磁光阵列:一种新型数据存储模式的实现

刘建梅,宋林峰,冯剑楠,唐建华,徐 宏,邵征宇,许长江,朱 明

(苏州互盟信息存储技术有限公司,江苏 苏州 215151)

摘 要:介绍了一种磁盘与光盘融合数据安全存储的新模式,简称磁光阵列或 ORAID(Optical Replicated Arrays of Independent Disks)。ORAID 模式包括:将数据同时存储到磁盘模块和光盘模块中;可根据数据的属性信息从磁盘模块中读取数据,在目标磁盘无法被读取时,可根据数据在光盘模块中对应的位置信息,将存储有目标数据的光盘中的数据读入预设的数据缓冲区;将存储有目标磁盘数据的光盘中的数据存储到磁盘模块中更换后的新磁盘中。该模式在使用上像基于硬盘的系统一样方便,读写性能与 RAID(Redundant Arrays of Independent Disks)一样;在数据安全方面,无论前端硬盘如何损坏数据都不会丢失,并且在硬盘坏损以及重建过程中保持数据的可提供性;ORAID 具有物理层面上的合规性,所存的数据不可更改(WORM,Write Once Read Only)并可在无迁移的情况下长期保存 50 年以上;ORAID 无须额外的备份系统。磁光阵列模式同样适用于固态硬盘与光盘的融合。

关键词:磁光阵列; 光存储; 融合存储

中图分类号:TP23 **文献标志码**:A **DOI**:10.3788/IRLA201645.0935002

ORAID——The implementation of a new data storage mode

Liu Jianmei, Song Linfeng, Feng Jiannan, Tang Jianhua, Xu Hong, Shao Zhengyu, Xu Changjiang, Zhu Ming

(Suzhou NETZON Information Storage Technology Co., Ltd., Suzhou 215151, China)

Abstract: A system concept called ORAID (Optical Replicated Arrays of Independent Disks) was presented in which the hard disks and optical media were converged into high performance WORM(Write Once Read Many) appliances in a way so that no additional tier was added to the existing storage system architecture. An ORAID product typically includes a full-range of non-redundant hard disk array on the front end and an optical library in the background where the optical library is hidden and therefore invisible to the user interface. The hard disk array can be formed as a NAS (Network Attached Storage) merged together with a database-driven hidden library. The user can access their file systems as usual with all the ease and comfort without additional integration efforts. ORAID effectively uses the lifetime of each individual hard disk in the array so the data migration overhead and the data loss risk can be significantly reduced. Unlike the conventional redundant technologies like RAID (Redundant Arrays of Independent Disks) or erasure coding where the number of defective disks allowed is limited, there is no correlation between the hard disks, and data on an ORAID system will never be lost no matter how

收稿日期:2016-08-05; **修订日期**:2016-09-03

作者简介:刘建梅(1970-),男,工程师,主要从事嵌入式软件开发方面的研究。Email:jianmei.liu@hit-netzon.com.cn

通讯作者:朱明(1957-),男,研究员,博士,主要从事光存储技术方面的研究。Email:ming.zhu@hit-netzon.com.cn

many HDDs (Hard Disk Drives) are defect. Because of its physical nature the WORM/compliance will be ensured for optical media and this feature can also be realized for the hard disk front end by using HASH coding. In case of disk failure or any HDD absence the related data on the optical media will be automatically activated so that the data availability will be guaranteed when hard disks are somehow not accessible, e.g. during the hard disk rebuild process or replacement work. In an ORAID system, data on HDD and optical media is mutually protected. A health check process is keeping the integrity and safeguards against media failure. ORAID implements an integrated hybrid media storage in which the explicit backup process can also be eliminated. The logical consequence of greatly reduced workload for the optical drives and robotics will lead to less maintenance overhead on the library side and significant improvement of the working lifetime of the optical drives.

Key words: ORAID; optical storage; converged storage

0 引 言

信息化社会的发展进程中，人类需要存储的数据量正在持续迅速增长。在现有相对应的存储技术中，磁盘仍然是应用主流，在需要高速度数据存储的应用中，半导体固态硬盘正在被更加广泛地应用。

为了能相对高效、可靠地使用大量磁盘来存储数据，现有的技术是以冗余的方法将数据组织到多块磁盘中组成一个磁盘阵列。如此，根据冗余度的大小，在多块磁盘中有一块或几块磁盘出现坏损时，存储在磁盘阵列中的数据不会丢失[1]。

这样组成磁盘阵列仍然不能确保数据的安全，为了不丢失数据，通常的做法是对磁盘阵列添加一个备份系统，其功能是在磁盘阵列出现故障时能够将数据恢复到新建的磁盘阵列中去，数据备份所用的设备通常为磁带库、虚拟带库、镜像磁盘阵列等。备份系统无论是设备成本、管理成本，还是日常运维管理开销都是十分昂贵的。

当磁盘阵列用于归档存储时，存在一个不符合法规的问题，即存储在磁盘阵列上的数据理论上是可以删除或更改的。符合法规的做法是将数据归档存储在一次可写的介质上，即通常所述的 WORM (Write Once Read Many) 介质。光盘是可以满足 WORM 特性的存储介质，而且数据保存的寿命很长，通常大于 50 年，因此非常适合于数据的长期存储和归档存储[2]。

采用磁盘与光盘互相结合的存储方法已存在多年，通常采用分层存储的方法。此时系统将经常使用的数据放在高速度的存储器中，如固态硬盘和 SAS 磁盘，对于不经常使用的数据则逐步将其存入光盘，这种存储架构的问题是系统复杂度高，对于处于高速存储层的数据还要额外实施备份保护。此外，磁盘存储设备的寿命一般在 3~5 年，因此升级换代的开销很大，在光盘存储方面，光盘库处于分级存储的最底层，如何融入现有的存储架构，须视应用的具体要求，如此形成众多的系统解决方案，这样的应用方式比较费时费力，应用的效果也很大程度上受应用解决方案的影响。

文中提出了一种新型的磁盘与光盘融合存储的模式，即磁光阵列(ORAID，Optical Replicated Arrays of Independent Disks)，其主要包括一组磁盘、一组光盘模块、一套高可靠的机械手和一套软件控制程序，以网络附加存储 (NAS, Network Attached Storage)的形式提供网络数据存储服务。它可以很方便地被应用到各种不同操作系统平台上的信息系统中去。

1 背景技术

磁盘泛指硬盘，硬盘主要有固态硬盘(SSD，Solid State Disk，新式硬盘)、机械硬盘(HDD，Hard Disk Drive，传统硬盘)、混合硬盘(HHD，Hybrid Hard Disk，一块基于传统机械硬盘诞生出来的新硬盘)。SSD 采用闪存颗粒来存储，HDD 采用磁性碟片来存储，HHD 是将磁性硬盘和闪存集成到一起的一种硬盘。绝大多数硬盘都是固定硬盘，被永久性地密封固

定在硬盘驱动器中。

磁盘伴随着信息系统的发展进步至今,已经达到了很高的技术水平，其主要优点是随机存储速度快,单位体积存储容量大;其主要缺点是寿命有限且不可预测,一般 3~5 年需要更换一次,由于磁盘的机械旋转会消耗较多的电能并发热，为此需要设置散热装置,通常用风扇将热排出到设备外,空调装置用于抵消设备所发出的热能,这样同样会消耗大量的电能。

随着数据中心的存储规模日益增大，一个新的问题也逐渐显现出来,而且正在变得日益严重:当冗余磁盘阵列中的某地磁盘坏损时，需要及时进行更换,对于更换后的硬盘,还要将原先的数据再恢复上去,这个过程被称为重建。问题在于这个重建的过程非常耗时,根据应用场景和磁盘容量的不同,重建的过程往往会持续数小时到数十天,在此期间,如果再次出现磁盘坏损的情况则会导致数据丢失，在规模较大的数据中心,这种重建工作会频繁发生,除了耗费大量管理开销外,数据丢失的系统风险也在增加。

蓝光光盘(BD,Blu-ray Disc)利用波长较短(405 nm)的蓝色激光读取和写入数据[3],并因此而得名。而传统 DVD 需要光头发出红色激光（波长为 650 nm)来读取或写入数据,通常来说波长越短的激光，能够在单位面积上记录或读取更多的信息。因此,蓝光极大地提高了光盘的存储容量,对于光存储产品来说,蓝光提供了一个跳跃式发展的机会。到目前为止，蓝光是最先进的大容量光碟格式,BD 激光技术的巨大进步，使用户能够在一张单碟上存储 100 GB 的文档文件[4],是 DVD 存储量的 20 多倍。

BD 主要具有以下优点：波长更短-蓝光波长为 405 nm，红光波长为 650 nm；容量更大-由于多层技术和波长更短，存储容量迅猛增加；兼容性好-良好的技术规范，外形尺寸与 DVD 完全兼容；速度更快-蓝光达到 12×(54.0 MB/s)，红光只有 24×(32.6 MB/s)；超硬涂层-防止划伤,可以经受住频繁的使用、指纹、抓痕和污垢;高清影片-海量的存储空间使存储高清影像成为可能;保存更久-可保存 50 年、100 年甚至更长时间[5]。

如图 1 所示,磁盘与光盘存在明显的优势互补,其融合存储的模式可以为长期归档存储提供理想的物理空间。

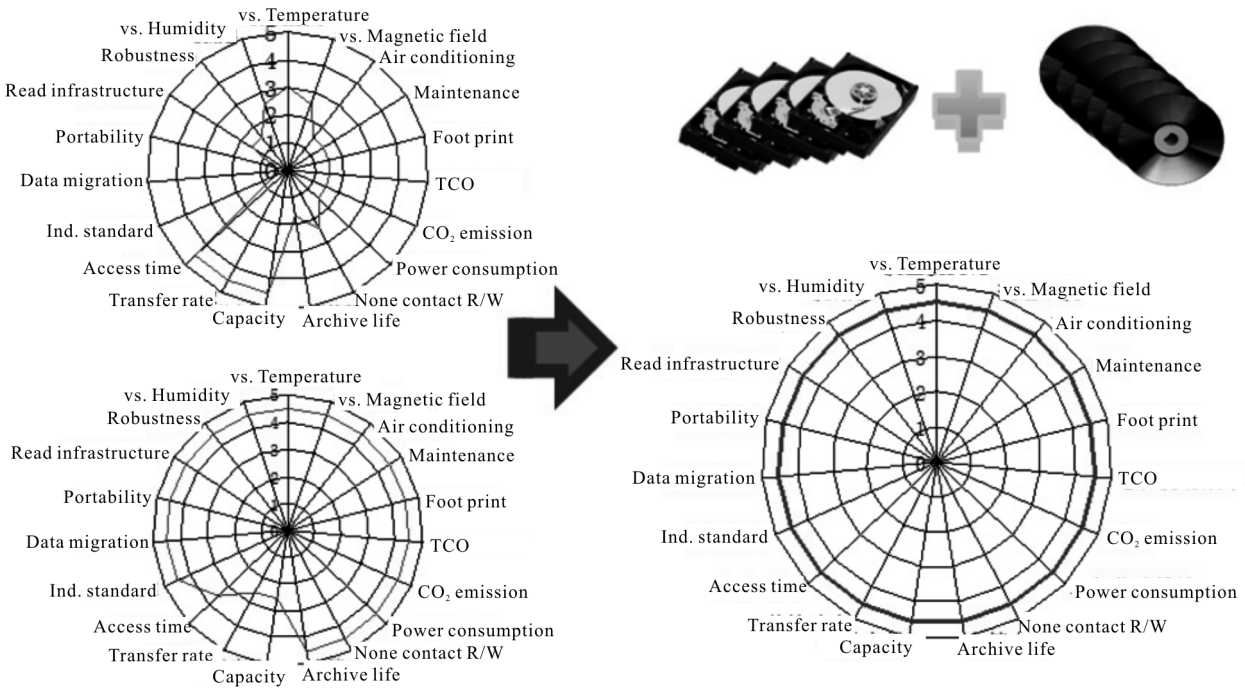

图 1 磁光融合存储的优势

Fig.1 Advantages of ORAID

2 磁光阵列的安全存储方法设计

(1) 将数据存储到磁盘模块中;

(2) 将上述磁盘中存储的数据在后台转存一份至光盘模块中的光盘中;

(3) 存储所述数据的属性信息以及所述数据在光盘模块中对应的位置信息;

(4) 根据数据的属性信息从磁盘模块中的目标磁盘中读取数据,在目标磁盘无法被读取时,根据所述数据在光盘模块中对应的位置信息,将存储有目标磁盘数据的光盘中的数据读入预设的数据缓冲区;

(5) 更新数据的属性信息和数据在光盘模块中对应的位置信息。

进一步地,还包括以下步骤:

(6) 将存储有目标磁盘数据的光盘中的数据存储到硬盘模块中更换后的新硬盘中。

ORAID 的设计原理如图 2 所示。

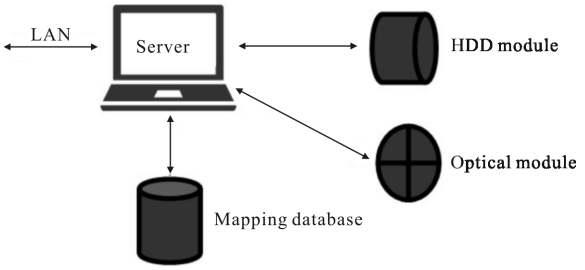

图 2 磁光阵列的组成

Fig.2 Composition of ORAID

3 技术实现

在 ORAID 系统中,磁盘模块和光盘模块处于同一个数据存储层中[6],两者之间由映射数据库建立了对应关系。其中光盘模块隐含于磁盘模块之后,用户看不到对光盘模块的任何操作。因此,光盘模块又被称为隐含光盘库,这是与分层存储的明显不同之处。

ORAID 技术实现的基础之一是随机多盘抓盘器技术。基本的技术要求是能在直接叠放的多张光盘中抓取任意数量的光盘, 从而实现快速随机调取光盘的任务。

3.1 光盘匣(Optical disk cartridge)

该光盘匣可放置 12 张 BD,无抽盘直接叠放,并带有锁定与解锁机构, 具有射频识别(RFID,Radio Frequency Identification Devices)标识,可防尘、避光防震,可交换。光盘匣关闭与打开时的图片分别如图3(a)、(b)所示。

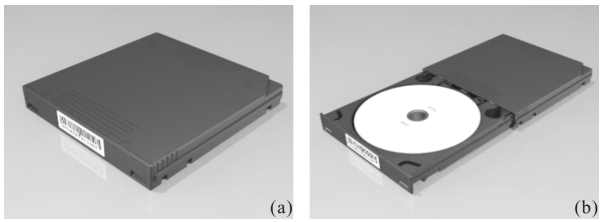

图 3 光盘匣

Fig.3 Optical disk cartridge

3.2 机械手操作单元

智能多盘抓盘器可随机抓盘, 实现了光盘匣与光驱之间的任意加载和卸载,如图 4 所示。

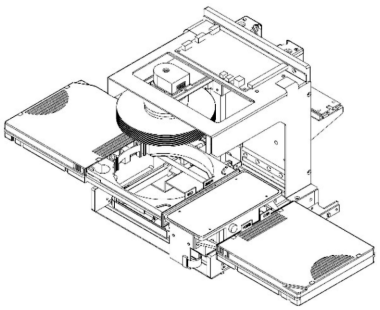

图 4 机械手单元

Fig.4 Unit of robot

3.3 设备形制

笔者在磁光阵列原理设计的基础上, 验证成功的样机 HDL24,如图 5 所示。

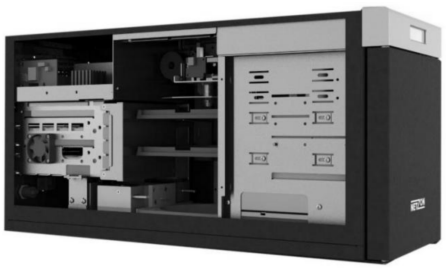

图 5 磁光阵列设备形制

Fig.5 Equipment shape of ORAID

3.4 软件控制程序

软件部分主要由以下模块组成:磁盘存储模块–用于将数据存储到磁盘模块中; 光盘存储模块–用于将所述磁盘中存储的数据存储至光盘模块中的光盘

中；磁光映射模块–用于存储所述数据的属性信息以及所述数据在磁盘和光盘模块中对应的位置信息；数据读取模块–用于根据数据的属性信息从磁盘模块中的目标硬盘中读取数据，在目标硬盘无法被读取时，根据所述数据在光盘模块中对应的位置信息，将存储在光盘中的数据读入预设的数据缓冲区；信息更新模块–用于更新数据的属性信息和数据在磁盘和光盘模块中对应的位置信息；数据重建模块–用于将存储有目标数据的光盘中的数据存储到磁盘模块中更换后的新磁盘中。

ORAID 技术具有以下优点，尤其适用于数据归档系统：

(1) 使用像基于磁盘的系统一样方便；

(2) 读写性能与 RAID (Redundant Arrays of Independent Disks)一样；

(3) 无论前端硬盘如何损坏数据都不会丢失；

(4) 长期数据安全(50 年以上)；

(5) WORM/物理层面上的合规性；

(6) 在硬盘坏损以及重建过程中保持数据的可提供性；

(7) 避免每隔 3~5 年的数据迁移；

(8) 无须额外的备份系统。

此外，基于 ORAID 技术的产品具有较高的经济性，尤其体现在长期的总体拥有成本(TCO)和更少的维护开销方面。

4 结 论

磁光阵列设置有磁盘模块和光盘模块，采用了异质冗余的技术，在将数据存储到磁盘模块的同时将数据写入光盘中备份，从而借助光盘存储的稳定性，增强数据存储的安全性。如果有磁盘发生故障无法读出，则从光盘模块将数据读入预设的数据缓冲区，并由此向外提供所要读取的数据；当硬盘模块中放入用于替换的新硬盘后，从光盘模块中读取存有故障磁盘数据的数据并存入新硬盘，从而在不影响数据读取的前提下实现硬盘的修复，有效地提高数据存储的效率和安全性。基于 ORAID 技术的产品在数据安全性、简单易用性和易维护性方面可为用户提供大的收益。

ORAID 原理与技术同样适用于 SSD/光盘的融合。ORAID 架构可以容易地横向扩展到现有设备，其中 HDD/光盘的配比可根据应用需求从 1:1 做任意的调整。ORAID 可以有效地降低云存储中的设备冗余度。在未来工作中将会研制更多基于 ORAID 的产品以适应更广泛的应用场景。

参考文献：

[1] Patterson D A, Gibson G, Katz R H. A case for Redundant Arrays of Inexpensive Disks (RAID)[C]//Proceedings of the ACM SIGMOD International Conference on Management of Data, 1998: 109–116.

[2] Zhao Weidong. Research on electronic archives Blu-ray storage practice [J]. ***Archives Sciences Study***, 2015 (3): 88–95. (in Chinese)
赵伟东. 电子档案蓝光存储应用探究 [J]. 档案学研究, 2015(3): 88–95.

[3] Stallinga S. Blu-ray disc [C]// Lasers and Electro-Optics Europe, CLEO/Europe, IEEE, 2005: 710.

[4] Thompson C. Optical disc system for long term archiving of multi-media content [C]// International Conference on Systems, Signals and Image Processing, IEEE, 2014: 11–14.

[5] Takeshima H, Yoshida H, Ueda C. Optical disk having a large storage capacity: EP, US5841757[P]. 1998.

[6] Feng D, Zeng L, Wang F, et al. ORAID: an intelligent and fault-tolerant object storage device [C]//International Conference on Embedded and Ubiquitous Computing. Springer-Verlag, 2005: 403–412.