

容灾备份技术、产品和应用现状的分析与研究

王浩铭, 穆道生, 周勇

(装备学院, 北京 101416)

摘要:该文首先结合不同应用背景给出了容灾备份技术的系统结构,对位于该系统结构中不同层次的容灾备份技术进行了详细的阐述,其次分析了本地容灾、异地容灾、数据容灾和应用容灾方案的工作原理和各自特点,最后给出对应的产品和方案并完成对不同容灾备份方法的分析与比较。

关键词: 容灾备份; SAN; ISCSI; RAID; 灾难检测

中图分类号: TP302.8 **文献标识码:** A **文章编号:** 1009-3044(2014)13-2914-05

Analysis and Research of Disaster Backup Technology, Product and Application

WANG Hao-ming, MU Dao-sheng, ZHOU Yong

(Equipment Academy, Beijing 101416, China)

Abstract: Firstly, a system structure of disaster tolerance technology was put forward combining with different application background and disaster backup technology of different levels of the system structure was detailed introduced in this paper. Secondly, the working principle and the characteristics of local disaster tolerant plan, remote disaster tolerant plan, data disaster tolerance plan and application disaster tolerance plan were analyzed. Finally, the corresponding products and solutions of different disaster backup method were put forward and compared.

Key words: disaster recovery and backup; SAN; ISCSI; RAID; disaster detection

1 概述

计算机技术的发展可以分为三个方向,第一个方向围绕计算处理技术展开,推动计算机技术的普及应用;第二个方向围绕信息传输技术展开,推动信息数字化时代的到来;第三个方向围绕信息存储技术展开,逐渐提升信息的安全性和高可用性。近年来随着社会对信息的依赖程度日益加深,对第三个方向也就是信息存储方向容灾备份技术的研究逐渐成为研究的热点。容灾备份是保护IT系统中重要数据和保障系统业务连续性所做的冗余工作,按照距离分为本地和异地容灾,按照功能分为数据级容灾和应用级容灾。该文介绍并分析不同应用背景下的容灾备份技术,同时例举出所对应的市场产品和解决方案^[1]。

2 容灾备份技术的体系结构

本文从容灾备份不同的应用背景,将容灾系统分为本地数据容灾、本地应用容灾、异地数据容灾和异地应用容灾^[2]。提出容灾备份技术体系结构,如图1所示。

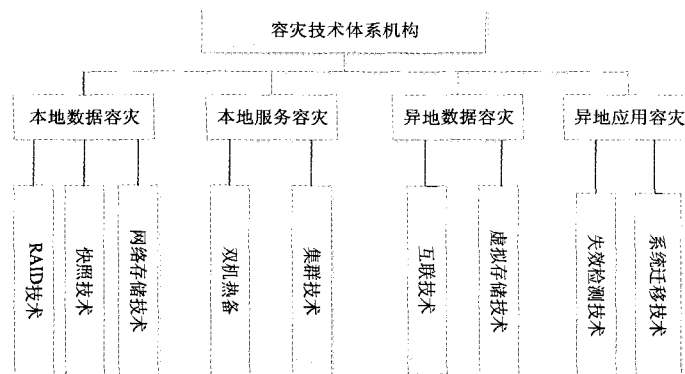


图1 容灾备份技术体系结构

收稿日期:2014-04-14

作者简介:王浩铭(1989-),男,吉林永吉人,研究生,主要研究方向为容灾备份应急服务应急服务;穆道生(1966-),男,黑龙江巴彦人,教授,硕士,主要研究方向为容灾备份应急服务应急服务;周勇(1989-),男,四川冕宁人,硕士研究生,主要研究方向为航天测控。

3 本地容灾

本地容灾适用于区域范围小、容灾级别并不很高的环境,通过对业务系统的有效冗余和对失效点的切换来实现。包括两方面内容,一方面通过各个环节要素自身的可靠性来实现本地数据容灾,另一方面通过生产中心冗余系统实现本地应用容灾。

3.1 本地数据容灾技术

RAID^[3](磁盘阵列)提出将若干磁盘条理化组成一组虚拟的整体存储空间,从而可以执行磁盘并行I/O操作,提高存储性能。一般分为六种结构分别是RAID0~5,RAID0在存储性能上得到提升,没有容灾作用;RAID1即镜像技术,通过额外数据拷贝达到容灾的效果。工作原理如图2所示:生产系统发出数据写命令,并传给镜像系统,镜像系统执行数据镜像任务,拷贝数据到介质,镜像技术也可以利用互联技术如SAN等达到远程镜像目的;RAID2利用海明码纠错布局,在RAID1的基础上增加一个校验盘起到容灾目的,存储粒度为bit;RAID3~5利用校验布局,当一个磁盘发生故障可以通过磁盘之间的校验关系还原失效磁盘数据。

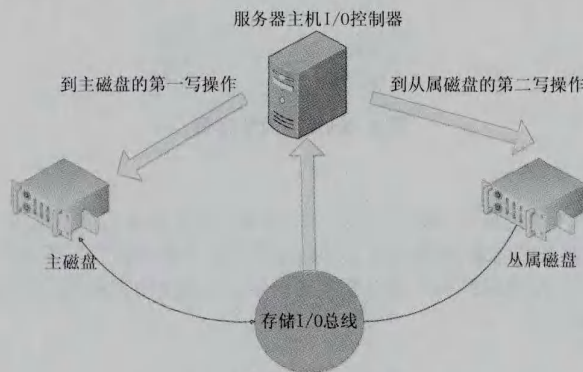


图2 镜像技术示意图

3.1.1 网络存储技术

传统存储方式为DAS(Direct-Attached Storage)方式,即服务器直接与存储设备连接,这种方式已经难以满足扩展性和高可用性的需求,因此网络存储技术得到迅速的发展,包括NAS(Network-Attached Storage),SAN(Storage Area Network)和基于IP的存储。

NAS是以网络文件服务器为模型,利用以太网的数据访问技术。用户可以直接使用它预先配置好存储空间和进行网络文件访问。NAS访问速度优于DAS,而且NAS发生故障并不影响主服务器的其他操作,具有一定容灾能力。NAS的结构如图3所示,拥有异构型强,利用现有LAN网络资源节约预算,易安装和广泛的连接性等优点。不足之处是文件访问速度不能满足在线处理,可扩展性差,数据备份占用LAN带宽等。

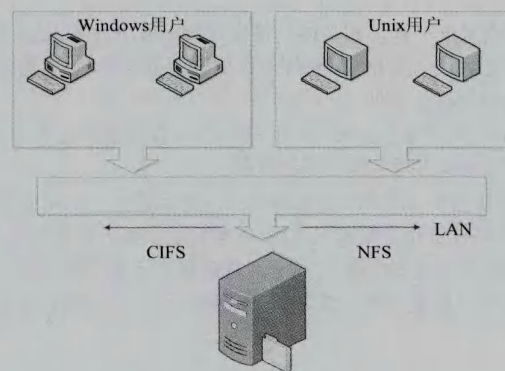


图3 NAS机构示意图

SAN一种利用FC(Fibre Channel)等高速互联协议连接起来的,可以在服务器和存储系统之间,或者在存储系统之间直接传送数据的网络。SAN如图4所示,在原有LAN的基础上建立一个专用的存储网络,分离了存储和传输避免网络拥塞。独特的结构设计和专用协议是SAN具有高速存取,良好扩展性等优点。但是利用FC协议传输的FC-SAN也存在局限性,包括FC协议下不同厂商设备互操作性较差,系统构建费用昂贵,存储系统异构兼容困难等,严重阻碍了SAN的发展。当传统FC-SAN发展遇到瓶颈是,通过IP网络协议完成SAN互联成为研究热,这一部分将在互联技术详细讲解^[4]。

3.1.2 快照技术

快照技术^[5](Snapshot)是指关于指定数据集合的可用拷贝,它包括相应数据在某个时间点(的映像),能够进行在线数据备份与恢复,为存储用户提供了另外一个数据访问通道。快照的实现层次一般是在卷管理层和物理层,分为第一次复写时复制和I/O重定向两种方式。快照通常完成本地存储系统信息备份到本地磁带库的功能,完成以数据快速复制过程。

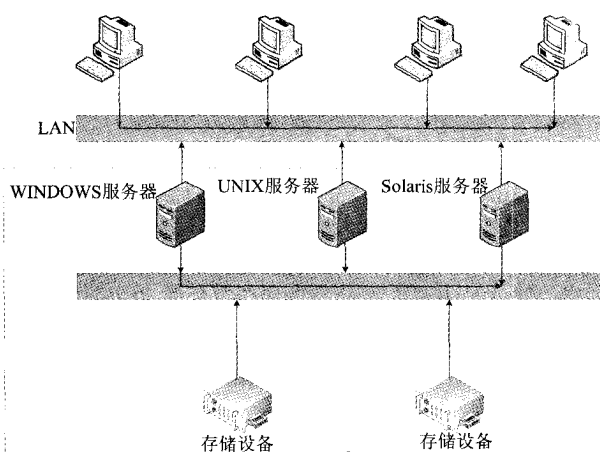


图4 SAN结构示意图

3.2 本地应用容灾技术

当灾难发生时,本地数据容灾技术虽然保证了数据的安全性,但是不可避免会造成系统应用的暂停,在这一点上已经不能满足人们对重要业务提供连续不间断服务的需求,因此需要本地应用容灾技术提高信息系统的高可用性,消除单点故障,当发生灾难时可以在用户不可察觉的最短时间迅速接管业务,确保服务连续性。本地应用容灾技术主要通过配置多机环境来达到高可用性,分为双机热备份技术和本地集群技术。

3.2.1 双机热备份

双机热备份技术是指在本地服务系统中,采用两台服务器进行冗余容错,并分为主从服务器。通常情况下,主服务器提供服务,而从服务器与主服务器保持一致的工作状态但是并不提供服务,当主服务器发生宕机,从服务器迅速向系统提供服务。在具体应用中,也可以选择双服务器互为备份同时向外输出服务。

3.2.2 本地集群

本地集群技术是由三个以上独立的计算机系统构成的一个相互协调的群组,作为一个整体对外提供服务。集群系统的其处理能力、I/O性能和可用性都有大幅提升。集群中每一台服务器称为一个节点,当其中某个节点发生故障时,集群会利用基于IP重定向的应用迁移技术将此节点业务转移到其他服务器执行,具体执行过程将在本文后续给出。

3.3 本地容灾产品介绍与方案分析

鼎联公司研发的Lander Cluster多节点集群一备多和多机互备模式容灾解决方案,是利用高可用集群软件对系统和数据进行重新规划,将数据库、文件系统存放在系统盘上。利用集群软件对存储资源进行控制,当遇到系统故障,则将故障点交给规定备援服务器工作,PTO恢复时间目标系统恢复时间在几十秒。

飞康公司CDP(Continue Date Protection)实时备份方案,利用飞康IPStor存储虚拟化平台和多时间点自动连续快照技术,实现I/O级别的备份粒度和分钟级恢复数据,而且极少占用主机和网络资源。

智网科技业务应急系统,利用差异快照技术,IP-SAN存储技术和内置虚拟化平台完成数据在线复制,应急系统自动化去主机信息实分钟级别现应急接管业务,是一种高性价比应用容灾手段,方案局限性在于RPO恢复点目标较大。

本地容灾可以有效解决单点故障问题,但是它的局限性在于不能避免大规模毁灭性的灾难,如地震、火灾、洪水战争等。单纯的本地容灾不能全面地保障业务连续,此时异地容灾技术成为全面容灾方案的主要手段,下面对异地容灾技术进行阐述。

4 异地容灾

异地容灾是指生产中心与备份中心在设计保持一定物理距离,通过远程容灾技术保障数据安全和业务连续性,在技术上分为两个层次:异地数据容灾和异地应用容灾。

4.1 异地数据容灾技术

异地数据容灾技术指将本地系统的数据在线备份到异地数据系统保存,当本地发生灾难时,异地数据可以用来完成数据重构,起到保护业务数据作用,但是不可避免的会发生业务暂停。异地数据容灾技术有虚拟存储技术和互联技术。

4.1.1 互联技术

IP存储技术作为主要互联技术^[6],通过IP网络完成类似SAN的块级数据传输,分为通过IP网络进行FC-SAN互联和直接构建以IP网络协议为传输协议的IP-SAN两种方式。与传统总线协议SCSI相比IP网络存储协议FCIP(Fibre Channel over IP)、IFCP(Internet Fibre Channel)和ISCSI(Internet SCSI)具有如下优点:主机与外设分离提高安全性、实现了资源灵活配置和使用、实现高性能I/O。FCIP和IFCP是利用网络将FC-SAN互联的协议,在IP网络和FC-SAN衔接处需要进行协议转换,增加系统复杂性。

而iSCSI定义了总线协议SCSI到IP上传输的映射,相当于在IP网络上传输SCSI协议,从而达到通过网络进行I/O操作的目的。如图5所示,iSCSI可以在IP网络上创建一个共享存储环境,利用发达的IP网络线路作为传输通道节约成本并克服了FC-SAN传输距离的限制,将I/O操作从局域网扩展到广域网。

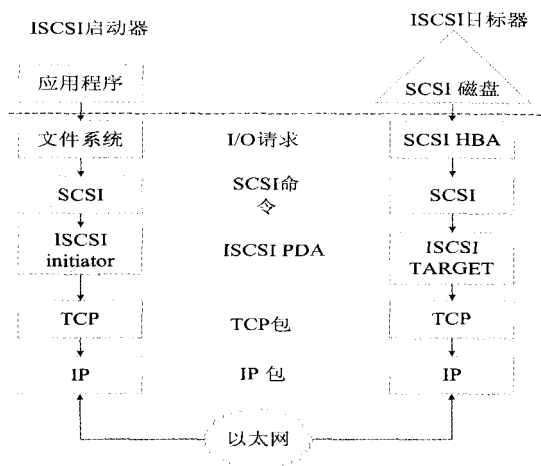


图5 iSCSI协议结构示意图

4.1.2 虚拟存储

存储体系从DAS发展到NAS、SAN甚至到IP存储,iSCSI协议将从总线连接扩展到网络连接,规模变得越来越大,结构变得越来越复杂,不得不面临存储子系统的异构共存问题,而虚拟存储技术正是高效管理大规模存储系统的有效手段。虚拟存储将不同存储介质和不同结构的存储子系统通过软硬件手段统一到用户存储应用的技术,可以实现存储管理自动化与智能化,从而提高存储效率、节约成本。

虚拟化^[7]实现需要在存储系统中加入一个新的虚拟化层,根据虚拟化层的位置可以划分为在服务器端、在存储设备端和存储网络上的虚拟化:

基于服务器的虚拟化将虚拟化层放在SAN中的服务器上,通过改造操作系统或者添加映射虚拟化层来实现。这种方法不需要特殊硬件,以软件模块的形式将虚拟化层嵌入到应用服务器操作系统中,不足之处在于受SAN软硬件环境制约,软件模块需要具备嵌入不同操作平台的能力。另外此虚拟化方法是通过多服务器分布式操作实现的,任何一个服务器的不当操作会影响数据资源的完整性。

基于存储设备虚拟化将虚拟化层放在存储设备的控制器中,这种虚拟方式充分考虑存储设备的物理特性,从虚拟化存储中解放应用服务器,达到很高的存储性能。但是这种方法局限于存储设备的同构环境,而且价格高昂,系统调试管理权限必须交给存储设备厂商。

存储网络上的虚拟化通过网络智能路由器、交换机,或者增加专用服务器等方案来实现,这种技术是目前虚拟化技术的主流。

1) 基于交换机的虚拟化将虚拟化层直接放置于交换机,通过改造或添加中间件的代码使交换机同时完成交换功能和虚拟化功能,为了提高交换机的性能普遍采用缓存技术和优化的缓存调度算法。

2) 基于存储路由器的虚拟化是随着SAN存储路由技术发展起来的,与存储网络中的交换机相似,具有路由功能和虚拟化功能,通过完成协议转换连接不同存储网络。这种方法不需要增加服务器软件模块,减少主机负载和复杂度,而且通过网络中多路由器还可以解决单点失效问题。不足之处在于路由器投资较大。

3) 基于专用服务器的虚拟化方式是在SAN中采用一台元数据服务器,只负责提供存储虚拟化。当应用服务器所需数据不在本地存储空间中,向专用虚拟存储服务器发出申请,专用服务器将所需数据相关信息发送给该应用服务器,且在此服务下次访问同样数据时无须申请便可直接存取,大大提高性能。

4.2 异地应用容灾

一套完备的容灾系统除了要提供异地数据级保护,还需要完成业务的异地转移提供应用级的容灾,这样才能保证业务高可用和连续性。应用容灾的目的是保证业务的无缝迁移,使用户察觉不到产生服务系统的位置变化。应用级的容灾是在数据级的基础上增加对整个业务的备份,相关技术包括:失效检测技术和服务迁移技术。

4.2.1 失效检测技术

失效检测表现一种未雨绸缪的主动性而不是灾难发生后“亡羊补牢”的被动性,目前主要灾难检测方法有心跳技术和检查点技术。心跳技术是系统各个重要节点每隔一段时间都要向外广播自身的状态,如果周期内节点广播消失则证明此节点失效。检查点技术为主动检测手段,是每隔一段时间会对系统重要节点进行检测,若周期内被检测节点没有响应,则认为检测节点失效。两种方法都存在周期问题,如果周期太频繁,会影响系统性能,占用系统资源;如果间隔时间过长,检测不灵敏,增加容灾难度和任务量。

4.2.2 系统迁移技术

系统迁移技术^[8]是保证灾难发生时实现系统透明的迁移,利用备用系统透明的代替生产系统,以保证业务连续性。系统迁移技术主要分为基于DNS的迁移、基于IP重定向迁移和基于集群的迁移。

基于DNS迁移技术利用动态域名解析系统完成业务迁移,实时性较差;

基于IP重定向迁移技术利用重定向设备实现容灾自动切换,智能性较高其,原理如图6所示。

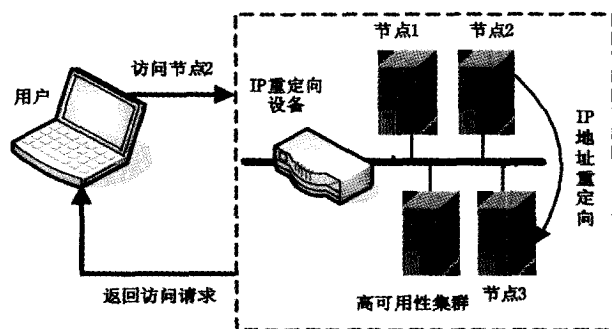


图6 IP重定向示意图

基于集群迁移技术指集群中任意节点出现故障,这个节点服务器的业务将由另一台服务器承担,保证业务连续性,且成本较低。

4.3 异地容灾产品与分方案析

4.3.1 异地数据级容灾产品与方案

VERTIAS利用卷管理器将数据以异步或是同步方式通过IP网络复制到远程系统,无须复杂硬件和专用线路,复制过程占用主机性能。

IBM利用ESS企业存储服务器和PPRC(Peer to Peer Remote Copy)点对点远程复制技术的数据容灾方案。PPRC以存储为基础,提供实时、同步与应用无关的数据容灾方案。通过专用ESCON通道完成本地ESS存储服务器到异地ESS的数据传输。

西北工业大学李战怀等人建立一个基于Linux平台的数据容灾系统——DDT(Data Disaster Tolerance System on Linux)。DDT系统通过Linux内核嵌入远程实时复制和灾难恢复机制,实现用户数据的动态复制和灾难恢复,并结合快照备份机制,实现很高的数据容灾能力。

四川李涛等人实现基于Internet的跨平台远程容灾系统,并提出一种基于Internet的大型文件镜像模型,具有同步异步切换功能,能够适应多种网络状况,具有良好容错能力和较强实时性。

这些模型和系统融合了SAN、NAS、镜像、RAID、集群等技术实现不同形式的复制过程。主要区别在于所使用的数据传输技术不同。

4.3.2 异地应用级容灾产品与方案

浪潮公司研发的HA Cluster集群应用容灾设计方案,通过任意热备服务器接管宕机服务器,集群热备服务器不是固定的,而是根据服务器设定的优先级和互斥性决定的。方案中在软件设计上用Heart Beat作为信息交换枢纽,实现全面监控和失效监测。在监测环节引入“浮动IP”概念,避免失效切换引起的目标服务器的切换。

数腾公司研发的CDAP持续数据应用保障设备可以提供实时容灾25台本地或异地服务器,产品基于底层卷完成实时数据备份,结合虚拟化技术和专业的迁移工具实现多点容灾和接管业务,具有一定的硬件结构一体化性和系统兼容性。

5 结束语

现有容灾备份手段存在运维负担大、技术要求高、兼容性差、应急演练难度大等问题,为解决信息系统面临越来越多的人为或自然不确定因素威胁,需要进一步对容灾备份技术深入研究。目前,随着前沿容灾技术如持续数据保护CDP、多点恢复等技术不断发展,将会构建更加安全、可靠的数据容灾系统。

参考文献:

- [1] 刘卫平.网络存储中的数据容错与容灾技术研究[D].西北工业大学, 2006.
- [2] 刘晓洁,赵奎,曾金全,等.容灾抗毁系统研究[C].全国网络与信息安全技术研讨会论文集(下册), 2007, 687.
- [3] 陈华英.磁盘阵列 RAID 可靠性分析[J].电子科技大学学报,2006,35(3):403-405.
- [4] Sacks D.Demystifying Storage Networking DAS, SAN, NAS, NAS Gateways, Fibre Channel, and iSCSI[J]. IBM Storage Networking, 2001: 3-11.
- [5] 袁晓铭,林安.几种主流快照技术的分析比较[J].微处理机, 2008,29(1):127-130.
- [6] 李琼,郭御风,刘光明,等.I/O 互联技术及体系结构的研究与发展[J].计算机工程,2006,32(12): 93-95.
- [7] 吴松,金海.存储虚拟化研究[J].小型微型计算机系统, 2003, 24(4): 728-732.
- [8] 张晓,张羽.远程数据库迁移技术研究[J].计算机应用研究, 2005, 22(4): 106-108.