

基于虚拟化的数据备份机制^①

连淑娟

(北京外国语大学 信息技术中心, 北京 100089)

摘要: 随着信息化建设的发展, 数据备份的重要性越发显得重要, 而传统的备份方式消耗资源多, 且不易管理; 而虚拟化技术的出现打破了这一传统的数据备份方式. 阐述了在虚拟化平台下几种新的数据备份机制, 并分析了每种数据备份机制的优缺点, 以便于读者在建设备份系统时可以根据实际情况, 选择合适的备份方案. 最后, 本文总结了虚拟化备份机制存在的问题, 并探讨了未来的发展趋势.

关键词: 虚拟化; 备份; 容灾; VMware; 存储

Data Backup Mechanism Based on Virtualization

LIAN Shu-Juan

(Information Technology Center, Beijing Foreign Studies University, Beijing 100089, China)

Abstract: With the development of information technology, the importance of data backup becomes more important. But the traditional backup methods consume more resources and become managed difficultly. The emergence of virtualization technology breaks the traditional data backup methods. This paper describes several data backup mechanisms under the virtualization platform, and analyzes the advantage and disadvantages of each mechanism, so that the reader can choose the right backup solution according to the actual situation. Finally, the paper summarizes the virtualization backup mechanism problems and discusses the future trends.

Key words: virtualization; backup; disaster recovery; VMware; storage

随着信息化应用技术的逐步成熟, 围绕着信息化产生的各种系统也越来越广泛地应用于日常工作当中, 这些系统的海量数据是整个系统运行的核心^[1]. 而对于企业的信息化建设来说, 仅对这些系统数据进行简单的备份是远远不够的, 还必须做到当业务系统数据遭到破坏后, 能尽快在短时间内恢复关键业务数据展开正常的工作才是重中之重. 目前广泛应用的虚拟化技术打破了传统的备份机制壁垒, 将物理资源转化为逻辑资源, 从而将数据容灾备份系统的建设带入了一个崭新的时代.

在虚拟化备份方面, Javaraiah 等介绍了一种基于云计算的在线备份机制; GuoDong 等设计和实现了基于 DHFS 的云数据备份系统; DuHongtao 提出一种镜像同步组概念, 也应用于云计算领域; ChengHao 等人提出了利用服务器空余资源参与虚拟备份的策略, 利于

节省资源.

本文综述了数据备份机制, 根据备份软件所处的层次分成了三大类, 并分析了其优缺点, 最后探讨了虚拟化备份技术的发展方向.

1 传统的备份机制

备份, 简言之, 就是为了防止计算机数据及应用等因计算机故障而造成的丢失及损坏, 从而在原机中独立出来单独贮存的程序或文件副本就称之为备份^[2].

传统的备份机制有: ①磁带备份, 即将数据实时传送到远程备份中心来制作完整的备份磁带或光盘. ②数据库备份, 即在与数据库主机分离的备份机上建立数据库的一个拷贝. ③网络备份, 即对系统的数据库数据和重要目标文件进行监控, 并将更新日志实时通过网络传送到备份系统, 备份系统根据日志对磁盘

① 收稿时间:2014-12-09;收到修改稿时间:2015-01-19

进行更新。④镜像备份,即通过光纤通道和磁盘控制技术将镜像备份到其他地方,镜像磁盘数据和主磁盘数据完全一致,更新方式为同步或异步^[3]。

以上几种备份方式属于冷备份,只适应于对数据可用性要求不高的情况下,而传统的热备机制克服了这一缺点,对于任何系统宕机或服务中断的情况,都能进行错误判定、故障隔离并快速启动备机来恢复被中断的服务,主要有基于共享存储和基于数据复制两种热备方式,具有高可用性,在国内使用较多,但是热备技术只限定了两台服务器。

不管是冷备还是热备,虽然在系统出现故障时能够进行系统恢复,但是这些机制只能处理计算机单点故障,对区域性、毁灭性灾难则束手无策。具体来说:①磁带备份连续写入需要更高的网络处理能力及更多的时间,而这往往不可行。②制约了服务器虚拟化技术的实施和发展。③耗费时间“堆砌”在一起的大量单点产品导致管理备份活动极其困难,数据恢复既缓慢又不细化,效率低下且缺乏确定性。④难以应对数据量激增给 IT 环境带来的超负荷处理挑战,海量数据导致备份时间长,恢复极其复杂。⑤不能全局性地解决冗余数据激增的问题,这一问题会导致对网络、存储和管理资源的过度消耗^[4]。

2 虚拟化平台下的备份机制

虚拟化技术的出现打破了传统数据备份机制的壁垒,克服了上述的缺点。虚拟化技术使得一个物理机器上可以运行多个独立的操作系统,而这些操作系统可以共享物理机器上的 CPU、内存、磁盘等资源,打破了软硬件的紧耦合性,带来了分区、隔离、封装等好处。正是由于这些好处,所以人们对虚拟化平台下的备份机制提出了更高、更可靠的要求。

市场上虚拟化平台管理软件主要有:VMware vSphere,微软的 Hyper-V 以及思杰公司的 Xen。其中,VMware 以 70% 的市场占有率领航整个虚拟化市场。本文也主要针对 VMware 虚拟环境进行讨论。针对 VMware 虚拟平台下备份机制的研究很多,归纳起来主要可以分为四大类:①虚拟机层次备份;②主机层次备份;③存储层次备份;④云备份。

2.1 虚拟机层次备份

这种备份机制是将虚拟机作为普通物理机进行处理,按照备份物理操作系统的方式来备份虚拟机上的

数据。方法是:在每个虚拟机内部均安装一个备份代理软件,并通过网络将数据备份到其他的备份服务器中,可以采用手动或者使用脚本进行复制^[7]。

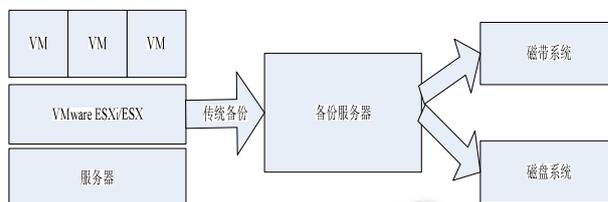


图 1 虚拟机层次备份

这种备份机制的优点是:①可以使用合适的软件来备份数据中心内的所有服务器(无论是虚拟机还是物理服务器),保证数据的可用性及一致性。这一点尤其表现在数据库的备份上,当数据损坏时,可任意时间点回退解决逻辑错误^[5]。②可以实现文件级的备份和还原,无论是数据库或系统文件,都可以轻松地执行增量备份,有效地避免了重复数据的产生。

这种备份机制的缺点也是显而易见的:①在每个虚拟机上都安装备份软件,需要占用一定的资源,且需要多套授权,价格昂贵。②无法对虚拟机被封装在文件中的这种特性加以利用。当将虚拟机中的个别文件进行备份时,无法将整个虚拟机进行备份和还原。③对物理主机来说,5%的备份负载是可以接受的,若备份多个虚拟机,则负载比例会成倍增加,物理主机不容许这么大的备份负载^[6]。

2.2 主机层次备份

这种备份机制是将主机内的每台虚拟机看作打包好的一个整体(包括各个文件,如:vmrk, .vmx, .vmxf, .vmsd, .log 等)进行备份,很好地利用了虚拟机被封装成文件的这种特性。在 VMware 环境下,最常用的是利用自带备份工具 VDR (VMware Data Recovery) 和 VDP (VMware vSphere Data Protection) 进行备份。

2.2.1 VDR 备份

VDR 是建立在 VMware vStorage API for Data Protection 的基础上,它与 VMware vCenter Server 集成在一起,使我们可以集中调度备份作业。VDR 在创建虚拟机备份的同时,不会中断虚拟机的使用或其提供的数据和服务。VDR 会管理已有的数据备份,并在这些备份过期后将它们删除,支持去重复功能来删除冗

余数据^[7]。

VDR 可以在任何接受 VMware ESX/ESXi 支持的虚拟磁盘上存储备份, 可以使用存储区域网络(SAN)、网络附加存储(NAS)等设备。

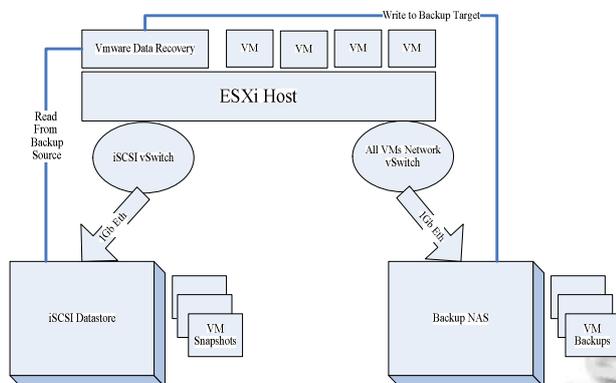


图2 VDR的体系结构

VDR 的备份和还原时凌驾于虚拟机之上的, 因此无论虚拟机运行着什么操作系统, VDR 都将其看作一个整体来备份, 不需要在虚拟机内部安装任何代理软件, 虚拟机对于备份和还原操作是不可感知的。

VDR 备份机制与 2.1 中的备份机制相比, 有如下几个优点: ①VDR 与 VMware vCenter Server 集成, 可以集中、高效地管理虚拟机的备份, 并简化备份作业的配置与管理, 提高了工作效率。②VDR 支持重复数据消除功能, 可以有效地节省用于备份的磁盘存储空间。③VDR 基于快照式备份, 显著降低备份虚拟机的成本, 并将备份时间窗口中的时间减至最低^[8]。④可以快速进行系统还原, 只要选中 VDR 中的还原点, VDR 将快速检索特定的数据库, 当覆盖原有的虚拟机时, VDR 只高效传输已发生改变的数据块, 缩短故障恢复时间。

VDR 同时也存在着以下劣势: ①不能进行单个文件恢复, 由于 VDR 还原时要求还原整个虚拟机, 即使只恢复单个文件也是如此。②数据的一致性差, 由于 VDR 无法实现实时备份, 所以当数据更新后还未进行备份时, 此时若出现虚拟机故障, 还原时将造成数据的不一致性。

2.2.2 VDP 备份

VDP 是一款基于磁盘的备份和恢复解决方案, 可靠且易于部署。VDP 和 VDR 一样, 也可以与 VMware vCenter Server 完全集成, 用来对备份作业进行高效的

集中式管理, 同时将备份存储在经过重复数据消除的目标存储位置中^[9]。

VDP 可以看作 VDR 的升级版, 它具有 3 个版本, 分别是: 0.5TB 级别的 OVA 版本; 1TB 级别的 OVA 版本; 2TB 级别的 OVA 版本。不同的版本仅代表它的 Deduplication Storage 的空间大小, 用户可以根据自己的实际情况来选择版本。

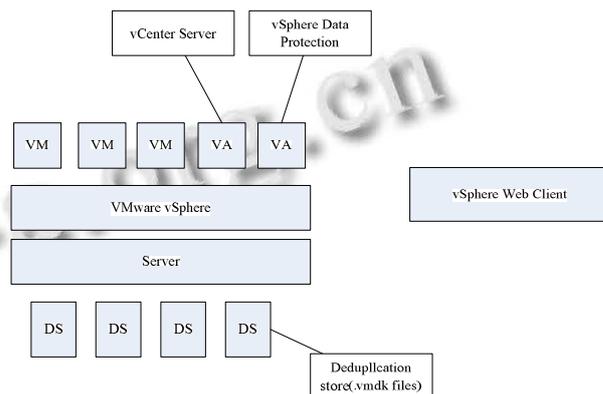


图3 VDP 体系结构

VDP 除了具有 VDR 的优点, 还具有其他的一些优势: ①可作为集成组件简单直接地安装到 vSphere 中, 可通过 Web 门户进行管理。②使用检查点和回滚机制保护备份。③从基于 Web 的界面中, 通过终端用户启动的文件级恢复(FLR)提供 windows 和 linux 文件的简化恢复。④对所有的备份采用获得专利的可变长度重复数据消除技术, 从而极大地减少备份所占用的磁盘空间。⑤通过使用更改数据块跟踪(CBT)和虚拟机快照, 最大限度地缩短了备份窗口。

VDP 存在诸多优点的同时, 也存在以下几个缺点: ①VDP 的文件级恢复有很多限制, 如目前还不支持 ext4、FAT16、FAT32、扩展分区、加密分区、压缩分区等; 在同一个恢复操作中不能恢复超过 5000 个文件夹或文件。②VDP 仅有三种配置(0.5TB、1TB、2TB), 而且一旦部署后将无法更改大小。③VDP 对于使用独立磁盘的虚拟机, 可以进行备份, 但是不能恢复, 因为独立磁盘不受快照影响。

目前, VMware 在 VDP 的基础上又提出了 VDP(A)(VDP Advanced), 它克服了 VDP 的一些缺点, 支持 2T、4T、6T、8T 四种级别, 而且 VDP(A) 在部署后还可以增加大小。当然, VDP(A) 对系统服务器的要求也会更高。

以上几种备份机制都是基于 VMware 平台的, 在市场上 Hyper-V 平台下主要有 Snapshot、Windows Server Backup & VSS、Data Protection Manager 等三种备份方式, 而在 Xen 平台下, 备份机制要远远落后于 VMware, 在此不再赘述。

2.3 存储层次备份

这种备份机制是以存储为基础、实时的、与应用无关的数据远程镜像功能, 如 NetApp SnapMirror、IBM Enhanced Remote Mirroring、Metro Mirror 等, 这种备份机制能达到最佳的性能, 并能通过 VMware Site Recovery Manager 完成异地的容灾备份架构^[10]。

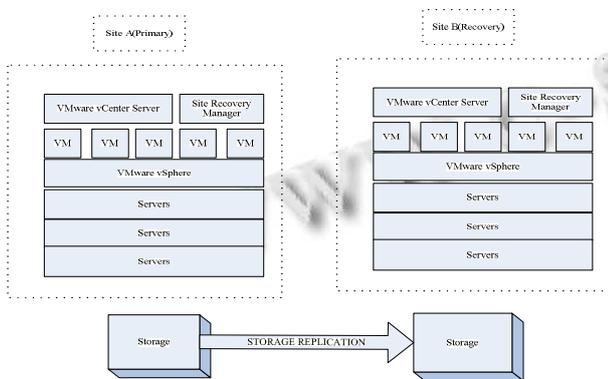


图 4 存储复制体系结构

这种备份机制一般分为两种模式: 同步复制模式和异步复制模式。

同步复制模式的优点是可保持数据的一致性, 因为主站点要接收到远程站点的 I/O 回应后才算完成整个 I/O 作业, 因此这种模式不会产生数据丢失。

同步复制模式的缺点是备份效率低下, 因为主站点要等待远程站点的回应, 所以会消耗很多等待时间, 同时也受网速及距离的限制。

异步复制模式的优点是备份效率高, 因为主站点完成 I/O 后即可完成整个作业, 无需等待远程站点的回应, 所以也不受网速及距离的限制。

异步复制模式的缺点是不保证远程站点与主站点有相同的 I/O 顺序, 所以有可能会造成数据的丢失, 数据一致性稍差^[11]。

2.4 云备份

在虚拟化备份领域, 云备份依然是个新课题. 这种备份机制是通过集群应用或分布式文件系统等功能, 将网络中大量不同类型的存储设备通过应用软件集合起来协同工作, 共同对外提供数据存储备份和业务访

问的服务^[12]. 这里的云备份指的是私有云备份, 这种方案可以概括为 CIA(Confidentially, Integrity and Availability): 安全性、完整性和可用性。

这种备份方案的优点是: ①可扩展节约成本, 可以依靠第三方云提供商的无限扩展能力, 使得我们可以预测并管理容量增长和费用. ②高效可靠, 云提供商会提供先进的技术, 如基于磁盘的备份、压缩、加密、冗余数据删除、服务器虚拟化、数据优化等, 还能提供监控、管理和报表服务等. ③恢复时间短, 虚拟机出现故障后, 可以通过网络传输从云端进行数据恢复, 节省时间. ④高可用性, 这种离站式备份能通过任何网络连接来访问, 在发生区域性灾难时增加了保险系数。

这种备份机制的缺点是: ①首次完全备份和完全恢复数据可能会过于消耗时间, 对生产系统的影响过大. ②根据现有的带宽资源, 企业每天都会有每天能上传到云端的数据量的合理阈值, 这一限制会对备份策略产生影响。

目前主流的几种虚拟备份工具可以把现有数据备份到云端, 如 Veeam Backup and Replication、PHD Virtual VMware Backup、Zmanda for ESXi、Acronis vmProtect 等。

3 备份机制优劣势对比总结

本文根据虚拟机备份程序所在的位置, 把备份机制分为了四种类型, 每种备份机制都有其优缺点, 简单总结成表 1。

表 1 虚拟机备份机制优劣对比

备份机制	优势	劣势
虚拟机层次备份	简单易用; 可文件级还原	价格昂贵; 备份主机负载大
主机层次备份	备份和还原效率高; 占用磁盘空间小	不能实时备份, 数据一致性差
存储层次备份	实时备份; 性能最佳	占用磁盘空间大
云备份	高可用性; 扩展性好	大小和带宽限制

在虚拟化环境下进行备份和物理环境中还是有所不同, 可能会遇到很多意想不到的问题, 有些问题需要注意并避免一些常见错误: ①确保虚拟机处于正确的静默状态. 在备份虚拟机时, 需要事先确认虚拟机已经静默完成, 以确保备份是完整连续的, 否则, 备份在恢复时可能会出现错误. ②VDR 和 VDP 虽然都是基于快照进行备份的, 但是虚拟机快照并不是备份,

这一点务必要搞清楚。虚拟机快照保留了虚拟机做快照时的状态,一旦被恢复到了其中一个还原点,就无法恢复到现在的状态。③谨慎地规划好备份周期。虚拟机本质上是共享主机的,而主机一般又共享存储设备,因此备份时需要考虑对资源的消耗情况。建议用户在备份时能减少并行虚拟机在主机和共享存储上运行的数量。同样,如果在主机上同时有太多的虚拟机备份并行运行,将会造成主机上的性能瓶颈^[13]。

4 结语

数据对于企业的巨大价值已经毋庸置疑,数据保护作为企业价值实现的重要一环已经被所有的企业所重视,这就将数据备份提到了一个更加重要的高度。结合虚拟化环境下备份机制的发展情况,我们可以预测未来备份机制的发展趋势:①持续数据保护。通过持续数据保护才能保证虚拟机在故障发生时,系统状态可以恢复到以前的任意时间点,最大程度地保证了数据的一致性^[14]。②继续改进重复数据删除技术,以减少需要在网络上移动和被存储的数据量,进一步节省磁盘空间。③虚拟机更快速高效备份。在大型的数据中心中,虚拟机可以在主机间迁移而且对应的数据可以随之迁移,这个虚拟机的备份带来了更大的挑战。如何在不影响服务器性能的情况下对虚拟机进行快速高效的备份也是将来的研究重点。

参考文献

- 1 李悦,梁祺.基于虚拟化技术的高校信息服务系统容灾备份的研究.信息技术,2013,7: 30.
- 2 向永谦,胡桥.基于虚拟化的服务备份技术,2012,08: 176-177.
- 3 <http://baike.baidu.com/view/600259.htm>.
- 4 <http://www.ciotimes.com/infrastructure/sjk/64571.html>.
- 5 <http://it.chinabyte.com/25/12624525.shtml>.
- 6 <http://www.docin.com/p-12643537.html>.
- 7 VMware Data Recovery 管理指南白皮书. <http://www.vmware.com>.
- 8 <http://www.flow-ever.com/html/DownViews.asp?ID=21&SortID=8>.
- 9 VMware Data Protection 管理指南白皮书. <http://www.vmware.com>.
- 10 <http://it.rising.com.cn/safe/2012-06-07/11664.html>.
- 11 <http://blog.csdn.net/jackxinxu2100/article/details/6701293>.
- 12 褚福勇,肖依,蔡志评,陈彬.虚拟机备份机制研究.计算机工程与科学,2009,9(31):131-134.
- 13 <http://it.rising.com.cn/safe/2012-06-07/11664.html>.
- 14 http://baike.baidu.com/link?url=Ylh-f1dNrk9defiIybpIgnX4tF9Yo-i3zLYhjk2kzIDVuZYrSc_nABfWSi8-DeFjT9Pin2SqYP0KjXTQI4_wa.