

# 基于 ILM 的海量存储研究

## Study of Massive Storage System based ILM

申丽珍 朱嘉嘉 (温州大学 浙江温州 325000)

**摘要:** 随着存储技术的发展,海量存储已经越来越普遍。如何管理并使用好这些海量信息,成为海量存储面临的一个新的问题。本文通过数据都有自己的生命周期,引入 ILM 信息生命周期管理的海量存储模式,试图找寻能够解决海量存储中降低成本,保持数据完整性,提高数据使用率的一种方式。

**关键词:** ILM 信息生命周期 海量存储 虚拟存储

### 1 引言

2006 年 IBM 存储系统解决方案峰会上以美国医疗行业为例,解析了数据存储问题面临的一些问题,如:病人的信息需要共享,但病人的隐私需要保护;不

断治疗信息的缺失。另据美国伯克利大学的一项研究显示,美国公司每年产生 40 亿页的归档文件,相当于 140TB 的数据。从以上我们可以看出目前存储的问题主要在于如何根据数据的价值,运用不同的管理模

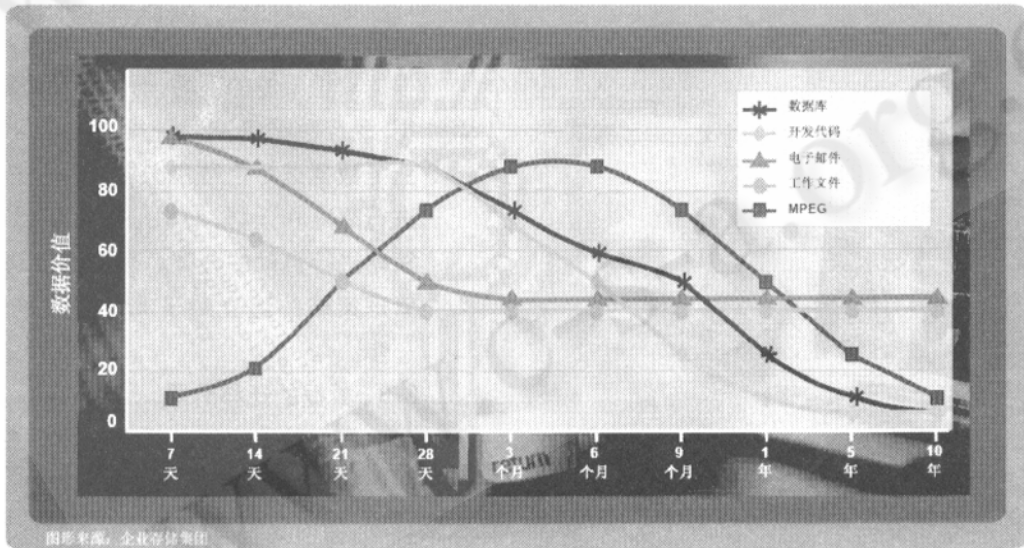


图 1 数据的生命周期

必要的金钱浪费在不必要的纸质介质上;医生常常要花费 20-30% 的时间在寻找和组织数据上;90% 的健康保健业务是通过电话,信件和传真,这些信息常常没有妥善保存;81% 的情况是医生没有获得有关病人的足够信息;20% 的化验及 X 光检查等是因为找不到原来的信息而重复做的;1/7 的住院是因为病人的以往的

式收藏和使用好海量数据。

研究发现数据有自己的生命周期(详见图 1),会随所处的“年龄”阶段的不同而在使用价值、频度、管理模式上起起落落。是否可以在海量存储技术中引入生命周期管理,使得存储更便于低成本的保存、高效率的利用呢?

## 2 生命管理周期(ILM)

信息生命周期管理(ILM)是以成本与信息不断变化的价值相对应的方法,从创建到最终处置对信息进行全程管理。有效的 ILM,还必须做到集中管理,以应

进行处理和分析。这样,用户将可以提高现有存储设备的利用率,利用自动化的 IT 数据管理技术实现自动的数据管理,减少企业的 IT 管理成本,满足政府和企业的数

## 3 ILM 的过程划分

信息的 ILM 的过程在存储上,通过理念与实践可以划分为这一套完整的模型,共分五个阶段。它们依次是:数据的创建阶段,数据的分类阶段,数据的迁移阶段和数据的保存、归档阶段,数据删除、销毁。(见图 2)

### 3.1 数据的创建阶段

数据的创建阶段是 ILM 的重要组成部分,ILM 系统的管理规则运用,在这一部分有着很好的体现。

### 3.2 数据的分类,安全保护阶段

基于存储成本的要求,数据并不都是生而平等的。数据应用的周期性规律,决定了可以根据其重要性的不同,而分为动态数据、静态数据、受控数据等不同种类。用户可以对不同的业务数据进行合理的分类,对一些非业务数据进行隔离和迁移,对核心的业务数据根据管理策略进行自动的分类。有效的自动数据分类和管理策略的建立对于 ILM 来说是极为关键的一步。

### 3.3 数据的迁移阶段

根据数据不同的重要性及功能,用户可以更好地迁移数据、维持设备的正常运转,降低数据恢复时间,同时获得更大的灵活性、性能和简易性,并且降低管理成本。简单来说就是根据数据的类型,自动将不同生命周期阶段的数据存放在最合适的存储设备上:如访问量较高的数据存放在一级存储上(SAN, DAS 存储设备),而对于访问量较低的则可放在二级存储上(磁带库或磁带机),对于长时间不用的数据可存放于三级存储上(一些离线存储设备,如光盘)。

### 3.4 数据的保存、归档阶段

用户可以通过文件存储系统和近线存储系统以及

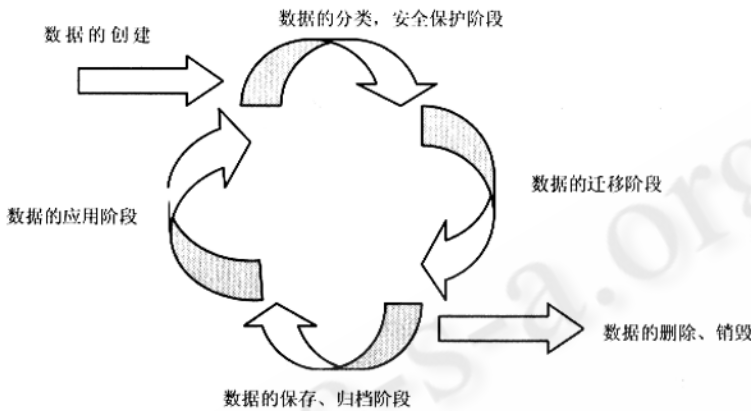


图 2 循环的生命周期管理(ILM)模型

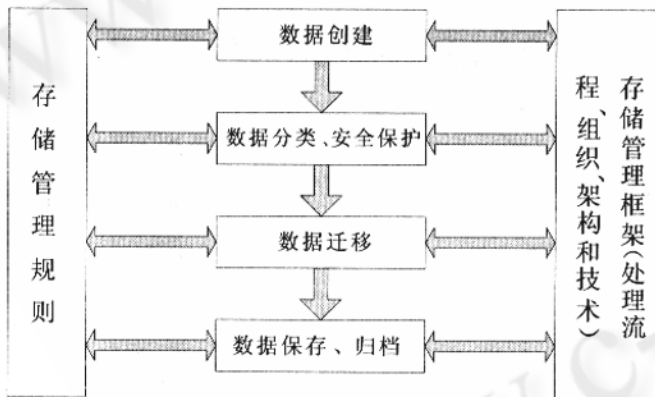


图 3 基于生命管理周期(ILM)存储架构思路图解

用和流程为指导。数据的生命可以分成如下个阶段:生成、保护(不丢失、一致)、读取(读取时间速度)、迁移、存档、删除。根据这样的数据生命轨迹,信息生命周期管理(ILM)就是识别“合适”的数据,将其存储在“合适”的介质上,在“合适”的时间用于相关应用,直至删除、销毁。

信息生命周期管理(ILM)就是对不同的业务数据进行贯穿其整个生命周期的管理,通过完整的信息生命周期管理解决方案,可以让不同类型的数据存放在适合的存储设备上,利用适当的技术手段对这些数据

在线、近线和离线相结合的方式,合理地保存、归档自己的数据,同时保持最高的数据可用性和管理简便性。

### 3.5 数据的应用阶段

在以上几个过程的基础上,用户能以最高的效率,最低的成本,从容应对竞争环境下不同的数据应用需求。同时保持数据的高可靠性,做好及时的备份与更新。

### 3.6 数据的删除、销毁

数据的最终归宿。大部分的数据由于利用价值的丧失,而最终被删除、销毁。

## 4 基于 ILM 的海量存储

### 4.1 基于 ILM 的海量存储思路

在仔细分析了生命周期管理(ILM)的属性后,不难看出使用 ILM 的存储架构中只要围绕其数据的不同生命阶段,相应结合存储管理规则和存储管理框架(处理流程、组织、架构和技术),就可以得到如下所示存储架构思路。

时数据写入,以及未来可能的数据读取操作,IP 网络存储的低效率必然无法完成该部分功能实现,而 SAN 网络存储效率不成问题,但是如何保证多个节点的同时读写将成为难点之一。);针对如此大的数据容量,如何经济有效的存储和管理这些数据(随着数据保存到磁盘中,保存的时间越久,数据的价值越低,被访问的可能性也就越发降低,如何充分利用资源);在视频数据的生命周期管理的过程中,如何保证数据的稳定、可靠以及自动的管理;如果视频流媒体文件可以保证其存储使用的无障碍,那么其他形式的文件问题就可迎刃而解。

基于上述的几点考虑,有如图 4 之基于 ILM 的海量存储架构。

在此架构中,为了更方便的引入 ILM 生命管理周期,这里也使用的虚拟存储的技术。虚拟存储技术就是把多个存储介质模块(如硬盘、RAID)通过一定的手段集中管理起来,所有的存储模块在一个存储池(Storage Pool)中得到统一管理,从主机和工作站的角度,看到就不是多个硬盘,而是一个分区或者卷,就好像是一个超大容量(如 1T 以上)的硬盘。这种可以将多种、多个存储设备统一管理起来,为用户提供大容量、高数据传输性能的存储系统称为虚拟存储系统。也就是说通过虚拟存储技术把同一存储平台、磁盘阵列、智能带库等各类型存储设备通过光纤网络连接起来,形成一个在管理者的眼中非常易于管理和操作的虚拟存储池。所有的数据都汇聚在这虚拟存储池中。虚拟存储技术也是海量存储得以实现的先决条件。

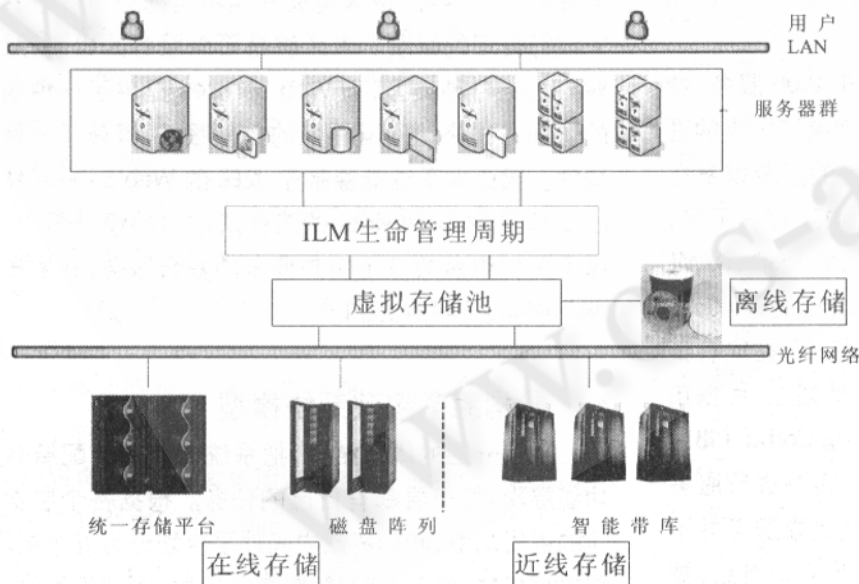


图 4 基于 ILM 的海量存储架构

### 4.2 基于 ILM 的海量存储架构

海量存储的难点是视频流媒体文件,这类文件要求保证视频数据能够迅速的存放到介质中,在使用时可以非常方便的调用出来(针对如此大规模的多点同

看到就不是多个硬盘,而是一个分区或者卷,就好像是一个超大容量(如 1T 以上)的硬盘。这种可以将多种、多个存储设备统一管理起来,为用户提供大容量、高数据传输性能的存储系统称为虚拟存储系统。也就是说通过虚拟存储技术把同一存储平台、磁盘阵列、智能带库等各类型存储设备通过光纤网络连接起来,形成一个在管理者的眼中非常易于管理和操作的虚拟存储池。所有的数据都汇聚在这虚拟存储池中。虚拟存储技术也是海量存储得以实现的先决条件。

这样再引入 ILM 生命周期管理的策略。策略应当可以自动对业务数据的访问周期进行统计并根据访问统计结果自动对数据进行分层

(下转第 51 页)

存储;引入文件系统虚拟化的解决方案可以根据不同的业务数据根据存储设备的特点建立不同的存储池,处于不同周期的业务数据可以自由的迁移到不同级别的存储池中,实现存储资源的优化,提高数据的访问效率;对于处于生命周期末期的文件类型业务数据,可以进行归档并保存在适合长期保管数据的存储设备中;通常归档的数据都是比较关键的历史数据,对归档的数据应提供更高层次的保护,只有在满足特定的条件前提下,系统管理人员才可以对这些数据进行修改,删除作业,以满足较高级别的要求。

## 5 小结

ILM 生命管理周期的存储解决方案将数据在不同存储层次之间移动,降低成本,保持数据完整性,提升数据可用性。该方案帮助用户根据数据重要性程度的变化,将数据迁移到更便宜的存储设备上,降低数据保存成本,使大量数据在整个生命周期内保持高可用的同时,释放昂贵的高性能磁盘空间。这无疑是海量存

储的一个有力支撑。

## 参考文献

- 1 FalconStor Software, Inc. "The Value of Storage Virtualization". MAY 2002.  
<http://www.falconstor.com/Whitepapers/IPStorVirtualizationWhitepaper.pdf>
- 2 C Lumb, A Merchant, G A Alvarez、Facade: Virtual storage devices with performance guarantees[ C]. USE-NIX Conf on File and Storage Technologies ( FAST ' 03 ), San Francisco, CA, 2003.
- 3 Store Age White Paper, "High - Performance Storage Virtualization Architecture ", <http://www.storage.com>
- 4 IBM Company, "whitepaper - ILM", 2006.
- 5 Hewlett - Packard Company, " HP Storage Apps sv3000 White Paper", 2002.