Research about Theory and Key Technology of Disaster Recovery

Weijie Han, Hui Yan, Yu Wang, Xiaodan Gu
Department of Information Equipment, The Academy of Equipment Command & Technology, Beijing, China, 101416
Email: visc_hwj@126.com

Abstract: The disaster recovery system plays an important role on ensuring the safety of key information systems. The theory and key technologies of disaster recovery are introduced systemically. Firstly, the current foreign and domestic research status of disaster recovery is analyzed. Secondly, the evaluation standards of disaster recovery including the primary and complementary are introduced. Thirdly, the key technologies of disaster recovery are studied and analyzed systemically. Finally, the 3-D disaster recovery architecture model is worked over synthetically.

Keywords: disaster recovery; data backup; data duplication; data storage; failure detection; application transfer

容灾理论及关键技术研究

韩伟杰，阎　慧，王　宇，顾晓丹
装备指挥技术学院 信息装备系，北京，101416
Email: visc_hwj@126.com

摘　要：容灾系统可以保障重要信息系统的安全性。系统介绍了容灾的理论和关键技术。首先对国内外容灾研究现状进行了分析，并介绍了其定量评价指标，重点对容灾的关键实现技术进行了分析和研究，最后综合性地研究了容灾系统的三维体系结构模型。

关键词：容灾；数据备份；数据复制；数据存储；灾难检测；系统迁移

1 引言

美国“9.11”事件发生以后，国内外对容灾的认识普遍得到增强。广义上讲，任何提高系统的生存能力和可用性的努力都可称之为容灾。通常所说的容灾一般是指异地远程容灾。远程容灾是指为了防止自然灾害、战争或人为破坏等原因带来的区域性灾难而导致的系统瘫痪、数据丢失和业务中断，而在原生产系统之外的另一地点建立备份系统，备份系统具有与原生产系统相同或相似的主机、网络和存储设备[1]。容灾系统对于保障重要信息系统的安全十分重要。

2 国内外研究现状

2.1 国外研究现状

针对容灾的重要性，国外的研究机构首先研究制定出了容灾的相关标准和规范。目前，国际上通用的容灾系统标准为 Share78[2]，它是由美国 SHARE 用户组和 IBM 公司于 1992 年共同提出的灾难恢复等级标准。它把容灾的级别定义为 7 级，不同的级别对应不同的恢复时间和投资开销。

Share78 具体包括以下设计指标：

（1）备份/恢复的范围；
（2）灾难恢复计划的状态；
（3）业务中心与容灾中心之间的距离；
（4）业务中心与容灾中心之间如何相互连接；
（5）数据是怎样在两个中心之间传送的；
（6）允许有多少数据被丢失；
（7）怎样保证更新的数据在容灾中心被更新；
（8）容灾中心可以开始容灾进程的能力等。

按照以上的设计指标，容灾系统应主要包括：数据备份系统、备用数据处理系统、备用网络系统、备用用户接入系统、备用基础设施、技术支持、运行维护支持、应急响应计划等 8 个容灾要素。

资助信息：本课题得到部委级基金项目支持。

国外研究机构针对容灾需求，研究开发出了一些具有代表性的容灾产品。

HP[3]开发了 OpenVMS 高可用集群系统，该系统可以提供高可用的、可扩展的、灵活的计算环境，可以支持 98 个节点，覆盖范围可以达到 500 公里，能够提供不间断的服务，可以容忍火灾、地震等灾难。

VERITAS[4]提供了 Volume Replicator，可以将数据以异步或者同步的方式复制到远程系统，通过 IP 网络进行传输，不需要复杂的硬件和构建专用传输线路，复制过程基于主机，使用 Flash snap 技术检查远程数据，与距离无关，独立于磁盘阵列，与 VCS / GCM 集成，构成完整的灾难恢复解决方案。

IBM[5]开发出了基于 ESS 企业存储服务器的 PPRC (Peer to Peer Remote Copy)复制技术的数据容灾方案，以及基于 IBM RS / 6000 服务器的 HAGEO(High Availability Geographic Cluster)异地集群技术的应用级容灾方案。

EMC[6]公司的远程数据备份软件 SRDF（Symmetrix Remote Data Facility）是一个在线的且独立于主机的数据镜像存储解决方案，可在多种操作系统下使用，它能够同时为大型机、UNIX、Windows 和 AS/400 系统提供完整的业务连续可用性能力。

2.2 国内研究现状

在 9.11 事件和印度洋海啸之后，我国充分认识到了重要信息系统容灾的必要性。为此，国务院信息化办公室[7]于 2005 年发布了《重要信息系统灾难恢复指南》，并于 2007 年发布了《信息系统灾难恢复规范》，用于指导信息系统的灾难恢复规划工作。《规范》将灾难恢复分为 6 个等级，并认为容灾系统应包括：数据备份系统、备用数据处理系统、备用网络系统、备用基础设施、技术支持、运行维护支持和灾难恢复预案等 7 个要素。在这 7 个要素中，前三个属于 IT 技术的范畴，后四个属于管理和服务的范畴。

按照规范要求，国内的研究机构也开发出了一系列的产品和解决方案[8]。

华为提供的 OceanStor 企业级数据容灾解决方案，通过将企业关键在线数据复制到异地的容灾中心，并在整个过程中保证数据的实时性和正确性，且在整个复制过程中将对生产业务系统的影响降到最低，由此保证企业业务系统在不可预料的灾难发生时，仍能保证业务数据的完整和业务系统的连续。

浪潮根据企业容灾需求，分别提出了基于主机复制技术的容灾方案、基于磁盘阵列复制技术的容灾方案、基于智能 SAN 虚拟存储设备复制技术的容灾方案和基于虚拟带库系统复制技术的容灾方案。

神州数码提出了不同阶段的容灾解决方案，并设计了一个完整的容灾体系。该容灾体系包括四个阶段：本地数据安全保护、本地应用的高可用性、异地数据安全保护、异地应用的连续性。这四个阶段是容灾系统建设的一个渐进的过程，用户可以根据自己的实际情况进行选择，分步建设，最终建成一个完善的容灾系统。

此外，四川大学计算机网络与安全研究所[9]研制的“及时雨灾难救援中心”，是基于 Internet 的远程数据备份方案，具有完全的自主知识产权。它集数据实时镜像、服务自动切换和数据安全传输于一身，可广泛应用于金融、证券、电信、政府、教育和企业等行业和部门。

3 容灾评价指标

从技术上看，衡量容灾系统主要有两个指标：RPO 和 RTO。RPO（Recovery Point Objective）：即数据恢复点目标，是指灾难发生时刻与最近一次数据备份时刻的时间间隔，即尚来不及对数据进行备份（导致数据丢失）的时间，代表了数据的丢失量；RTO（Recovery Time Objective）：即恢复时间目标，主要指的是系统所能容忍的业务停止服务的最长时间，也就是从灾难发生到业务系统恢复服务功能所需要的最短时间周期。RPO 针对的是数据丢失，而 RTO 针对的是服务丢失，二者没有必然的关联性。RTO 和 RPO 的确定必须在进行风险分析和业务影响分析后根据不同的业务需求确定。

此外，NRO（Network Recovery Object，网络恢复目标）和 DOO（Degrade Operation Object，降级运作目标）对容灾系统也是至关重要的性能指标。NRO 代表灾难发生后，网络切换需要的时间。DOO 代表的是恢复完成以后到第二次故障或灾难的所有保护恢复以前的时间间隔，反映了系统发生故障后降级运行的能力。

4 容灾关键技术

容灾可以分为数据容灾和应用容灾两种类型。数据容灾是指建立一个备用的数据系统，该备用系统对生产系统的关键数据进行备份。采用的主要技术是数据备份、数据复制和数据存储技术。应用容灾则是在

978-1-935068-18-1 © 2010 SciRes.

数据容灾之上，建立一套与生产系统相当的备份应用系统。在灾难发生后，将应用迅速切换到备用系统，备份系统承担生产系统的业务运行任务。主要的技术包括负载均衡、集群技术。数据容灾是应用容灾的基础，没有数据的一致性，就没有应用的连续性，应用容灾也无法保证。

构建一个容灾系统主要使用的技术包括数据备份、数据复制、数据存储、灾难检测和系统迁移等。

4.1 数据备份

数据备份就是把数据从生产系统备份到备份系统中的介质中的过程。常用的备份技术方法主要有以下几种：

（1）主机备份

主机负责将数据备份到和主机直接相连的存储介质上（一般是磁带）。这种技术仅能适应于单台服务器备份，并且在灾难恢复过程中，系统恢复的时间较长。

（2）网络备份

系统中备份数据的传输以网络为基础。根据备份系统中备份服务器和介质服务器是否在同一个 LAN 中，可以将网络备份分为基于局域网的备份和远程网络备份。

（3）专有存储网络备份

存储系统独立于备份系统，备份过程可以在存储局域网中实现。根据备份过程中对应用服务器的影响，专有存储网络备份可以分为 LAN-Free 备份和 Server-Free 备份。

4.2 数据复制

数据复制技术是容灾系统的核心。数据复制技术是通过不断将生产系统的数据复制到另外一个不同的备份系统中，以保证在灾难发生时，生产系统的数据丢失量最少。

按照备份系统中数据是否与生产系统同步，数据复制可以分成同步数据复制和异步数据复制。同步数据复制就是将本地生产系统的数据以完全同步的方式复制到备份系统中。由于发生在生产系统的每一次 I/O 操作都需要等待远程复制完成才能返回，这种复制方式虽然可能做得数据的零丢失，但是对系统的性能有很大的影响。异步数据复制则是将本地生产系统中的数据在后台异步地复制到备份系统中。这种复制方式会有少量的数据丢失，但是对生产系统的性能影响较小。

4.3 数据存储

目前，比较重要的存储技术有直接附加存储（DAS）、网络附加存储（NAS）、存储区域网络（SAN）以及 IP 存储网络等。

DAS，即存储设备通过光纤或铜线之类的连接介质直接与服务器相连，I/O 请求直接访问设备。DAS 的主要特征在于对直接附加的服务器提供快速数据存取。

NAS 是一种能够提供灵活的、可伸缩的解决方案的存储类型，能够满足文件共享需求。NAS 设备是一种运行专门设计用于处理文件服务的操作系统的服务器。网络附加存储的主要特点是可以通过 TCP/IP 等 LAN 协议从局域网上直接访问存储设备。

SAN 是一种专用网络，能够提供高性能与高度可用的存储子系统。SAN 由专门的设备组成，例如主机服务器中的主机总线适配器、帮助路由存储流量的交换机、磁盘存储子系统与磁带库。上述设备通过光纤或铜线相互连接。SAN 的一个主要特点是存储系统通常可提供多台主机同时使用，从而能够提供可扩展性与灵活性。

基于 IP 的存储是指在 IP 网络中实现类似于 SAN 光纤通道的“块级”数据处理，它将 SCSI 协议映射到 TCP/IP 协议上，使得 SCSI 的命令、数据和状态可以在传统的 IP 网上传输。IP 存储网络技术主要有三种：FCIP（Fiber Channel over IP）技术、iFCP（Internet Fiber Channel Protocol）技术和 iSCSI（Internet SCSI）技术。

4.4 灾难检测

对于可能遇到的各种灾难，容灾系统需要能够自动地检测灾难的发生。目前，灾难检测一般都是基于超时的，通过判断被检测进程发出的消息数据是否能在一定的时间内到达，来分析被检测进程是否失效。灾难检测主要有两种方法：Pull 模型和 Push 模型。

基于 Pull 模型的灾难检测方法是，检测进程主动向被检测进程发送“Are you alive？”的消息，被检测进程收到这样的询问消息后，就返回一个应答消息，报告自己是存活的。检测进程在发出询问消息后会启动一个定时器，设定一个超时值。如果检测进程在规定的时间内没有收到被检测进程的应答消息，就认为被检测进程失效。

基于 Push 模型的灾难检测方法是，被检测进程持续地按一定时间间隔主动向检测进程发送类似于“I am alive”的存活消息，又称心跳消息。检测进程会启动一个定时器，设定一个超时值。如果检测进程在定时器超时前没有接收到被检测进程发送的心跳消息，就认为失效。

4.5 系统迁移

在发生灾难时，为了能够保证业务的连续性，必须能够实现系统的透明迁移，也就是能够利用灾备系统透明地代替生产系统，主要的系统迁移方法包括：基于 DNS 的迁移技术、基于 IP 重定向的迁移技术以及基于集群的迁移技术。

基于 DNS 的迁移技术是通过动态域名解析系统将发生故障的生产中心的业务由容灾中心的业务应用系统接管，该方式存在延迟，实时性较差。

基于 IP 重定向的迁移技术是一种重要的实现应用容灾的迁移技术，它主要使用 IP 重定向设备，使用户的连接在生产中心和容灾中心之间自动切换，以实现容灾抗毁与业务连续性。

基于集群的迁移技术主要利用集群的特性，当集群中的一台节点服务器发生故障时，这台服务器上所运行的应用程序将在集群中的另一台服务器上被自动接管。如果发生的灾难影响到了容灾系统的生产中心，集群技术会自动将业务系统转移到集群中的其它部分，关键业务不会受到影响。

5 容灾系统体系结构模型

容灾系统主要利用先进的软、硬件设备和环境，以容灾恢复等级（量化指标）为中心，综合运用各种技术手段（灾难备份与恢复技术）和管理措施（灾难恢复计划与措施），实现系统的灾难恢复要求，其各要素之间的关系可用如图 1 所示的三维模型来表示。

一个容灾系统的体系结构（Disaster Recovery Architecture，简称 DRA）可定义为一个三元组：DRA=（O，T，P），其中：

O={<O_1,O_2,O_3,O_4>| O_1,O_2,O_3,O_4 分别表示不同类型指标 RPO、RTO、NRO、DOO 的值约束}，即 O 表示要达到的各种灾难恢复指标；

T={T_i}表示为达到各种灾难恢复指标，必须采用的相关灾难备份与恢复技术 T_i；

P={P_j}表示为保证达到该系统的各种灾难恢复指标，必须实施的各种灾难恢复计划和措施 P_j。

在灾难恢复体系结构的三维模型中，可根据三元组定义若干的规则点，由若干规则点构成的集合则形成各级的灾难恢复解决方案，为整个体系结构的各个层次、各个组件的协调工作起到核心控制作用。

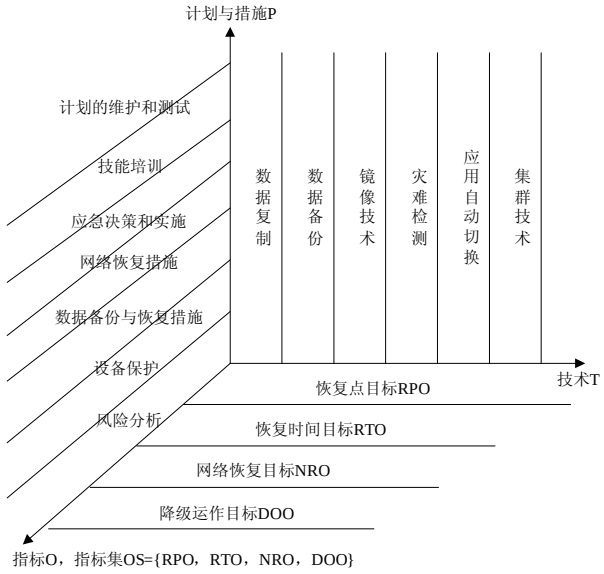

Figure 1. 3-D model of disaster recovery architecture

图 1. 容灾体系结构三维模型

6 总结

容灾系统是保障重要信息系统安全的重要屏障。一个完整的容灾系统应采用必需的灾备技术，遵照相关的标准来实施。随着信息系统结构的日益复杂和海量数据存储的出现，如何有效将生产中心的应用数据及时备份到容灾中心，并实现复杂业务的实时迁移，是未来容灾技术研究的重点。

References (参考文献)

[1] Wang Dejun, Wang Lina. Research of Disaster Tolerance System. Computer Engineering, 2005, 31(6):43-45.
[2] National Institute of Standards and Technology. SP 800-34: Contingency Planning Guide for Information Technology Systems.
[3] http://www.hp.com.
[4] http://www.veritas.com.
[5] http://www.ibm.com.
[6] ttp://www.emc.com.
[7] The state council informatization office. Disaster recovery specifications for information systems. 2007
国务院信息化办公室. 信息系统灾难恢复规范. 2007.
[8] Yang Xiaohong, Li Jian, Yang Weiguo. Researching and analyzing of information system disaster tolerance technology. Computer Engineering and Desigh, 2005, 26(10):2727-2729.
[9] Yang P, Li T, Zhao K, et al. An internet based disaster recovery system. Disaster Recovery Journal, 2005, 18(1).