



云备份解决方案

司 伟,王 峰

(中国电信股份有限公司北京研究院 北京 100035)

摘 要:云备份作为一种新兴的数据备份方式,为用户提供了把本地数据远程备份到云端的服务,正日益受到业界的关注。结合云环境下的数据备份需求,从云备份解决方案的“端—管—云”3个方面对云备份客户端、云备份网络通道和云备份资源池的设计思路进行了全面、深入的分析,并针对不同的业务场景和用户需求,提出了公有云备份和混合云备份的切实可行的实施方案,为相关领域的研发和部署提供了重要参考。

关键词:公有云备份;混合云备份;解决方案

doi: 10.11959/j.issn.1000-0801.2015170

Cloud Backup Solution

Si Wei, Wang Feng

(Beijing Research Institute of China Telecom Co., Ltd., Beijing 100035, China)

Abstract: As a new data backup form, cloud backup service provides users with putting the local data backup to the remote cloud. The cloud backup is being more and more focused by the industry. Combined with the demand of data backup under cloud computing environment, three aspects of “end-pipe-cloud” cloud backup solution were comprehensively analyzed, including cloud backup client, cloud backup network channel and cloud backup resource pool. In addition, according to the different business scenarios and user demand, feasible public cloud and hybrid cloud backup solutions were proposed, which could provide an important reference for the research and deployment of related field.

Key words: public cloud backup, hybrid cloud backup, solution

1 引言

随着云计算技术的日趋成熟,越来越多的企业和个人把业务应用托管在数据中心,企业对业务连续运行的要求也不断提高,甚至越来越多的系统要求“零数据”丢失。为了提高数据的安全性,作为数据保护的最后一道防线——备份的重要性不言而喻。

云备份是指用户利用云服务提供商提供的数据备份

资源服务实现数据的备份和恢复。作为一种新兴的数据备份方式,云备份为用户提供了把本地数据远程备份到云端的服务,具有很多优点,例如,它能够依托于云服务提供商的“无限”扩展能力为用户提供按需调用的数据备份资源、利用多点分布的云资源池满足用户异地容灾高可用的需求、通过数据中心内部的高速网络传输数据提供高性能的传输等,这使得云备份拥有广阔的发展前景。

本文针对云备份系统的设计与实现,首先从“端—

收稿日期:2015-02-09;修回日期:2015-07-09

论文引用格式:司伟,王峰.云备份解决方案.电信科学,2015170

Si W, Wang F. Cloud backup solution. Telecommunications Science, 2015170



管—云”3个方面阐述了云备份解决方案的总体架构,然后重点分析了云备份系统实现中的关键技术和设计思路,最后分别针对公有云备份和混合云备份提出了切实可行的实施方案。

2 云备份系统架构

云备份不同于传统的数据备份方案,它依托于云服务实现用户数据的备份,其使用的云资源池能够按需满足用户的备份资源需求,简化用户的使用难度。云环境中的备份有其特殊性:首先,备份数据量大,备份频率高;其次,服务器虚拟化对存储和网络提出了更高的要求,共享存储必须能为数十个,甚至数百个虚拟机提供数据交换的能力;最后,存储和物理服务器之间的网络连接也需要支持由无数独立的虚拟机产生的数据通信。

典型云备份系统的架构如图1所示。

典型的云备份系统主要包括“端—管—云”3个部分,即由云备份客户端、云备份网络通道和云备份资源池组成。

- 云备份客户端:云备份客户端具有多种形态,例如传统解决方案在每台备份设备上部署的代理、基于虚拟化 Hypervisor 部署的无代理备份组件以及在用户侧部署的专用代理软硬件设备等,不同的客户端形态会直接影响云备份方案部署和使用的复

杂度。

- 云备份网络通道:根据不同的业务场景,云备份网络通道可以分为数据中心内部网络通道和数据中心外部网络通道两种类型,其中前者主要用于公有云的备份服务,而后者则主要用于混合云的备份服务,它们均有不同的网络设计与实现方案。
- 云备份资源池:由云服务提供商提供,拥有池化的计算和存储资源,能够支持多种形态的云备份服务,其实现的关键是在按需为用户提供备份资源的同时,尽可能优化用户备份的数据规模,实现用户数据的高效备份和恢复。

(1)云备份客户端

云备份客户端是影响云备份解决方案实施复杂度的关键。传统的数据备份方案普遍采用的是基于代理的备份客户端,即在系统管理员想要进行备份操作时,需要首先在待备份的节点上逐个安装备份代理软件。这种方式下,管理员需要进行繁琐的操作,完成对多种代理客户端的维护和升级,这导致了极高的复杂度。在云环境下,以虚拟机为代表的待备份节点数量庞大、类型众多,势必进一步导致管理问题的恶化。因此,基于代理的云备份客户端不再适用于云计算环境,基于无代理的客户侧备份模式应运而生。

无代理备份不需要在每台待备份节点上都安装备份代理,而只需要在承载虚拟机的物理服务器或者云数据中

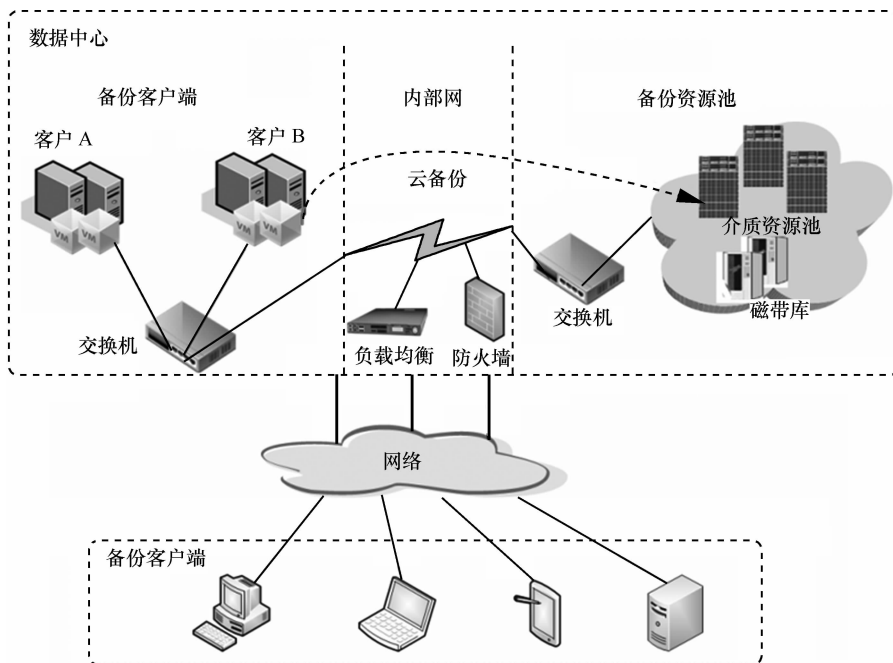


图1 云备份总体架构

心中的某几台设备上安装备份代理,就可以对整个数据中心的虚拟机进行备份管理,相当于集中了所有备份代理的工作,缓解了每台待备份节点上安装代理软件的复杂度。当前无代理备份主要有以下两种方案。

第一种方案是在待备份虚拟机节点所在虚拟化平台的 Hypervisor 中增加备份组件,通过调用 Hypervisor 提供的 API(application programming interface,应用程序编程接口)实现虚拟机相关数据的备份与恢复。这种方式在部署虚拟化环境时即可完成待备份节点的功能部署与配置,缺点是待备份节点必须经过虚拟化层才能实现对物理设备存储数据的访问,从而造成额外的资源消耗,特别是多个备份任务并行时,有可能产生性能瓶颈问题。此外,不同的虚拟化平台的备份组件有可能不兼容,在大规模异构虚拟化平台部署时会有一定的复杂度。

第二种方案是在待备份节点所在的用户侧,部署专门的硬件备份设备作为统一的代理,用户的多个待备份节点通过网络直接与备份设备相连,备份设备可以集中地为不同的待备份节点创建彼此独立的实例,并进行相应的配置,有效地降低了对分散部署的备份代理的管理复杂度。该方式普遍适用于物理环境和虚拟化环境的备份,更具有普适性。专用的备份设备对系统的成本会有影响,因此,如果对性能要求不是很严格,可以利用虚拟机承载方式取代物理硬件备份设备,实现降低系统成本的同时保证部署的灵活性。

以上两种备份方案各有优缺点,用户在实际应用时需要针对待备份节点规模、系统成本和运维复杂度等多个方面综合评估,进行最优选择。

(2) 云备份网络通道

如前所述,云备份网络通道可以分为数据中心内部的网络通道和数据中心外部的网络通道。其中,数据中心外的网络通道主要是指通过 Internet,将备份数据从远程用户侧待备份节点传送到云数据中心以及反向的数据恢复过程,其关键点在于广域网加速等比较成熟的网络技术。因此,本节将重点阐述数据中心内部的数据备份通道方案。

在传统的备份服务中,普遍采用的是基于主机(host-based)的架构或者基于局域网(LAN-based)的架构。这两种备份系统虽然架构简单,但都不具有专门的备份网络设计,而是备份与业务共享业务网络。当备份业务繁忙时,有可能会对正常的业务运行造成影响,从而成为系统的瓶颈。因此,在云备份解决方案中引入了 LAN-free 架构

和 server-free 架构。LAN-free 架构中数据无需通过局域网即可直接进行备份,用户只需将备份设备连接到 SAN(storage area network,存储区域网络)中,各服务器就可把需要备份的数据直接发送到共享的备份设备上,实现了控制流和数据流的分离。它的备份速度较快,但也存在一些问题,如备份过程仍旧需要占用 CPU 处理时间、数据恢复依赖于用户应用、架构的实施会导致成本高等。server-free 架构作为 LAN-free 架构的一种延伸,能够使数据在 SAN 结构中的两个存储设备之间直接传输。由于不需要在服务器中缓存数据,显著减少了对主机 CPU 的占用,服务器负担大大减轻,提高了操作系统的工作效率。但是,server-free 架构也有可能导致兼容性问题,同时在实施上该方案会具有更复杂的实现及更高的成本。

LAN-free 和 server-free 架构都是基于 SAN 的方案设计,它们利用独立于业务网络的专用存储网络实现数据的备份,能够极大地避免对正常业务的影响,更加适用于满足云环境下的备份服务要求,会是当前云备份系统中相对理想的备份网络通道的解决方案。

(3) 云备份资源池

云资源池是用户数据备份的最终位置,虽然它在理论上具有无限的容量,但是在实际建设中,仍旧要充分考虑如何减少备份数据量的规模。一方面是因为云环境下的备份数据量巨大;另一方面,备份数据的保留时长更长,对于一些极其重要的数据还会考虑异地的数据容灾备份。在云备份资源的利用上,最基本的思路是通过权衡备份数据的规模大小、恢复时间长短、主机和网络资源占用等因素选择适当的备份策略,在尽量减少备份数据量的同时,提供高效的备份恢复能力。当前,云备份领域最主要的备份策略包括完全备份、增量备份和差异备份 3 种类型,如图 2 所示。

在实际应用中,云备份资源池的备份策略可以组合应用,以获得更优化的效果,例如采用完全备份与差异备份组合或者完全备份与增量备份组合。

3 云备份关键技术

云备份解决方案的设计与实现需要针对实际的业务需求,在虚拟机备份、物理机备份、数据压缩去重、备份介质等方面开展技术攻关,以更好地满足云备份业务的需求。

(1) 虚拟机备份

虚拟机的数据备份是云备份的主要需求之一。由于虚

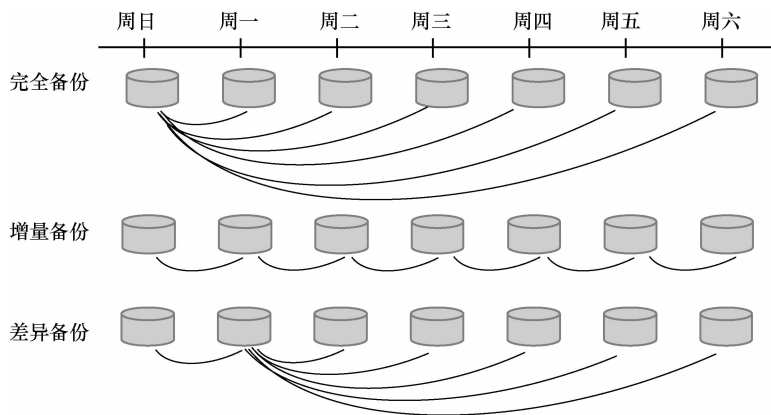


图2 3种云备份策略

拟化平台的存在,备份操作可以通过访问虚拟化平台层来获得待备份节点需要备份的数据,虚拟机的操作系统可以不参与备份过程,从而不会产生额外的性能损耗。最直接的备份过程是将虚拟机磁盘从其对应数据存储空间上挂载过来,实现虚拟机的镜像级备份。此外,常用的基于磁盘的数据保护技术还有快照和持续数据保护(CDP)。虚拟机备份的主要挑战是尽可能地缩短备份窗口,并将网络流量控制在合理的水平。

(2)物理机备份

云计算环境下,虽然虚拟机及其中的应用系统、数据是主要的备份对象,但是对于某些需要安装在物理机上的特殊应用,例如银行系统数据库等,仍需要对整个物理机进行备份。在某些情况下,能否支持高效的物理机备份是衡量一个云备份解决方案成功与否的关键。云环境下物理机备份的关键难点在于,如何通过统一的管理平台实现其与虚拟机相关操作的统一管理。因此,如前文所述的引入集中化的专用备份设备,而无需在每台服务器中部署代理的方式,将会发挥更大的效用。无论是物理机的备份还是虚拟机的备份,除了以服务器本身为粒度进行数据备份外,还可以针对服务器中安装的应用软件以及用户数据开展更细粒度的备份。

(3)数据压缩和去重

在云备份资源池中,除了合理地运用备份策略,优化备份数据量的规模,还可以通过对备份数据进行压缩和去重,尽量减少对存储空间以及网络带宽的占用,这也是云备份系统实现的要点之一。数据压缩是有效减小数据规模的方法,压缩比是衡量压缩效果的重要指标,压缩比越高意味着压缩效果越好,压缩比与备份数据类型有着密切的关系,例如视频、图片等数据的压缩比很低,而对于数据库数据、文本数据压缩比就比较高。除了采用压缩方法,去重

也是减小备份数据量的重要手段,即对于多个备份集中有重复数据的情况,可以对重复数据进行删除,并在数据恢复时填充被删除的数据。

(4)备份介质

传统备份系统中普遍采用磁带作为备份介质,其具有存储容量大、价格低廉、数据不易丢失等优点,但是它的数据串行存取特性导致了较低的性能。在海量数据的云环境下,磁带无法承担巨大的并行数据读写压力,会造成性能瓶颈。因此,云备份系统需要建立数据的分层管理机制。例如在备份过程中,采用磁盘作为备份的第一存储介质,对于需要长期保存的数据则采用磁带作为备份存储介质。随着存储技术的不断发展,利用分布式存储集群构建云备份资源池正在成为业界关注的热点。该类存储系统以通用的x86服务器集群为基础,通过其上部署的分布式中间件实现对分布在各台服务器上的磁盘资源的统一管理,它支持存储数据的并行读写,具有极高的扩展性和性能,同时有效地降低了系统成本。分布式存储集群将会成为未来云备份系统中的关键组成部分。

4 云备份实施方案

针对不同的业务场景和用户需求,云备份的部署和应用可以有多种方案。总体而言包括两种:基于已有的公有云服务构建备份系统的公有云备份方案;通过引入专用备份设备实现私有云数据向公有云资源备份的混合云备份方案。

4.1 公有云备份方案

云主机、云存储是最基本的公有云服务,利用这些服务搭建云备份服务平台,能够充分地发挥公有云的资源优势,帮助用户便捷地实现其数据的备份,同时减轻用户的使用复杂度。

公有云备份服务普遍采用无代理的备份模式,代理组件由公有云服务提供商统一部署在用户的虚拟化环境中,用户根据其数据占用的备份存储空间进行按需付费。用户的待备份节点与备份数据存储介质之间通常具有高速的网络连接,以有效地提升备份性能。在部分场景下,云备份系统可以通过通用接口与外部的云存储或者云备份服务资源池对接(如 Amazon 公司的 S3 对象存储服务),实现更好的资源扩展性。

根据备份资源池的不同,可以提供支持多种形态的云备份服务,包括基于云主机的备份、基于云存储的备份、基于第三方通用接口的备份等,如图 3 所示。

(1) 基于云主机的备份

基于云主机的备份利用云主机自带的存储资源进行备份,该部分存储通常直接部署在承载虚拟机的物理服务器上,以提供较高的访问性能,但是存储的容量普遍较小,因此仅供用户备份少量数据。这种备份方式的优点是不占

用大量的网络资源,用户可以选择在网络资源不紧张的时间段将数据迁移到其他存储上。

(2) 基于云存储的备份

基于云存储的备份是将数据备份到云主机外挂的数据中心的集中存储上,它既可以是共享 SAN 存储,也可以是分布式存储。优点是不像基于云主机的存储有数据量大小的限制,用户可以按需使用;缺点是当用户备份的数据量较大、多用户同时备份时,对网络带宽会带来压力。

(3) 基于第三方标准接口的备份

除了将数据备份在本机存储和共享存储上,用户的数据也可以通过标准接口直接备份到其他数据中心的存储上。采用这种方式,数据中心运营者就不需要预先购买超大存储,而是按需使用其他服务商提供的存储。这种方式需要考虑备份的网络传输成本和压力。

综上所述,公有云备份服务普遍采用已有的云服务构建,通过不同的服务组合满足各种备份服务需求。

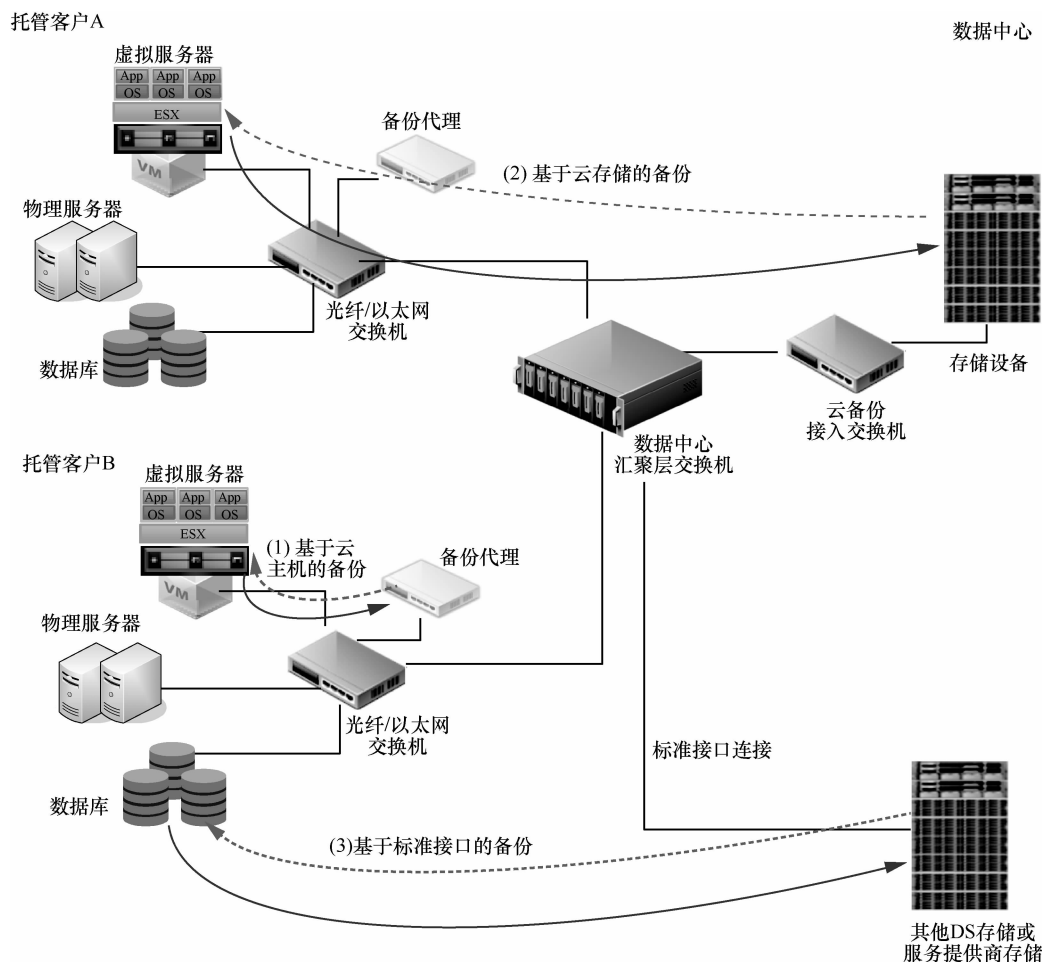


图3 公有云备份服务方案



4.2 混合云备份方案

混合云服务是当前云计算服务的热点,用户通常在拥有自己私有云的同时,借助公有云服务实现其所需资源的按需扩展,例如将部分不敏感、不保密的数据备份到公有云中。

混合云备份与公有云备份存在很大差异,主要体现在以下几个方面:

- 用户私有云管理平台不能与公有云管理平台打通,使得用户无法在其数据备份过程中直接调用公有云资源;
- 私有云中通常存在物理服务器等资源,因此混合云备份内容既包括虚拟机备份,也包括物理机备份;
- 私有云中的数据向公有云备份的过程需要经过VPN(virtual private network,虚拟专用网络),对带宽需求敏感;
- 数据的备份具有双向要求,例如用户部署在公有云中的虚拟机可以将其数据备份回私有云中。

针对上述需求,混合云备份必须采用不同的思路。通过在私有云中引入专用的备份设备,实现对待备份节点的集中化管理和配置是一个行之有效的方案。如图4所示,部署在私有云中的专用备份设备负责和公有云资源建立备份数据传输通道,同时为待备份节点提供相应的配置实

例,以满足不同的备份指标需求,例如备份时间、备份带宽等。当备份过程启动时,待备份节点通过备份设备以既定的备份策略将数据传递至目的存储资源中,实现混合云的备份。混合云备份服务对于私有云和公有云之间VPN通道的带宽消耗非常敏感,因此需要借助于压缩和去重技术,尽量减少备份数据的传输总量,避免额外的带宽占用。此外,还可以通过合理的备份调度机制减小网络拥堵,例如网络繁忙时,先将数据临时备份在本地,然后利用网络闲时带宽择机完成数据的传输与备份。

如前文所述,用于混合云备份的专用备份设备可以是专门的硬件设备,也可利用虚拟机承载实现,这取决于具体的备份需求。

混合云备份服务一方面保证了关键数据的安全性,另一方面又满足了用户对于弹性计算和存储资源的需求,将是未来混合云服务体系中的重要组成部分。

5 结束语

云备份是当前云计算领域的热点,本文基于云备份“端—管—云”的方案架构,全面、深入地分析云备份客户端、云备份网络通道和云备份资源池的设计思路,探讨了

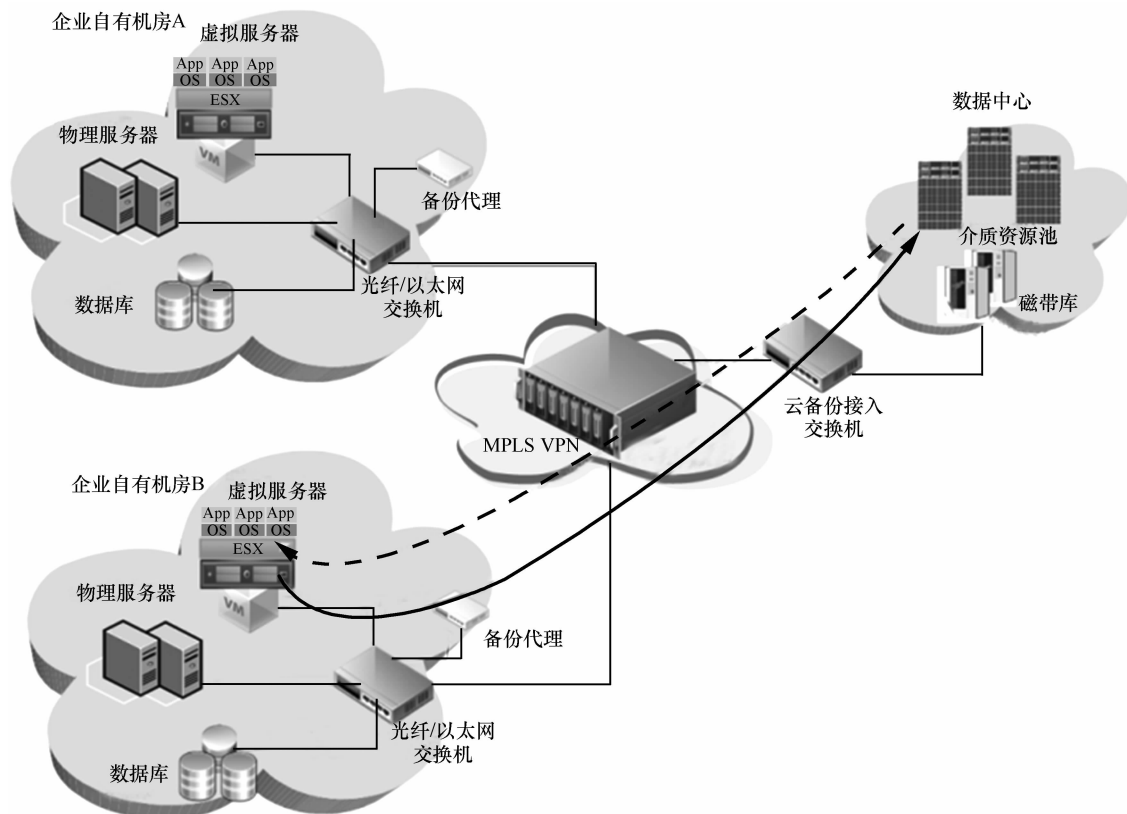


图4 混合云备份服务方案

虚拟机备份、物理机备份、数据压缩和去重等相关关键技术,并针对公有云和混合云的业务场景和用户需求,提出了切实可行的实施方案,有效地解决了云备份现有问题,为云备份系统的研发和部署提供了重要的参考。

云备份技术和服务方兴未艾,当前还有很多难点需要解决,例如更细粒度的应用软件备份、更便捷的混合云备份方案、结合带宽限制的备份策略制定等,这些都是有待进一步研究的问题。

参考文献

- 何欢,何倩.数据备份与恢复.北京:机械工业出版社,2010
He H, He Q. Data Backup and Recovery. Beijing: Mechanical Industry Press, 2010
- 郭涛.按云计算需求设计备份软件架构.中国计算机报,2012-02-13
Guo T. Design backup software architecture according to the needs of cloud computing. China Computer News, 2012-02-13
- 范中磊.网络存储技术漫谈之四:数据备份与恢复.记录媒体技术,2010(4):61~64
Fan Z L. On the network storage technology four: data backup and recovery. China Mediatech, 2010(4): 61~64
- Zeng W Y, Zhao Y L, Ou K R. Research on cloud storage architecture and key technologies. Proceedings of International Conference on Information Systems (ICIS), Phoenix, Arizona, USA, 2009
- Ghemawat S, Gobioff H, Leung S T. The Google file system. Proceedings of the 19th ACM Symposium on Operating Systems Principles, New York, USA, 2003
- Chockler G, Keidar I, Guerraoui R. Reliable distributed storage. IEEE Computer Society, 2009, 42(4): 60~67
- Lin G, Dasmalchi G, Zhu J. Cloud computing and IT as a service: opportunities and challenges. Proceedings of the 6th International Conference on Web Services, Bangalore, India, 2008
- 李君鹏,周宇.基于云存储的在线备份服务创新.中国科技论坛,2010(3):35~37
Li J P, Zhou Y. The innovation of the online backup service based on cloud storage. Chinese Science and Technology Forum, 2010(3): 35~37
- 谢长生,韩德志,李怀阳.容灾备份的等级和技术.中国计算机用户,2003,18(3):30~32
Xie C S, Han D Z, Li H Y. Disaster backup grade and technology. China Computer Users, 2003, 18(3): 30~32
- 王鹏.问道云计算.北京:人民邮电出版社,2011
Wang P. Asked Cloud Computing. Beijing: Posts and Telecom Press, 2011
- 顾炯炯.云计算架构技术与实践.北京:清华大学出版社,2014
Gu J J. Cloud Computing Architecture Technology and Practice. Beijing: Tsinghua University Press, 2014
- 张德丰.云计算实战.北京:清华大学出版社,2012
Zhang D F. Cloud Computing Practice. Beijing: Tsinghua University Press, 2012
- 史爱武.赢在云时代:企业云计算战略、方法和路线图.北京:清华大学出版社,2013
Shi A W. Win in the Cloud Era: Enterprise Cloud Computing Strategy, Method and Route Map. Beijing: Tsinghua University Press, 2013
- 俞东进.杭州市云计算产业发展报告(2012年).杭州:浙江大学出版社,2013
Yu D J. Hangzhou City Cloud Computing Industry Development Report (2012). Hangzhou: Zhejiang University Press, 2013

[作者简介]



司伟,女,中国电信股份有限公司北京研究院高级工程师,主要研究方向为云计算虚拟化技术、云备份、云存储。

王峰,男,中国电信股份有限公司北京研究院高级工程师,主要研究方向为数据中心网络、云管理平台、分布式计算技术。

2015170-7

(上接 2015177-6)

- 王鹏.问道云计算.北京:人民邮电出版社,2011
Wang P. Asked Cloud Computing. Beijing: Posts and Telecom Press, 2011
- 李青,男,中国电信股份有限公司上海研究院高级工程师、专业室主任,主要研究方向为电信增值业务与产品开发、呼叫中心、IDC、云计算/大数据关键技术研究与应用等。
- 周文红,女,中国电信集团公司网运部高级主管,主要研究方向为电信业务平台、云计算/大数据技术应用及相关维护管理工作等。

2015177-7