

# 基于 C-ILM 的工程存储管理体系架构

常 培<sup>1</sup>, 王红梅<sup>2</sup>, 白中英<sup>1</sup>

(1. 北京邮电大学计算机科学与技术学院, 北京 100876; 2. 国家卫星气象中心, 北京 100081)

**摘 要:** 在分析信息生命周期管理的概念和特点的基础上, 该文提出了一种新的信息生命周期管理立方体模型, 该模型描述了信息生命周期管理的 3 项技术基础(分级存储、自动化网络存储及虚拟存储)、信息生命周期管理的软硬件层次结构及信息生命周期管理的 6 个阶段(创建、保护、访问、迁移、归档及回收)。基于该模型, 建立了一个卫星地面应用系统的存储管理体系架构。

**关键词:** 信息生命周期管理; 信息生命周期管理立方体模型; 虚拟 SAN; 分级存储; 存储管理体系架构; 卫星地面应用系统

## Project Storage Management Architecture Based on Cube-information Lifecycle Management

CHANG Pei<sup>1</sup>, WANG Hong-mei<sup>2</sup>, BAI Zhong-ying<sup>1</sup>

(1. School of Computer Science and Technology, Beijing University of Posts and Telecommunications, Beijing 100876;

2. National Satellite Meteorological Center, Beijing 100081)

**【Abstract】** On the basis of analyzing the concept and feature of ILM, this paper puts forward a new cube model(C-ILM). The model depicts three kinds of basic technology(hierarchical storage, ANS, virtual storage), hardware-software hierarchy and six phases (creation, protection, access, transfer, archive, deletion) of ILM. And a storage management architecture of satellite ground application system based on C-ILM is implemented.

**【Key words】** information lifecycle management (ILM); cube-information lifecycle management(C-ILM); virtual SAN; hierarchical storage; storage management architecture; satellite ground application system

在卫星地面应用系统工程项目中, 数据存储系统负责工程中各类数据的存储、备份和应用服务, 是整个工程的数据管理中心。如何合理地存储和有效地管理数据, 为业务系统和应用系统提供高质量的数据服务, 成为其设计的一个重点问题。

信息生命周期管理(information lifecycle management, ILM)是数据存储管理供应商 Storage Tek 针对海量数据存储需求推出的先进存储管理理念。目前公认的定义是: 将自动化网络存储基础设施与其综合的解决方案结合在一起, 然后以此为基础制定出新的信息管理策略。这种新的信息管理策略就称为信息生命周期管理<sup>[1]</sup>。ILM 的目标是确保在信息生命周期的各个阶段将信息存储在最适合的介质上, 提供所要求的服务级别<sup>[2]</sup>。它是一种综合了信息流程、存储技术、基础设施和应用软件的旨在高效管理数据和信息的策略。

### 1 C-ILM 及工程存储管理体系架构实现方案

基于 ILM 的理念, 本文提出一种新的 ILM 立方体模型(C-ILM), 并通过该模型实现了某卫星地面应用系统的存储管理。如图 1 所示, C-ILM 由技术基础维、软硬件层次结构维和信息生命周期维组成。技术基础维是存储系统实现 ILM 所必须采用的 3 项存储技术, 分为分级存储、自动化网络存储、虚拟存储 3 个层次; 软硬件层次结构维从存储系统的基础设施和上层软件的角度, 分析了实现 ILM 需要的软硬件支持; 信息生命周期维是信息在一个生命周期中随着价值的变化所处的各个不同的阶段, 包括创建、保护、访问、迁移、归档、回收 6 个阶段, 建立在技术基础维和软硬件层次结构维之上。

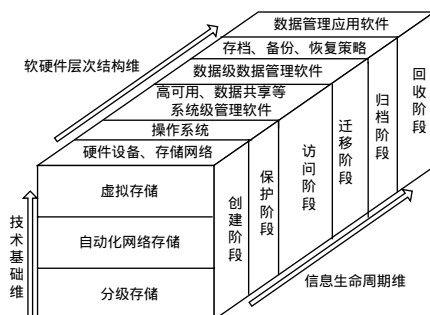


图 1 信息生命周期管理立方体模型 C-ILM

### 2 技术基础维

在技术基础维中, 分级存储、自动化网络存储和虚拟存储是逐级递进的关系。自动化网络存储建立在分级存储之上, 而虚拟存储建立在分级存储和自动化网络存储之上。在自动化的网络存储环境中, 只有对存储设备分级, 才能充分发挥存储网络的优势, 使数据在存储网络的各个不同节点之间按需流动; 既然存储网络中有分级存储, 就有不同厂商、不同类型的存储设备, 因此, 如何将所有异构的存储设备“黑盒”化, 对业务系统和应用系统提供透明的访问, 虚拟存储提供了更高层次的逻辑视图, 将所有的异构存储设备统一成一个

**作者简介:** 常 培(1981 - ), 女, 硕士研究生, 主研方向: 网络存储理论与技术, 计算机体系结构; 王红梅, 高级工程师; 白中英, 博士生导师

**收稿日期:** 2006-09-25

**E-mail:** changpei.bupt@gmail.com

大的存储池，集中进行管理。

## 2.1 分级存储

分级存储(hierarchical storage, HS)就是根据数据的重要性、访问频次等指标，采用不同的存储方式，将数据存储在不同性能的存储设备上。卫星地面应用系统采用三级存储体系结构：在线存储系统，近线存储系统及离线存储系统<sup>[3]</sup>。合理的三级数据存储结构能够从根本上缓和速度与容量的矛盾，存储体系结构如图 2 所示。在线、近线及离线存储系统的介质容量根据数据总量的大小及业务系统和应用系统的需求来定。

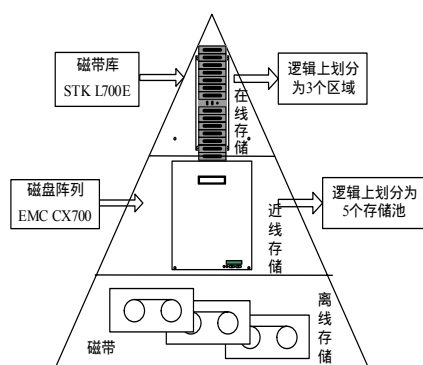


图 2 存储系统三级存储体系结构

### (1)在线存储系统

在线存储系统中的存储的数据时刻保持“在线”状态，用户可以随时访问，访问延迟时间较近线和离线存储系统短，但存储的数据量与期限受存储设备容量的制约。卫星地面应用系统采用 EMC CX700 磁盘阵列作为在线存储设备。将磁盘阵列按需求逻辑上划分为 3 个区域：分系统业务数据私用区，数据库库体存放区及数据在线交换及存储区。

在线数据管理规则采用在线存储时限、磁盘空间、访问频率(比如最近最少使用算法)3 种策略。比较特殊的是数据库库体存放区，对需要永久保存的数据，若该区空间不足，则利用数据库软件 Oracle 的自带功能将历史数据备份到指定的备份空间。

### (2)近线存储系统

近线存储系统是以容量大、价格便宜、性能较低的介质存放那些访问频率低、并且对访问速率要求不高的在线数据。它的访问延迟时间介于在线和离线存储系统之间。卫星地面应用系统采用 STK L700E 自动磁带库作为近线存储设备。为便于管理，使用企业级备份管理软件(VERITAS NBU, Veritas)。配置工具将磁带库划分为 5 个存储池：0 级数据主存储池，1 级数据主存储池，2 级数据主存储池，3 级数据主存储池及 4 级数据主存储池。每种数据类型对应主和备 2 个存储池，每个存储池可动态定义若干磁带，并在 Veritas 中定义每个存储池的存档策略(policy)和存档计划(schedule)。依据数据类型将数据存入预先定义的存储池中。存储池的划分策略依照“高内聚，低耦合”的原则，即同一个存储池中的数据属于同一类型，不同存储池中的数据属于不同类型。一方面便于同种类型的数据集中管理；同时也考虑到 I/O 性能，磁带是顺序读取，所以，当访问同一类型的若干数据时，可以在一个存储池或一盘磁带上找到所有的数据。

近线数据管理规则采用设置近线存储时限，将超过时限的数据离线保存。

### (3)离线存储系统

离线存储系统依据预先制定的离线策略将近线存储系统中的部分数据(一般为超过近线存储时限、且需要长期或永久保存的数据)离线保存。离线就是将磁带从磁带库中取出，放到指定的磁带柜中。当需要访问该类数据时，先将磁带重新入库再回调到在线磁盘阵列，所以访问延迟时间较长。

离线数据管理规则采用将磁带离线后的信息(比如柜名、层数和列数)记入数据库中，以便使用时快速找到数据所在磁带的位置。

## 2.2 自动化网络存储

自动化网络存储是信息生命周期管理的基础<sup>[4]</sup>。网络存储技术将网络技术与存储技术相结合，使存储系统独立于服务器，实现了服务器数据处理和数据存储的分离。而网络存储的自动化则依赖于网络设施之上的软件来完成的。

鉴于 SAN 传输速率高、传输距离远、服务器和存储设备可动态添加、文件迁移和备份及恢复都不需要通过服务器等优点，卫星地面应用系统采用 SAN 技术构建数据存储系统。服务器与存储设备之间、存储设备与存储设备之间通过光纤通道(fiber channel)、光纤交换机和集线器连接在一起，组成一个高速的存储局域网<sup>[5]</sup>。网络架构如图 3 所示。网络中 8 台 IBM 主机端(已对 3 台物理服务器进行了逻辑分区)均通过 2 台 EMC DS4700M 光纤交换机联入 SAN，EMC CX700 的 8 个光纤端口(对应 8 个主机端)和 STK L700E 的 3 个光纤端口(对应 3 个磁盘驱动器)与该 2 台 EMC DS4700M 光纤交换机连接。

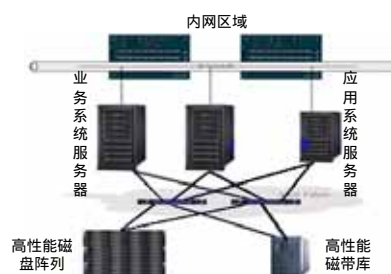


图 3 存储系统的 SAN 架构

## 2.3 虚拟存储

虚拟存储(virtual storage)就是将服务器操作系统的存储描述与实际物理存储设备相分离，服务器只与分配给它的虚拟卷打交道，而不关心其数据是在哪个物理设备上。虚拟卷与物理实体之间的映射关系，是由安装在服务器上的存储管理软件(称为基于主机的虚拟化)或存储设备上的管理器(称为基于存储设备的虚拟化)，或加入存储网络的虚拟化设备(称为基于网络的虚拟化)来实现的<sup>[6]</sup>。

卫星地面应用系统采用基于 SAN 网络的虚拟存储。SAN 中的虚拟存储服务器和一些网络中间件把应用服务器和存储设备相分离，将不同厂商的存储设备逻辑上统一成一个存储池。物理磁盘经过 RAID 控制器，生成一个个的逻辑卷(LUN)，根据其属性对 LUN 分类，低速存储设备上的卷分在“低性能组”，高速存储设备上的卷分在“高性能组”。在这些组上建立服务器能够看到的虚拟卷，这些虚拟卷可以跨在多个 LUN 上以获得足够的性能和空间。每个虚拟卷被指定给 SAN 上的一个或者多个应用服务器。所有的存取操作均由虚拟存储服务器统一调度、控制和管理。通过建立一个基于 SAN 的文件系统，不同的操作系统可以共享这个文件系统，从而实现了信息系统集成和信息资源共享的目标。

## 3 软硬件层次结构维

软硬件层次结构维从底层的基础设施到上层的各种软件阐述了实现 ILM 所需要的软硬资源支持。下面给出卫星地面应用系统的软硬件配置实例，说明这些硬件和软件如何相互配合为最上层的卫星地面应用系统数据管理应用软件提供服务的。

硬件设备诸如服务器(IBM P570/550)、磁盘阵列(EMC

CX700)、磁带库(STK L700E)以及连接存储设备和服务器的 SAN 网络作为 C-ILM 的底层基础设施。操作系统(IBM AIX5.3)构建在基础设施之上,控制和管理其下层的硬件资源和上层的软件资源。高可用集群管理软件(HACMP 5.2)、文件系统共享软件(GPFS)、系统运行监视软件等系统级管理软件运行在操作系统之上,为上层数据级数据管理软件和数据管理应用软件提供安全可靠的集群、共享网络运行环境。Oracle 数据库软件、Veritas 存储软件等数据级数据管理软件运行在系统级管理软件之上,完成存储、备份及相应的数据信息的入库操作。根据业务需求制定相应的存档、备份、恢复策略,为最上层数据管理应用软件提供相应的策略支持。

## 4 信息生命周期维

在信息生命周期维中,信息经历了创建、保护、访问、迁移、归档、回收 6 个阶段。信息创建阶段用存取速率高、性能较好的存储介质来及时保存生成的数据,并提取相关的元数据信息;信息保护阶段依据数据的等级采用不同的数据保护措施,如备份技术、复制技术等;信息访问阶段实现在线、近线或者离线数据的访问;信息迁移阶段将访问频率降低但仍要求高性能 I/O 的数据从一线存储介质转移到二线存储介质;信息归档阶段将访问频率低且需要长期或永久保存的数据存储到容量大且性能低的三线存储介质上;信息回收阶段将没有价值的数据进行回收或者销毁。

### 4.1 C-ILM 信息生命周期维工程设计

依据 C-ILM 信息生命周期维的设计思想,卫星地面应用系统的存储管理按以下 6 个阶段实施:

(1)信息创建阶段。所有送到存储系统进行存档的数据先放到磁盘阵列的临时存储区(临时存放数据的区域),当存储系统收到数据到达通知后,将数据从临时存储区拷贝到滚动存储区(有存储时限的区域)或者永久存储区(可以永久保存数据的区域)进行在线存档,并进行数据的一致性检验(包括文件名、大小、生成时间等信息)。对检验合格的数据提取各类属性信息,检验不合格的数据从临时存储区拷贝到回收站区(不合格产品存放区)。

(2)信息保护阶段。通过建立全方位、多层次的企业级的备份系统来建立可靠的信息资源容灾环境:1)使用硬件备份来防止硬件故障。主机采用双机集群结构的高可用配置,正常情况下,双机同时运行,承担不同的处理任务,异常时由定义的备用机接替主机运行业务,实现非对称的双机冗余备份;2)由于软件故障或人为误操作造成数据的逻辑损坏,因此使用 SAN 的存储备份系统和硬件容错相结合的方式实现对系统的多级防护,有效防止了数据的物理损坏和逻辑损坏。

(3)信息访问阶段。采用三级存储体系结构:在线,近线和离线,满足业务系统及应用系统对数据高速存取的要求。在生命周期初期,数据经常被存取,也是数据最有价值的时期,因此,将数据存储在本成本最高的在线磁盘阵列;当数据进入生命周期中期时,其价值也随之下降,此时将数据移至近线磁带库存储;当数据进入生命周期末期时,则移至离线设备存放,一旦需要,可以随时将其回调到在线磁盘阵列。

(4)信息迁移阶段。为满足业务系统和应用系统的存储需求,使用性能稍低的二线磁盘阵列作为一线磁盘阵列的补充,将访问频率稍低的数据进行二线磁盘阵列存储。此阶段的信息迁移都位于在线存储系统,只是在 I/O 性能和成本之间作了均衡。

(5)信息归档阶段。依据预先制定的在线存储时限将磁盘

阵列上数据存储到磁带介质上,此过程又称为近线存档。

(6)信息回收阶段。依据制定的存储管理策略,对业务系统和应用系统没有使用价值的信息要进行回收或者销毁。

### 4.2 C-ILM 信息生命周期维工程实现

下文简要说明卫星地面应用系统信息归档阶段的实施过程。假定需求分析中要求近线存档(即归档)包括主存档和备份存档(即一个文件在磁带库的主存储池和备份存储池各存一份)。在应用程序中调用 Veritas 的命令行接口实现磁盘阵列到磁带库的存档。首先确定基于 Veritas 的存档方法,使用 10 个大小为 1 856 729KB 的文件进行试验。

**方案 1** 主存档、备份存档以列表从 Veritas 后台批量提交。

$t_{主}=23'43''$ ,  $t_{备}=22'41''$ ,  $t_{总}=46'24''$ 。

**方案 2** 主存档、备份存档以单个文件从 Veritas 后台提交。

$t_{主}=25'56''$ ,  $t_{备}=26'40''$ ,  $t_{总}=52'36''$ 。

**方案 3** 主存档、备份存档在策略(policy)配置中绑定后以单个文件在 Veritas 后台提交。 $t_{总}=26'07''$ 。

考虑存档消耗的时间因素,采用在 Veritas 的 policy 中配置主备绑定存储(即每提交一个存档文件,Veritas 会在相应的主、备存储池各存一份)。近线存档实现的伪码如下:

```

输入  strFilename 待存档文件名
      strPolicy    存档策略名
      strSchedule 存档计划名

Operation
AMSArchive(strFileName,strPolicy,strSchedule)
/* 超时信号(SIGALRM)和中断信号(SIGTERM)处理,
ICNTIMEOUT 为超时时限,SignalDisp()为信号处理函数*/
alarm(ICNTIMEOUT);
signal(SIGALRM,SignalDisp(ALARM));
signal(SIGTERM,SignalDisp(TERM));
/*依据存档策略名 strPolicy 和当前机器时间生成此次存档作业的
关键字 strKeyword*/
GenerateKeyword(strPolicy,strKeyword);
/*依据输入参数生成相应的存档命令 strArchCmd
("/usr/opensv/netbackup/bin/bpbackup -L /home/changpei/bplist.log
-p%s -s%s -k '%s' -S dms2 -w 00:10:00 %s2>/home/changpei/ bplist.err")*/
GenerateCmd(strPolicy,strSchedule,strKeyword,strFileName, strA
rchCmd);
/*执行存档命令 strArchCmd*/
SystemCall(strArchCmd);
/*取本地机器时间作为存档时间 strTime*/
GetLocalTime(strTime);
/*取错误信息文件的属性,structFileStatus 定义为 struct stat 类型*/
stat("/home/changpei/bplist.err",&structFileStatus);
/*错误信息文件大小不为 0 时,本次任务主、备都失败*/
If(structFileStatus.st_size!=0) Then
/*解析错误信息文件的内容,并上报任务状态消息*/
AnalyzeErrFile("/home/changpei/bplist.err");
/*否则,主备存档全部成功或者主、备其中一个失败*/
Else
/*依据输入参数生成相应的取磁带信息命令
strImageCmd("/usr/opensv/netbackup/bin/admincmd/bpimagelist -L
-d %s 00:00:00 -e %s 23:59:59 -keyword '%s'>/home/changpei/image
list.txt")*/
GenerateCmd(strTime,strKeyword,strImageCmd);
/*执行取磁带信息命令,执行此命令必须使用 root 权限*/
SystemCall(strImageCmd);

```

(下转第 89 页)