

4.2 重复数据删除存储原理

4.2.1 系统架构

重复数据删除存储系统包括文件访问协议、文件服务、内容分析、数据块过滤以及块存储等层次。如图4-2所示，顶端为文件访问协议层，可以支持若干种文件访问协议，如NFS、CIFS、VTL或iSCSI等。通过文件访问协议层将存储子系统网络化，实现高速共享存储的目标。在访问协议层以下是包括各种标准的文件服务接口的通用文件服务层，它负责管理文件系统的元数据和名空间。传统的存储系统在文件服务层之下即为块操作接口，而重复数据删除系统在文件和数据块之间，定义了一种抽象的数据对象——块（Chunk）。块是对文件进行内容分析的基本单位，它可以是一个文件，也可以是一个数据块。为了进行重复数据删除，内容分析层按照数据划分策略将文件划分成若干个块，同时计算出每个块的特征值。这样，文件可以通过其所含数据块的特征值列表来表示，而不是逻辑块地址信息。数据块过滤层比对块的特征值来确定块是否为重复的。块存储则负责存储唯一的块，并可以结合数据压缩来存储块。为支持快速查询，重复数据删除系统建立了块索引来保存块的特征值以及块到逻辑块的对应关系等信息。

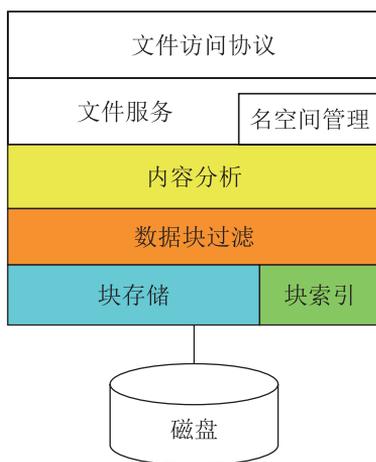


图4-2 重复数据删除存储系统结构