

基于 EMC DATA DOMAIN 的数据备份系统设计

黄 荣

福州大学信息化建设办公室

摘要: 快速增长的校园网络数据信息和其高可用性对信息中心存储备份技术提出了较高的要求。以我校信息中心为背景, 探讨了存储备份相关知识, 以及我校的建设情况。

关键词: 存储; 备份; 网络; 设计

引言

近年来随着我校的数字化校园和信息化公共服务平台的建设, 计算机存储信息量不断增长, 同时存储在信息系统中的数据已成为学校的重要资源, 数据的丢失以及信息系统的失效都将给学校造成重大的损失。数据存储、数据备份和灾难恢复日益成为引人关注的重要研究课题, 尤其是数据备份占有重要地位。为保护不断增长的数据, 防止数据丢失, 数据备份技术作为数据保护的重要方式, 也得到了长足的发展, 更加受到人们的重视^[1]。

一、信息系统网络结构

我校的信息系统主要包括邮件系统、数据库系统及其他的相关应用服务器系统, 涉及的操作系统平台主要是 windows、

Linux 以及 VMware 虚拟化平台, 数据库以 Oracle 为主。

随着学校信息化建设的快速发展, VMware 虚拟化平台系统的部署越来越多, 服务器虚拟化能快速部署应用系统, 方便计算资源的动态调配和使用, 但同时也带来了备份方面的严重挑战——传统备份手段占用了太多的 ESX 服务器资源, 备份时间太长等一系列问题。因此, 虚拟机的保护有别于传统的物理机, 应当采用重复数据消除的磁盘备份方式来解决虚拟环境的数据保护。考虑到学校信息中心后续向云平台建设的扩展与迁移, 还应该考虑针对学校师生提供基于学校云平台的个人桌面数据保护功能。信息系统体系架构如图 1 所示。

二、备份系统的要求

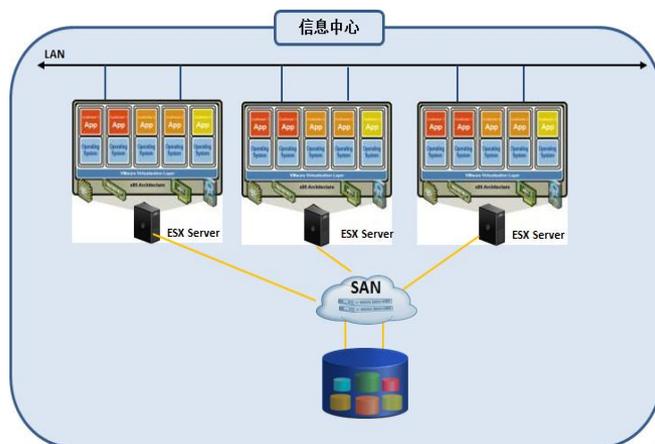


图 1 信息系统体系结构

对数据进行备份是为了保证数据的一致性和完整性，消除系统使用者和操作者的后顾之忧。不同的应用环境要求不同的解决方案来适应，一般来说，一个完善的备份系统，需要满足以下原则：

1. 稳定性

备份产品的主要作用是为系统提供一个数据保护的方法，于是该产品本身的稳定性和可靠性就变成了最重要的一个方面。首先，备份软一定要与操作系统 100% 的兼容，其次，当事故发生时，能够快速有效地恢复数据。

2. 全面性

在目前现有的信息系统中，包括了多种操作平台，包括 windows、Linux、VMware，并安装有各种应用系统，如 OA 系统、邮件系统等。因此，选用的备份软件，要能有效的支持各种操作系统、数据库和典型应用。

3. 自动化

很多系统由于工作性质，对何时备份、用多长时间备份都有一定的限制。因此，备份方案应能提供定时的自动备份，在自动备份过程中，还要有日志记录功能，并在出现异常情况时自动报警。

4. 高性能

随着个人多媒体移动业务的不断发展，数据越来越多，更新越来越快，在休息时间来不及备份如此多的内容，在工作时间备份又会影响系统性能。这就要求在设计备份时，尽量考虑到提高数据备份的速度。

5. 容灾考虑

在建设备份系统的时候需要考虑到容灾的要求，如将备份后的数据通过技术手段，存放一份或者多份在其他地方，如容灾中心，以防数据中心发生不可预测的灾难等。

三、备份系统的组成

备份软件，优秀备份软件包括加速备份、自动操作、灾难恢复等特殊功能，对于安全有效的数据备份是非常重要的。

备份网络，备份网络可以是 SAN，也可以是 LAN /WAN，或 SAN+LAN/WAN 混合方式，它是数据传输的通道，数据备份的效率高低与备份网络有密切关系。

备份介质，介质是数据的负载物，它的质量一定要有保证。

四、备份系统的功能

数据备份系统主要解决业务系统中数据备份工作中存在的备份可靠性差、效率不高，介质管理不科学等问题。因此一个合理的备份系统，至少应该具有以下几个功能[2]。

集中式的数据存储和管理，备份系统的客户端（需要备份数据的设备）可分布在多个机房，只要机房之间有以太网链路或传输资源，都可以实现备份系统地集中管理和数据的集中存储，链路的带宽根据数据量和备份窗口的大小确定。

备份策略自动执行，备份系统的软硬件安装完成后，将根据业务需求和网络状况定制备份策略，包括备份的时间、对象、频率、类型等，定制的策略将自动执行无须人工干预。

文件系统的备份恢复，文件灵活备份，可选择任意文件夹、文件及其组合进行备份。

数据库的备份与恢复，在线完成数据库及相关日志的备份，包括全备份和增量备份；备份的数据库类型包括 oracle、SQL、DB2 主流数据库；可以将数据库备份恢复任何一台机器；可灵活根据实际情况用全备份+增量备份+逻辑日志备份恢复数据库到某个时间的状态。

在线式的数据索引，数据备份到磁盘阵列和磁带库后，可以通过备份系统的管理界面对数据进行管理，可以根据需要恢复某个备份。

归档功能，可以对需要长期保存的数据进行归档。

五、基于 EMC Data Domain 的存储备份系统设计

1. 备份系统体系结构

依据我校信息系统现有建设情况，我们拟建立一套基于现有硬件和网络资源，对信息系统中的关键业务数据进行集中管理备份的保护。

根据系统建设原则和建设目标，新建备份系统的先进性、开放性、扩展能力、管理能力，对于统一备份平台的成败至关重要。我们将采用业界领先，成熟稳定的备份软件以及专业备份设备来构建统一的备份平台。Data Domain 备份架构建立专用备份服务平台，为信息中心各个业务系统提供数据保护服务。建成后的存储备份系统体系结构如图 2 所示。

2. 备份系统备份策略

备份系统的建立是为了能够存储历史数据，同时在数据丢失后，能够恢复需要的数据。备份策略的定制应遵循以下原则：备份对业务和网络的影响尽可能小；备份/恢复窗口尽可能短；恢复的时间点能够达到要求；占用备份介质空间尽可能少。根据实际要求，进行平衡，并且通过实践，总结出适合的合理备份策略[3]。

备份的方式可以分为四种：全备份、增量备份、增量备份、合成备份。

全备份：每次备份定义的所有数据，优点是恢复快，缺点是备份数据量大，数据多时可能做一次全备份需很长时间。

增量备份：备份自上一次备份以来更新的所有数据，其优点是每次备份的数据量少，缺点是恢复时需要全备份及多份增量备份。

增量备份：备份自上一次全备份以来更新的所有数据。

合成备份：将全备份与增量备份组合生成新的全备份，目前只针对文件备份。

3. 备份系统备份容量规划

容量规划与如下因素有关：备份源数据量；数据增量；备份策略；保存周期。

根据我校信息系统关键业务备份策略和保留策略，可以计算出，需要配置多少的存储空间，具体结论表 1 所示。

根据 EMC Backup System Sizer 按照不同类型文件具有不同的重删效果。通过备份策略和保留策略，可以统计出需要的备份设备空间为 15.54TB，考虑到一定的冗余性，配置 20TB 的备份空间。

4. 备份作业管理

对于备份作业，备份系统管理员应定期执行以下操作：

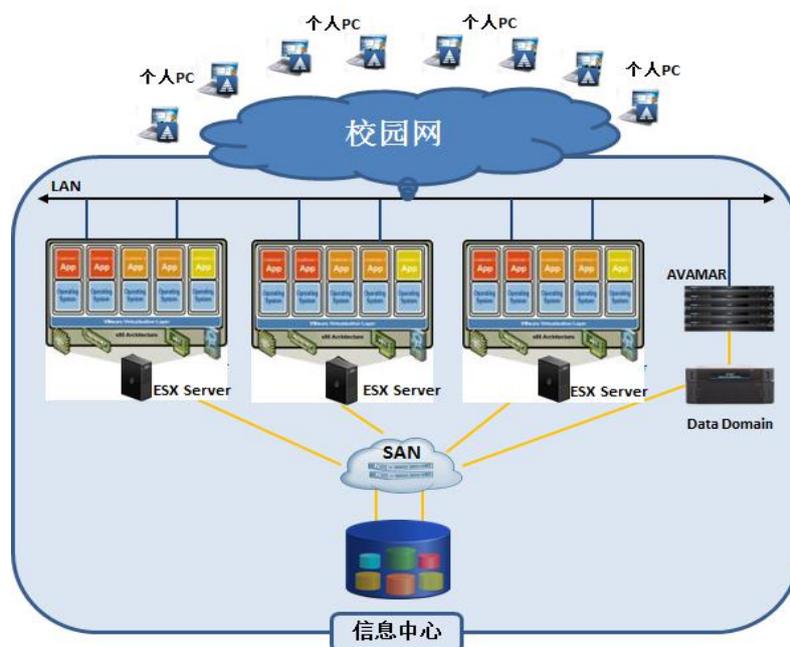


图 2 存储备份系统体系结构

表 1 备份策略

数据类型	数据量	预估增长率	备份策略	备份保留策略	
虚拟机	15TB	1% (天)	每周一次全备份, 每天增量	本地保留 4 个全备份	
邮件	4TB	5% (天)	每周一次全备份, 每天增量	本地保留 4 个全备份	
数据库	2TB	5% (天)	每周一次全备份, 每天增量	本地保留 4 个全备份	

通过观察备份作业运行状态来判断备份作业是否正常, 如果发现作业异常或者失败应尽快排除故障使备份作业正常运行。

当备份介质使用率超过 80% 时, 可以考虑是否需要调整备份数据保存策略或者对备份介质进行扩容。

通过备份软件告警功能可快速跟踪并查找到即将导致或者已经导致备份系统故障的信息, 并通过该信息排除备份系统故障。

备份系统管理员应根据实际业务变化, 适时调整备份策略[4]。

六、结束语

基于 EMC Data Domain 构建的存储备份系统, 具有总体超过 20 倍的重复数据消除率, 解决了整个信息系统备份数据爆炸性增长的所面临的难题。高性能的吞吐量, 备份速度可以达到最高 13TB/每小时, 相对于磁带设备和传统的虚拟磁带库, 备份性能大大提升。存储设备可方便实现扩容, 单台设备可达到最大 130TB 的可用容量, 其可容纳消重前数据量达到 1.3-6.6PB, EMC 存储备份系统可以很好地满足我校信息中心的建设需求。

参考文献

- [1] 韩德志, 谢长生, 李怀阳. 存储备份技术探析. 计算机应用研究, 2004(6):1-4.
- [2] 陆渝. 数据备份技术. 重庆工商大学学报 (自然科学版), 2003(20):89-91.

- [3] Yinjin Fu, Hong Jiang, Nong Xiao, Lei Tian, Fang Liu, Lei Xu. Application-Aware Local-Global Source Deduplication for Cloud Backup Services of Personal Storage IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, 2014, 25(5):1155-1165.

- [4] 谢汶兵, 马晓东, 李中升, 牛夏牧. 基于备份控制流信息的缓冲区溢出监测技术. 计算机工程与应用, 2015(4):1-8.