

# 分布式存储系统中用户磁盘空间分配策略

谭子军, 何连跃

(国防科学技术大学计算机科学与技术学院, 长沙 410073)

**摘要:** 针对用户文件的分布式存放导致的磁盘空间管理问题, 提出一种在分布式存储系统中动态分配磁盘空间的策略, 在保持用户磁盘配额大小不变的情况下, 根据各存储节点的数据量差异, 按需分配用户实际所用的存储空间。与传统的磁盘分配机制相比, 该方法简便灵活, 更能适应网络存储数据的动态变化, 有效提高磁盘空间资源的利用率。

**关键词:** 分布式存储系统; 磁盘配额; 物理空间; 按需分配

## User Disk Space Allocation Policy in Distributed Storage System

TAN Zi-jun, HE Lian-yue

(School of Computer Science and Technology, National Defense University of Science and Technology, Changsha 410073)

**【Abstract】** Aiming at the problem that allocating disk space result of the distribution of user files stored is concerned by the people, this paper presents a policy which is named dynamic allocation of disk space in distributed storage system. It maintains the size of the user disk quota unchanged, and allocates the actual storage space for user on demand according to the amount of data difference in each storage node. Compared with traditional disk allocation mechanism, this method is more simple and flexible, and it can rather adapt to the dynamic changes of the network storage data, and improve effectively the utilization of disk space resources.

**【Key words】** distributed storage system; disk quota; physical space; allocation on demand

### 1 概述

在传统的网络存储系统中, 所有的数据和元数据都集中在存放在存储服务器里, 随着客户访问连接数的增加, 单服务器成为整个系统性能的瓶颈, 无法满足大规模存储应用的需要。因此, 分布式存储技术正成为存储领域新的发展趋势和研究方向。分布式存储系统采用可扩展的结构, 将分布在网络中的存储资源组织起来, 用多台存储服务器共同分担存储负荷, 使之构成大容量的磁盘空间, 并采用元数据服务器为用户定位检索存储信息, 同时管理元数据和系统资源, 而后台的服务器集群向用户提供实际的数据存储服务。这种体系结构不但提高了系统的可靠性、可用性和工作效率, 还易于扩展和移植。

### 2 麒麟天机安全存储系统

麒麟天机安全存储系统(Kylin Tianji Security Storage System, KTSSS)是国防科技大学计算机学院软件所银河麒麟课题组自主研发的网络数据安全存储服务系统, 目的是为用户提供透明的数据加密存储服务, 同时实现用户之间的数据秘密共享<sup>[1]</sup>。该系统采用集中式存储方式, 目前课题组正在将其改进到分布式多服务器结构。分布式天机系统的物理结构如图1所示。

KTSSS 由以下几类节点部署而成: 普通用户客户端, 管理用户客户端, 安全存储根服务器, 安全存储文件服务器, 证书管理系统。其中, 根服务器和文件服务器是整个系统的核心部分, 根服务器只有一台, 集中存放着与用户和服务器相关的元数据, 起到信息索引和状态监控的作用; 文件服务器是由多台安全存储服务器组成的机群, 用户通过连接这些服务器来访问自己存储的文件数据。

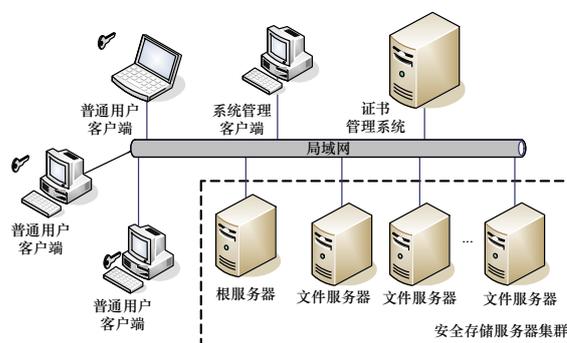


图1 分布式天机安全存储系统的物理结构

在 KTSSS 中, 用户是以加密保险箱的形式来存储数据的, 并可以对保险箱进行创建、删除、查询或设置属性、管理共享用户等。保险箱只有它的创建者才能打开, 对于其他用户来说是隐藏不可见的, 这就保证了文件数据的安全性。但保险箱也能共享给其他用户, 创建者可以对这些共享用户设置不同的访问权限。拥有相应的权限后, 其他用户就能看见此保险箱并对其访问。文件保险箱创建在哪台文件服务器上由系统根据一定的策略决定, 这台文件服务器称为该用户的相关文件服务器。文件保险箱创建后, 用户可以在保险箱里继续创建文件或文件夹, 实行查看、编辑或删除数据

**基金项目:** 国家“863”计划基金资助项目“分布加密存储软件结构及其关键技术”(2007AA012408)

**作者简介:** 谭子军(1983-), 男, 硕士研究生, 主研方向: 分布式存储系统; 何连跃, 副研究员

**收稿日期:** 2009-10-20 **E-mail:** tzj1983@163.com

的操作。一个用户可拥有许多保险箱，这些保险箱可能在不同的文件服务器上，而系统需要给用户设定磁盘配额时就带来了分配和管理方面的问题。

### 3 用户磁盘配额问题

单服务器天机系统中，所有保险箱都放在一个服务器里，固定的磁盘配额由文件系统的磁盘配额机制实现。保险箱数据分布结构如图 2 所示。

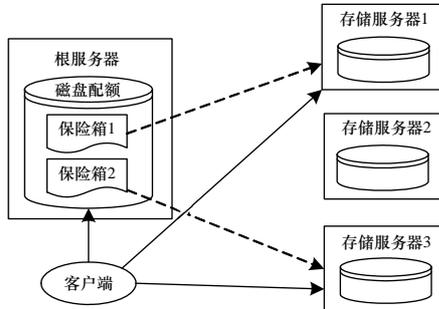


图 2 保险箱数据分布结构

然而，当 KTSS 扩展到多服务器结构后，由于一个用户的保险箱可以分散存放到不同服务器上，用户磁盘配额如何在不同的文件服务器上实行分配，成为有待解决的问题。事实上，保险箱内的数据因为应用不同，所占存储空间大小也各不相同，对某一个用户来说，必然有些文件服务器上保险箱数据量大，所需磁盘空间也较多；而有些文件服务器上数据量小甚至为空，所需的空间便少。况且，用户对数据的修改是随机而频繁的，使得保险箱的数据量会随时变化，因此，分配的存储空间还要根据数据被访问情况而动态调整。此外，一台文件服务器上可能存放着用户的多个保险箱，磁盘空间的分配当以服务器为基本单位，细分到保险箱是没有意义的。可见，使用一种合理的方案给与用户相关的文件服务器分配磁盘限额，使之既能满足各个服务器的差异性，又能适应数据变化的动态性，这对提高存储空间的利用率有重要意义。

磁盘分配一般可以用到等分法、动态统计法<sup>[2]</sup>等。等分法是将用户磁盘配额平均分割后再配给各个相关服务器。这种方法显然不可行，对于数据量较小的保险箱所在的文件服务器，就存在着空间的浪费；而那些占空间较大的保险箱所在的文件服务器，用户再往里加入数据就有可能超出限额，让访问无法完成。相比之下，动态统计法能周期性地监测每个文件服务器中保险箱的数据改变量，然后以最近一段时期内保险箱的数据量变化率为依据对其分配磁盘空间，变化率为正且较大的，说明保险箱数据增加地很快，必须预留更多的空间；变化率很大但为负的，说明保险箱数据减少地很快，就得剥夺更多的空间；而那些变化率较小的，说明保险箱数据量变化不大，为其预留或剥夺的空间也较少。这种策略动态适应性较强，比等分法更优越，但它毕竟是一种通过提前估计保险箱数据变化趋势而作出的决策，由于用户对数据的访问具有很大的不确定性，使得这种预测不一定准确。比如由于数据量下降地较快，文件服务器被剥夺了较大的空间，然后某段时间用户突然往里写入了大量的数据，所占空间一下子超出限额，造成用户访问无法成功，要么得等待系统作出分析，临时为该文件服务器增加配额。

由此可见，上述 2 种方法均不能时刻满足用户对存储空间的需求变化。为解决这个问题，笔者提出了一种存储空间按需分配(Storage Space Allocation on Demand, SSAoD)<sup>[3]</sup>的原

则来为一个用户在相关的不同文件服务器上动态分配磁盘空间的方法。

### 4 存储空间按需分配原理

SSAoD 的基本思想为：对于一台文件服务器，当某个用户在上面创建第一个保险箱时就为其分配与当前用户磁盘配额余额一样大小的存储空间，使之能最大限度地满足用户的需求，但没有占用物理磁盘。只有当用户写入数据时，才会占用实际的物理空间。同时，该文件服务器会根据用户访问保险箱的情况，按与根服务器同步更新的方式自动地调整所需磁盘空间。

SSAoD 机制能较好地解决存储系统中“空间刚好用尽”的问题。它采用动态分配的方法，用户定义保险箱时，分配的存储空间都是虚拟的，不会立即占用物理空间，当发生数据访问时就按这样一种原则实行分配——用时才分配、需要多少分配多少<sup>[4]</sup>、超过限额不予分配。每次用户写数据，系统便按当前所需大小给文件服务器分配存储容量，这样仅占用了已被写入实际数据的物理空间。此外，每次用户删除数据或保险箱时，系统会释放所占用的空间，以供其他保险箱或相关文件服务器使用，有效提高存储空间资源利用率。

用户所有未使用的空间会集中在一起，放在全局空间中，共享给用户所有相关文件服务器中的局部空间，凡局部新增的空间都从全局空间中获取；凡局部释放的空间都在全局空间里回收，如此可以保证各个文件服务器的存储空间能动态地扩大和缩小。然而，系统中每个文件服务器只了解自身占用的空间，对其它文件服务器中的磁盘使用情况是不可见的。因此，必须要有一个能处在用户层次的高层视图，以便能对整个磁盘配额有一个总体的认识。本文将全局空间放在根服务器中，便可掌握所有文件服务器磁盘空间使用情况，更能够对磁盘资源作统一的调度和灵活的分配。

对 4 个重要名词进行定义如下：

- (1) 用户磁盘配额：管理员给注册用户分配的磁盘总限额，其大小由管理员决定，一般是固定不变的。
- (2) 用户磁盘空间占用量：某用户在所有相关文件服务器中已全部占用的磁盘空间。
- (3) 文件磁盘配额：系统为某用户在某台相关文件服务器里分配的磁盘容量，其大小是动态可变的。
- (4) 文件磁盘空间占用量：某用户在某台相关文件服务器中已占用的磁盘空间。

KTSS 下用户存储空间的动态分配机制如图 3 所示。

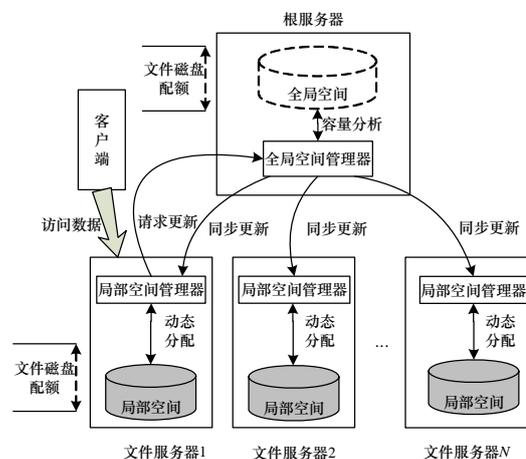


图 3 用户存储空间动态分配机制

