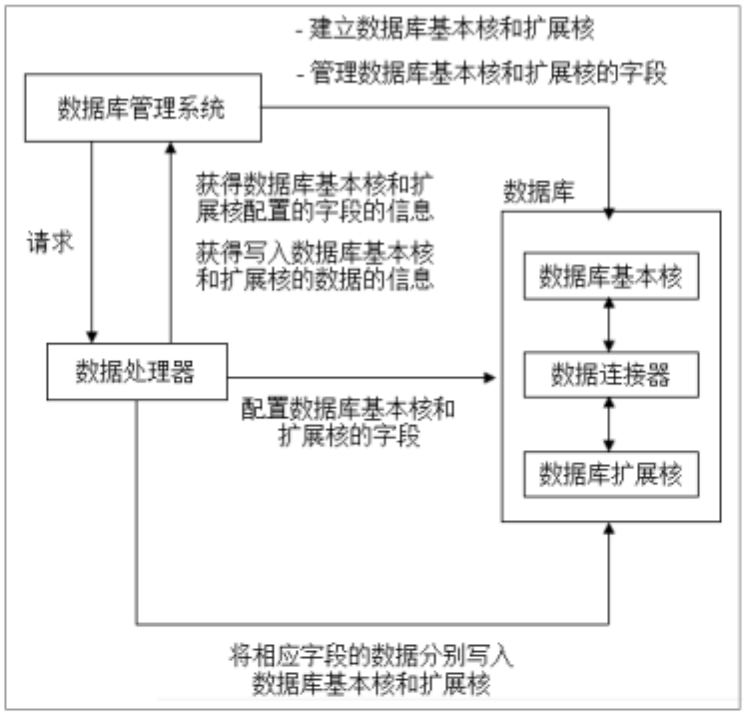

图 2.8 数据库双核存储系统完成数据库基本核和数据库扩展核进行数据存储的流程示意图

2.2.4 数据库双核存储系统在电子档案的应用

数据库双核存储系统是根据记录中各字段的数据类型和长度，配置数据库基本核的字段和数据库扩展核的字段。数据库基本核主要用于存储著录信息和小文件（电子档案原文），数据库扩展核主要用于存储超大文件（电子档案原文），包括电子文档、图像、音视频等，将记录的字段分为两部分，一部分字段在数据库基本核，另一部分字段在数据库扩展核，数据库基本核的字段组成基本核子记录，数据库扩展核的字段组成扩展核子记录，基本核子记录和相应的扩展核子记录形成完整的记录。可以将记录中的全部字段的各种格式的数据写入数据库，完善了数据库管理和存储超大文件（电子档案原文）的功能和结构。由于数据库基本核和数据库扩展核是不可分割的两部分，基本核子记录和扩展核子记录也是不可分割的两部分，保证了电子档案的完整性和安全性。

数据库双核存储方法能够在输出数据库的电子档案记录时，自动输出基本核子记录和扩展核子记录，利用数据库能够对各种类型的电子档案进行汇聚整合。

数据库双核存储方法能够在光盘上建立数据库，并在光盘上的数据库中设置

数据库基本核和数据库扩展核。根据记录中各字段的数据类型和长度，配置数据库基本核的字段和数据库扩展核的字段，记录的字段分为两部分，一部分字段在数据库基本核，另一部分字段在数据库扩展核，数据库基本核的字段组成基本核子记录，数据库扩展核的字段组成扩展核子记录，基本核子记录和相应的扩展核子记录形成完整的记录，采用这种方法，将电子档案记录中的全部字段的数据写入光盘上的数据库。由于数据库基本核和数据库扩展核是不可分割的两部分，基本核子记录和扩展核子记录也是不可分割的两部分，保证了光盘上电子档案的完整性和安全性，实现了利用在光盘上建立数据库的方法大规模存储、管理和访问大文件（电子档案原文），如视频、图片、图像等。

2.3 异构类型数据库存储技术

在一些情况下，非关系数据库处理结构化数据的效率低于关系数据库，在这种情况下，不宜采用非关系数据库管理结构化数据。

本课题研发的异构类型数据库存储方法，用于解决电子档案中的著录信息与电子档案原文通过单一类型数据库进行数据处理时导致数据处理性能下降的技术问题，解决电子档案中的不同结构类型数据无法结合数据库类型利用磁盘和光盘进行数据存储的技术问题。

2.3.1 异构类型数据库存储系统结构

异构类型数据库存储系统包括服务器，与服务器数据连接的磁盘存储装置和光盘存储装置，服务器上设置关系数据库管理系统，非关系数据库管理系统，数据处理器，数据库连接器，其中：关系数据库管理系统用于响应数据请求，完成磁盘存储装置中关系数据库管理和数据管理；非关系数据库管理系统用于响应数据请求，完成磁盘存储装置和光盘存储装置中非关系数据库管理和数据管理；数据处理器用于响应数据请求，配置关系数据库的字段和非关系数据库的字段，并将相应字段的数据分别写入关系数据库和非关系数据库；数据库连接器用于响应数据请求，建立关系数据库和非关系数据库的数据连接；存储于数据库中的记录通过数据处理器将字段分为两部分，一部分字段在关系数据库，另一部分字段在非关系数据库，关系数据库的字段组成关系数据库的子记录，非关系数据库的字段组成非关系数据库的子记录，关系数据库的子记录和相应的非关系数据库的子记录通过数据库连接器形成完整的记录。异构类型数据库存储系统的数据库文件

的物理结构示意图见图2.9。

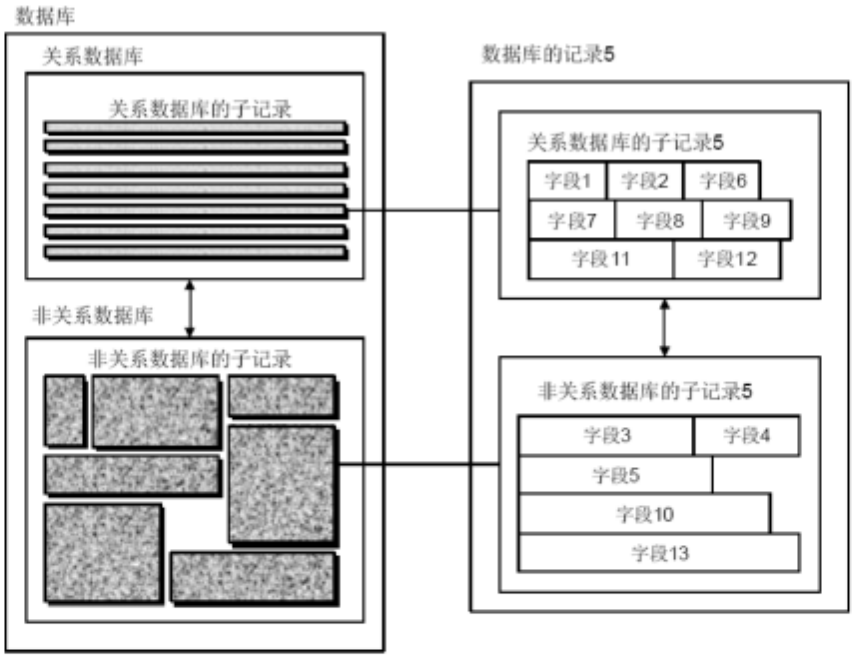

图2.9 异构类型数据库存储系统的数据库文件的物理结构示意图

2.3.2 异构类型数据库管理系统

异构类型数据库管理系统由关系数据库管理系统和非关系数据库管理系统组成。关系数据库管理系统包括关系库字段结构生成模块、关系库记录索引生成模块、关系库存储空间分配模块、关系库存储空间状态模块、关系库记录链接模块、关系库结构生成模块，非关系数据库管理系统包括非关系库字段结构生成模块、非关系库记录索引生成模块、非关系库存储空间分配模块、非关系库存储空间状态模块、非关系库记录链接模块、非关系库结构生成模块、非关系库记录拆分模块，数据处理器包括字段配置模块、关系库数据录入模块、非关系库数据录入模块，数据库连接器包括数据库连接模块，其中：

关系库字段结构生成模块，读取关系数据库中每一条子记录的字段结构信息，包括子记录中各字段的数据类型、长度，并写入关系数据库管理系统中；

关系库记录索引生成模块，记录关系数据库中每一条子记录的索引信息，包括子记录中各字段的修改时间、修改内容，并写入关系数据库管理系统中；

关系库存储空间分配模块，记录为每一条子记录所分配的标准存储单元在关系数据库中的位置信息，并写入关系数据库管理系统中；

关系库存储空间状态模块，记录关系数据库中已分配标准存储单元中未使用的

空间信息，并写入关系数据库管理系统中；

关系库记录链接模块，将关系数据库中各子记录的字段结构信息、索引信息、标准存储单元的位置信息和空间信息合并，形成关系数据库特征数据，并写入关系数据库管理系统中；

关系库结构生成模块，将关系数据库的库结构形成独立数据文件，关系数据库管理系统根据独立数据文件在磁盘存储装置中建立相同库结构的关系数据库；

非关系库字段结构生成模块，读取非关系数据库中每一条子记录的字段结构信息，包括子记录中各字段的数据类型、长度，并写入非关系数据库管理系统中；

非关系库记录索引生成模块，记录非关系数据库中每一条子记录的索引信息，包括子记录中各字段的修改时间、修改内容，并写入非关系数据库管理系统中；

非关系库存储空间分配模块，记录为每一条子记录所分配的标准存储单元在非关系数据库中的位置信息，并写入非关系数据库管理系统中；

非关系库存储空间状态模块，记录非关系数据库中已分配标准存储单元中未使用的空间信息，并写入非关系数据库管理系统中；

非关系库记录链接模块，将非关系数据库中各子记录的字段结构信息、索引信息、标准存储单元的位置信息和空间信息合并，形成非关系数据库特征数据，并写入非关系数据库管理系统中；

非关系库结构生成模块，将非关系数据库的库结构形成独立数据文件，非关系数据库管理系统根据独立数据文件在光盘上建立相同库结构的非关系数据库，或在其他磁盘存储装置中建立相同库结构的磁盘非关系数据库；

非关系库记录拆分模块，根据非关系数据库管理系统的指令进行非关系数据库的子记录拆分，数据拆分以子记录为单位进行，读取数据库特征数据，确定子记录中每个字段的数据位置和数据量，标记出符合指令参数的子记录，并将标记信息写入非关系数据库管理系统中。

字段配置模块，根据记录中各字段的数据类型和长度，配置关系数据库的字段和非关系数据库的字段，并将字段配置信息分别写入关系数据库管理系统和非关系数据库管理系统中；

关系库数据录入模块，根据字段配置模块配置的关系数据库的字段，按照关系数据库管理系统的指令，将相应字段的数据写入关系数据库；

非关系库数据录入模块，根据字段配置模块配置的非关系数据库的字段，按照非关系数据库管理系统的指令，将相应字段的数据写入光盘上的非关系数据库或磁盘上的非关系数据库。

数据库连接模块，按照关系数据库管理系统和非关系数据库管理系统的指令，连接关系数据库的子记录和相应的非关系数据库的子记录，形成完整的记录。

2.3.3 数据存储的流程

在存储装置中完成数据在关系数据库和非关系数据库进行存储的步骤如图2.10所示，包括：

数据前向转移时：

关系数据库管理系统向数据处理器发出配置关系数据库的字段请求，数据处理器根据记录中各字段的数据类型和长度，配置关系数据库的字段；

非关系数据库管理系统向数据处理器发出配置非关系数据库的字段请求，数据处理器根据记录中各字段的数据类型和长度，配置非关系数据库的字段；

关系数据库管理系统向数据处理器发出将相应字段的数据写入关系数据库的请求，数据处理器将相应字段的数据写入关系数据库；

非关系数据库管理系统向数据处理器发出将相应字段的数据写入非关系数据库的请求，数据处理器将相应字段的数据写入非关系数据库。

数据后向转移时：

关系数据库管理系统向数据处理器发出配置关系数据库的字段请求，数据处理器返回关系数据库配置的字段信息，关系数据库管理系统获得关系数据库配置的字段信息；

非关系数据库管理系统向数据处理器发出配置非关系数据库的字段请求，数据处理器返回非关系数据库配置的字段信息，非关系数据库管理系统获得非关系数据库配置的字段信息；

关系数据库管理系统向数据处理器发出将相应字段的数据写入关系数据库的请求，数据处理器返回写入关系数据库的数据的信息，关系数据库管理系统获得写入关系数据库的数据的信息。

非关系数据库管理系统向数据处理器发出将相应字段的数据写入非关系数据库的请求，数据处理器返回写入非关系数据库的数据的信息，非关系数据库管理

系统获得写入非关系数据库的数据的信息。

数据库连接器将写入关系数据库的数据的信息和写入非关系数据库的数据的信息形成完整的记录的数据。

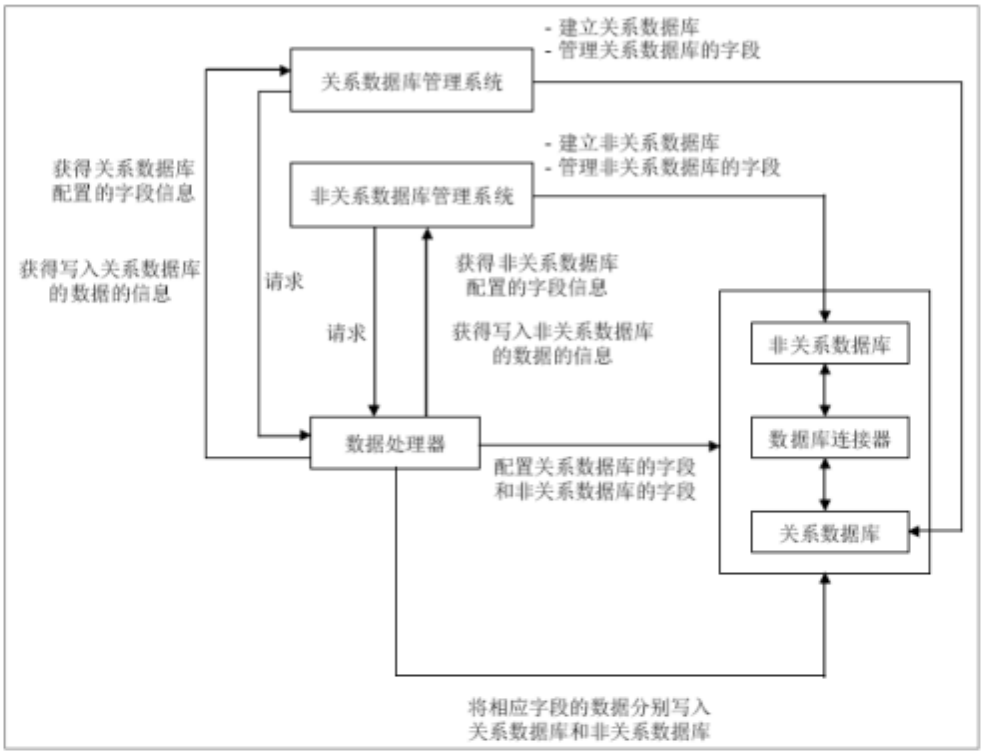

图2.10 异构类型数据库存储系统在存储装置中进行数据存储的流程示意图

2.3.4 异构类型数据库存储系统在电子档案的应用

本课题研发的异构类型数据库存储方法可以将记录的字段分为两部分，著录信息的字段在关系数据库，电子档案原文的字段在非关系数据库，不再将电子档案原文挂接在关系数据库上，使得所有的数据都存储在数据库中。关系数据库的字段组成关系数据库的子记录，非关系数据库的字段组成非关系数据库的子记录，通过连接关系数据库与非关系数据库，关系数据库的子记录和相应的非关系数据库的子记录形成完整的记录，保证了电子档案的完整性。充分利用关系数据库和非关系数据库的优势和特点，通过检索磁盘上关系数据库中在著录信息，查找到非关系数据库中同一记录的电子档案原文，提高了关系数据库的存储能力和安全性。对于已有的关系数据库，可采用本课题研发的处理数据的方法，将挂接在关系数据库上的电子档案原文转储到非关系数据库。

利用本课题研发的方法，关系数据库存储电子档案记录中的著录信息，非关系数据库存储电子档案记录中的电子档案原文，关系数据库建立在磁盘上，非关系

数据库管理系统将一个庞大的磁盘上的非关系数据库拆分成若干个结构定义一致、数据完整的子数据库，每一个包含子数据库的光盘都可以接受非关系数据库管理系统的管理，使得磁盘上的关系数据库的子记录可以与相应光盘上的非关系数据库的子记录形成完整的记录。利用光盘存储装置容量巨大，数据保存安全性高的特点，大规模存储电子档案原文，为实现利用磁盘存储装置和光盘存储装置进行分级存储电子档案提供了良好的途径。同时，可以显著降低电子档案数据库存储系统的构建成本，降低能源消耗。

2.4 数据库拆分与数据库合并技术

随着档案行业信息系统的扩建和新建，大量计算机和现代化电子设备的投入使用，信息采集更加准确和及时，获取的信息量不断增加，数据的类型和格式不断增多。历史纸质档案数字化与现时电子档案融合在一起，电子档案的应用广度和深度大大提高，对电子档案的功能和综合利用提出更高的需求。例如：一个事件涉及多个类别或多个部门的档案；有时需要对涉及一个事件的不同类别的档案专门归档，形成一个专题档案；有时一个档案可能归属多个类别的档案，等等，这就要求电子档案系统具有足够的灵活性和广泛的适用性，对电子档案高效存储、科学分类、有效汇聚整合，需要解决电子档案数据库的拆分和合并的技术问题。

2.4.1 数据库拆分

（1）数据库拆分是以记录为单元进行拆分。一条记录可以包括结构化数据和非结构化数据，数据库拆分时，将一条记录的结构化数据和非结构化数据同时拆分出来。

（2）一个数据库被拆分成若干个结构定义一致、数据完整的子数据库，同时生成相应的子数据库主文件（见图 2.11）。

（3）对每个子数据库主文件单独进行索引，生成与子数据库主文件相对应的子数据库索引文件（见图 2.11）。

（4）数据库中的各记录关联数据不需要跨子数据库存储或关联，使得数据库在存储过程中可以保持数据的完整性，每一个子数据库作为正常的数据源接受数据库管理系统的管理。

（5）可按记录数、容量大小、档案的内容、档案的内在联系拆分数据库。

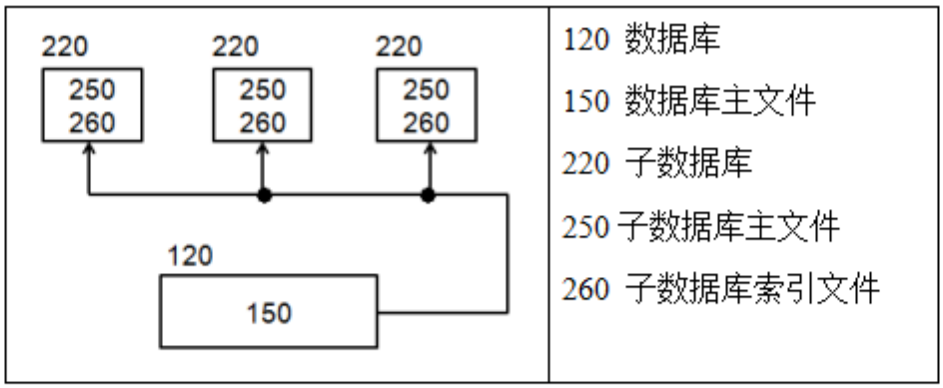

图 2.11 数据库拆分示意图

2.4.2 数据库合并

（1）数据库合并是以记录为单元进行合并。一条记录可以包括结构化数据和非结构化数据，数据库合并时，将一条记录的结构化数据和非结构化数据同时合并到数据库。

（2）若干个结构定义一致的数据库合并生成一个大数据库，同时生成相应的大数据库文件（见图 2.12）。

（3）对合并生成的大数据库主文件进行索引，生成与大数据库主文件相对应的大数据库索引文件（见图 2.12）。

（4）生成的大数据库作为正常的数据源接受数据库管理系统的管理。

（5）可按记录数、容量大小、档案的内容、档案的内在联系合并数据库。

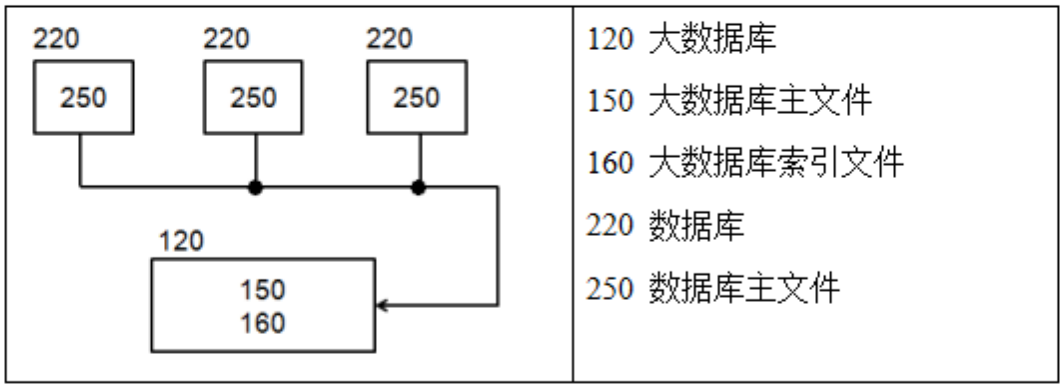

图 2.12 数据库合并示意图

2.5 基于光盘的数据库存储技术

2.5.1 光盘存储数据的现有技术

在现有技术中，光盘存储数据主要有以下三种方法：

（1）光盘存储数据

这种方法是将各种格式的数据刻录在光盘上，包括电子文件和电子档案。一般采用目录数据库（关系数据库）管理光盘上的数据。将光盘上数据的著录信息存储在目录数据库中，通过检索目录数据库中的著录信息查找光盘上的数据。

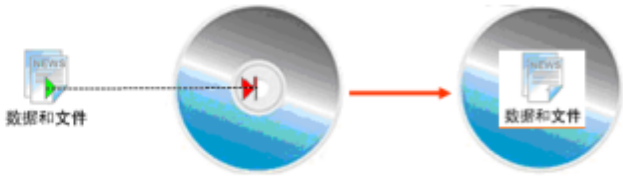

图 2.13 光盘存储数据示意图

（2）光盘存储数据块文件

这种方法是将数据装入数据块，形成数据块文件，然后将数据块文件刻录在光盘上。一般采用目录数据库（关系数据库）管理光盘上的数据块和数据。将光盘上数据块和数据的著录信息存储在目录数据库中，通过检索目录数据库中的著录信息查找光盘上的数据块和数据。

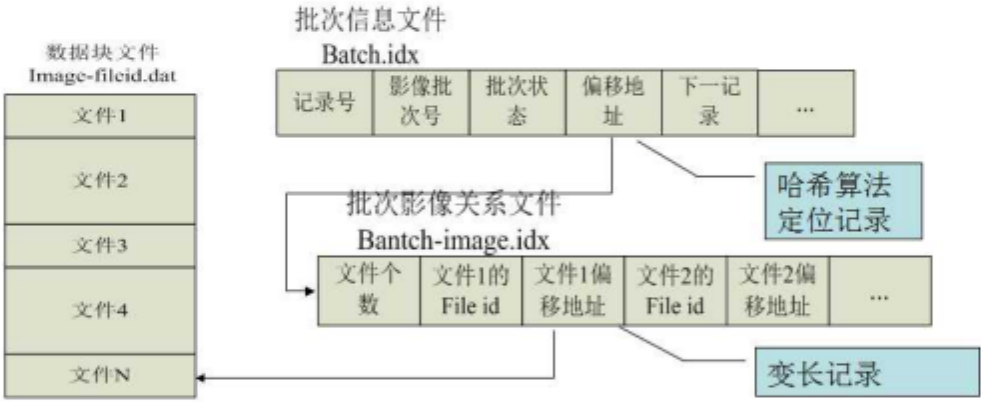

图 2.14 数据块文件结构示意图

（3）光盘存储数据库数据

这种方法用于将磁盘上的数据库数据备份到光盘上。光盘上的备份内容包括物理备份和逻辑备份。物理备份包括数据库数据和控制文件，逻辑备份是标识符。一般采用关系数据库管理光盘上的物理备份和逻辑备份。

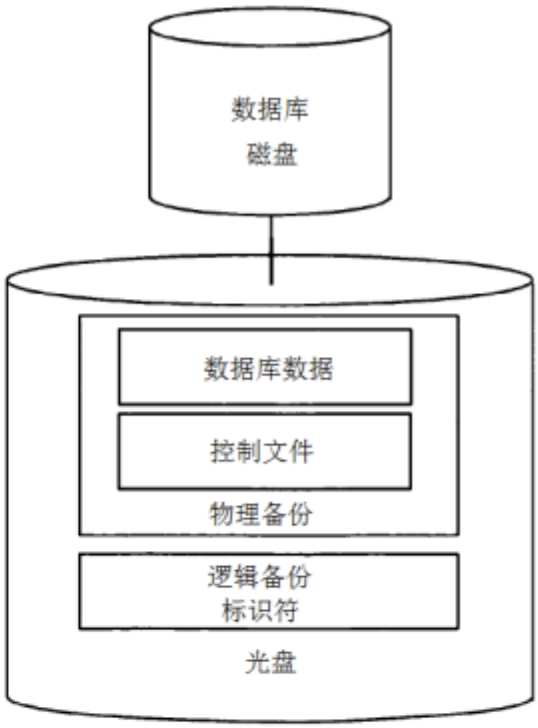

图 2.15 光盘备份数据库结构示意图

以上三种方法实质上是将关系数据库建立在磁盘上，将数据存储在光盘上，采用关系数据库管理光盘上的数据，仅可对关系数据库中的著录信息进行检索，不能对光盘上的数据内容进行全文检索。

2.5.2 数据库建立在光盘上

本课题研发的关键技术之一是要突破光盘上存储数据的传统方法，改变以往“将数据存储在光盘上”的传统模式，代之以“将数据存储到数据库，把数据库存储在光盘上”的新模式，按照在磁盘上建立数据库的方法在光盘上建立数据库，实现对光盘上的数据内容进行全文检索。在光盘上建立数据库有两种可能途径和方法，一种方法是在光盘上建立数据库，并将数据装入光盘上的数据库，另一种方法是将磁盘上的数据库转移到光盘上。

第一种方法：在光盘上建立数据库。这种方法与在磁盘上建立数据库的方法相同。首先要解决在光盘上建立数据库的技术问题，其次，要解决将数据装入在光盘上的数据库中的技术问题。

目前，数据库存储技术中的数据库类型划分为关系型和非关系型。关系数据库所管理的数据是能够用平面（二维）行列表结构进行逻辑表达，一行代表一个记录，每列的数据相当于不同记录中相同字段的数据。一个库可以设计成由多张

二维表组成。不同表之间的联系通过关系来实现。因此，不能将关系数据库建立在光盘上。非关系数据库是把数据组织在不限空间的文件中，突破了关系数据库严格的表结构，解决了关系数据库模型简单，不易表达复杂嵌套数据结构的问题。非关系数据库能够容纳各种格式和类型的数据，将结构化数据、半结构化数据和非结构化数据全部装入数据库。使非关系数据库独立于操作系统，使数据库管理软件与数据库分离。因此，将非关系数据库建立在光盘上成为可能。

非关系数据库的结构如同气球，数据库文件大小随装入的数据量大小而变化，数据装入量增加，数据库文件增大。本课题根据光盘存储数据的性质，解决了光盘上数据库文件增大的技术问题，这样，就能够将数据装入光盘上的数据库。

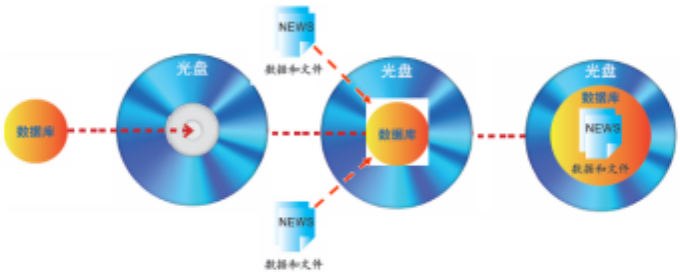

图2.16 在光盘上建立数据库示意图

第二种方法：将磁盘上的数据库转移到光盘上。需要解决大容量数据库转移到光盘上的技术问题。

当光盘容量小于磁盘数据库的容量时，需要利用数据库拆分技术将磁盘数据库拆分为若干个子数据库，使每个子数据库的容量小于光盘的容量，拆分后生成的若干个子数据库可以转移在多张光盘上，每张光盘上的数据库结构与磁盘上的数据库结构完全相同。

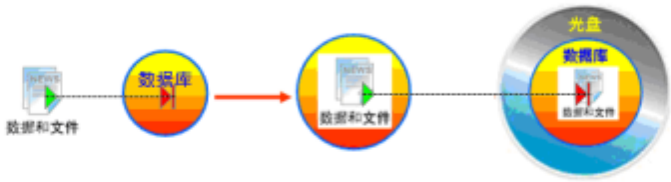

图 2.17 磁盘上数据库转移到光盘上示意图

这两种在光盘上存储数据库的方法是将物理存储与逻辑存储有机地结合在一起。光盘上存储数据库是物理存储，数据库存储数据是逻辑存储。

2.5.3 光盘上数据库全文检索

全文索引是全文检索的基础。全文索引是对数据库中的非结构化数据和从非结

构化数据中抽取出的全部字符进行索引。抽取字符（抽词）是非常重要的技术环节。

全文索引的方法是：将数据（结构化数据和非结构化数据）装入数据库主文件；然后对导入数据库主文件的全部字符进行全文索引，生成索引文件。数据库主文件和索引文件是两个独立的文件，且一一对应，既可以放置在磁盘的相同位置（例如 D 盘），也可以分别放置在磁盘的不同位置（例如 D 盘和 E 盘）。

本课题研发的关键技术之一是对光盘上的内容进行全文检索。利用数据库文件和索引文件可以分别放置在不同位置上的特性，设计出将数据库主文件放置在光盘上的方法，既可以将索引文件放置在光盘上，也可以放置在磁盘上，且一一对应，通过对索引文件的全文检索，实现对光盘上内容的全文检索。

本课题研发的检索词索引生成模块，根据数据库中每一条记录的索引信息，建立与标记信息对应的记录的检索词索引数据，包括检索词出现的频率和在每一条记录中的位置，特别是在光盘上的位置。采用的方法是数据库管理系统通过检索词索引生成模块形成与光盘上的数据库主文件相应的检索词索引数据，然后通过数据库管理系统将其存储在相应数据库的光盘上，或存储在磁盘上。

对于光盘上的子数据库，是对每个子数据库主文件单独进行索引，生成与子数据库主文件相对应的子数据库索引文件，且一一对应。数据库管理系统可以直接在子数据库中完成该记录范围的数据检索和查询，不需要对同一记录的不同字段数据在各子数据库间进行数据检索，保持了各子数据库的数据完整性。每一个子数据库作为正常的数据源接受数据库管理系统的管理。

2.6 基于光盘库的数据库存储与全文检索技术

由于磁盘数据库容量大，而光盘的最大容量是 300GB，因此，要将庞大的磁盘数据库存储在光盘上，就必须将磁盘数据库拆分为若干个子数据库，并分散存储在若干张光盘上，这需要解决两个技术问题，一是解决对大规模存储在光盘上的数据库和子数据库进行统一管理和访问的技术问题，二是解决管理和访问大规模数据库光盘的技术问题。本课题利用现有技术中的光盘库作为大规模在光盘上存储数据库的硬件基础。光盘库存储数据库，是给光盘库中每张光盘都创造了与磁存储相似的环境，将光盘库中光盘存储数据的方式由存储数据转变为存储数据库。

2.6.1 光盘库

光盘库是一种带有自动换盘器（机械手）的光盘网络共享设备，是一个海量数据存储设备系统。光盘库可容纳35—12000张光盘，容量达到3.5TB—1.2PB，并配置多台光驱。用户访问光盘库时，自动换盘器从光盘槽取出所需的光盘并送入光驱中进行读写。自动换盘器的换盘时间通常在秒级。光盘库管理系统对光盘库硬件进行操作，包括对机械手和光盘驱动器等的管理，对光盘的文件管理、光驱读写、数据传输等。通过光盘库管理系统可以看到光盘库中每个光盘槽的状况，例如：光盘的位置，光盘槽上是否有盘，光盘的种类，是CD光盘，还是DVD光盘，或是BD蓝光光盘，是一次性写入光盘还是可重写光盘，是空盘还是有内容的盘，光盘卷标（盘名），光盘被激活还是未被激活等状态参数。光盘库管理系统视光盘库为一个整体，与光盘库中装有光盘的数量无关。光盘库管理系统与操作系统的文件结构、目录结构等系统级数据结构无缝连接，使得光盘库在服务器上映射为一个盘符，相当于硬盘上的一个分区，例如E盘或Z盘。每个光盘的卷标（盘名）相当于硬盘上的一个文件夹。可以在系统的资源管理器上看到光盘库所存光盘的卷标，卷标下的文件夹名和文件名，用系统的使用方式对光盘进行检索读取。用户访问文件时并不需要知道光盘在光盘库中的具体位置。

2.6.2 基于光盘库的数据库存储系统

基于光盘库的数据库存储系统，包括服务器、光盘库和磁盘存储器，服务器与光盘库间建立数据连接，服务器与磁盘存储器间建立数据连接，服务器上的操作系统中安装有数据库管理系统和光盘库管理系统，数据库管理系统用于响应数据请求，完成存储装置的数据库管理和数据管理，光盘库管理系统用于完成光盘库与操作系统数据结构的连接。

2.6.3 在光盘库中光盘上建立数据库

在光盘库中光盘上建立数据库有两种方法。一种方法是直接在光盘库中光盘上建立数据库，并将数据装入光盘上的数据库。另一种方法是采用两步法在光盘库中光盘上建立数据库。

第一种方法：利用基于光盘库的数据库存储系统直接在光盘上建立数据库的步骤如下：

步骤1，数据库管理系统通过光盘库管理系统获取光盘库中光盘存储介质的容

量参数；

步骤 2，数据库管理系统利用库结构生成模块 u8 在光盘存储介质上建立数据库文件；

步骤 3，数据库管理系统利用光盘建库与录入模块向光盘存储介质上数据库文件中增加记录，数据库管理系统通过字段结构生成模块将写入相应记录的字段结构信息保留，通过记录索引生成模块将相应记录的索引信息保留，通过存储空间分配模块将相应记录的标准存储单元位置信息保留，通过存储空间状态模将记录标准存储单元的空间信息保留，通过记录链接模块形成相应的数据库特征数据保留；

步骤 4，重复步骤 3，更新保留的数据库特征数据；

步骤 5，当数据库达到光盘存储空间容量值（指光盘容量或小于光盘容量的设定值）时，数据库管理系统将保留的数据库特征数据写入光盘上的数据库，在光盘上完成数据库建立与记录存储；

步骤 6，数据库管理系统通过检索词索引生成模块形成与光盘上的数据库相应的检索词索引数据；

步骤 7，重复步骤 1 至 6，直至完成数据存储。

该方法可以通过光盘库直接在光盘上完成数据库建立和增加数据，使得基于光盘库的数据库存储系统可以用于数据安全性要求高，数据响应周期长的在线数据存储，部分替代在线磁盘存储设备。

第二种方法：利用基于光盘库的数据库存储系统采用两步法在光盘上建立数据库的步骤如下：

步骤 1，数据库管理系统通过光盘库管理系统获取光盘库中光盘存储介质的容量参数；

步骤 2，数据库管理系统根据光盘存储介质的容量通过记录拆分模块完成磁盘上的数据库的记录拆分（数据库拆分），形成子数据库的标记信息；

步骤 3，数据库管理系统通过光盘建库与录入模块在光盘库中的相应光盘上建立子数据库文件；

步骤 4，数据库管理系统通过字段结构生成模块向光盘库中各个光盘上的子数据库文件中写入相应记录的字段结构信息；

步骤 5，数据库管理系统通过记录索引生成模块向光盘库中各个光盘上的子数据库文件中写入相应记录的索引信息；

步骤 6，数据库管理系统通过存储空间分配模块向光盘库中各个光盘上的子数据库文件中写入相应记录的标准存储单元位置信息；

步骤 7，数据库管理系统通过存储空间状态模块记录标准存储单元的空间信息；

步骤 8，数据库管理系统通过记录链接模块在光盘库中各个光盘上的子数据库形成相应的子数据库特征数据，完成子数据库建立与存储。

步骤 9，数据库管理系统通过检索词索引生成模块形成与各光盘上的子数据库相应的检索词索引数据。

该方法可以通过光盘库完成现有海量数据形成的光盘上的子数据库数据的统一管理，光盘上的子数据库成为操作系统中文件结构的有机组成部分，使得数据库的拆分、数据的变化在光盘上可以实现。

2.6.4 光盘库中数据库全文检索

光盘库中光盘上的数据库独立于计算机操作系统。服务器上的数据库管理软件与光盘库中光盘上的数据库相互分离，光盘上只有数据库，没有数据库管理软件。数据库管理软件通过光盘库管理软件，对光盘库中所有光盘上的数据库进行管理。

在现有技术中，无法同时对多张光盘上的多个数据库（多盘多库）进行全文检索，因此，需要将各光盘上的数据库主文件相对应的索引文件存储在磁盘上，对磁盘上的多个索引文件进行全文检索。如果数据库主文件和索引文件都在磁盘上，容易实现多库（对多个结构相同或结构不同的数据库）检索，因为对索引文件检索时，数据库管理系统会将检索获得的信息传递给数据库主文件，磁盘上数据库主文件会立即响应。如果数据库主文件在光盘上，当对磁盘上的索引文件进行检索时，数据库管理系统要将实施检索的信息传递给光盘上的数据库主文件，使得光盘库的机械手抓取相应于索引文件的数据库主文件所在的光盘，放入光驱进行读取，读取获得的信息返回数据库管理软件，才能获得检索击中的记录。读取一张光盘花费的时间约 10 秒，也就是说，检索一个索引文件需要花费约 20 秒，同时检索 10 个索引文件则需要花费 3 分钟以上，这种方法在实际应用中不可取。为

此，本课题研发出光盘库全文检索模块。

光盘库全文检索模块将检索分为两步，第一步对磁盘上的索引文件进行检索，数据库管理系统保留实施检索的信息，不传递给光盘上的数据库主文件，不需将光盘库中数据库主文件所在的光盘放入光驱进行读取就可以获得检索击中的记录数，仅显示检索击中的记录数。检索一个索引文件所需的时间仅为毫秒级，同时检索 10 个索引文件不超过 1 秒。第二步显示检索的结果，数据库管理系统将要显示检索结果的信息传递给光盘上的数据库主文件，使得光盘库的机械手抓取相应于索引文件的数据库主文件所在的光盘，放入光驱进行读取，获得检索的结果，读取一张光盘花费的时间约 20 秒。根据检索结果，直接读取光盘上数据库主文件中记录内容，不需要将记录还原到磁盘上。显示当前页结果时，仅调取当前页的结果所在的光盘，不调取其他页的结果所在的光盘。当前页的结果分布的光盘数量超过光驱数时，不出现无限循环读取光盘上数据库的记录的现象。

3 基于非关系数据库的电子档案存储规范

《基于非关系数据库的电子档案存储规范》具体内容见附件 1。

4 基于非关系数据库的电子档案存储系统

4.1 用户管理系统

用户管理系统设计严密，建立与管理用户简便灵活，便于掌握。其主要特点是：

（1）用户分为四类：系统管理员、数据库管理员、用户管理员和一般用户。

（2）系统管理员创建和管理数据库管理员，但无权管理数据库管理员管理的数据库；

（3）系统管理员创建和管理用户管理员，但无权管理用户管理员管理的一般用户；

（4）根据需求创建数据库管理员和用户管理员，数据库管理员和用户管理员的数量均无限制；

（5）系统管理员创建和管理一般用户，一般用户的数量无限制。

（6）用户管理员创建和管理一般用户，一般用户的数量无限制。