Universe Scientific Publishing
Education and Learning,教育与学习（2）2020，11
ISSN:2705-0408(P);2705-0416(0)

高性能分布式在线学习平台视频存储系统的设计与实现

刘　骞

（云南开放大学　云南昆明　650223）

【摘　要】　本文介绍了干部在线学习平台的特性，分析了主流独立存储系统的问题，并根据云南省干部在线学习平台的经验，结合 PCIe SSD、分布式架构、DFS 等技术，提出一个构建高性能分布式在线学习平台视频存储系统的设计方案与实现方法。

【关键词】　PCIe；Dfs；视频存储；在线学习平台

DOI：10.18686/jyyxx.v2i11.38416

近年来，随着政府对干部在线教育越来越多重视，在线培训已经成为政府以及国家公职人员提高修养，获取业务知识的一个重要手段。很多省以及地方组织部将在线培训成绩作为领导干部年终考评的依据。各级政府不仅越来越依赖在线教育这种培训手段，也更多的利用干部在线学习平台来进行大规模的领导干部以及公务员培训。尤其在疫情防控常态化的背景下，干部在线教育的作用就更加明显。

1　背景

以云南省干部在线学习平台为例。平台呈现出，注册用户高（数十万用户），在线率高，并发量大等特点。同时，干部在线学习平台课程一般以视频流媒体为主，对于大量用户的访问学习，视频系统要想支撑好，就需具备较高的性能。单靠服务器或者磁盘阵列的方式构架的存储系统无法应付高并发访问。所以配备专业的存储系统势在必行，如企业级的 SAN（存储区域网络）或者 NAS（网络附加存储）。但是这又带来了诸多问题，例如，设备一次性投资大，企业级存储系统动辄都是上百万的价格，后期运维成本很高；系统复杂，管理难度大，需要专业技术人员进行维护，日常运行中若出现停机的情况需要花费很长时间来关闭和启动设备，增加了业务恢复时间；系统利用率不高，很多企业级的功能不一定都能用上；设备容易存在性能瓶颈，如 NAS 存储系统存储控制器（俗称 NAS 机头）就有扩容限制，同时硬盘也是存储瓶颈，若扩容也无法支持新业务发展，那么就不得不考虑投入更多的资金来购买更高级别的产品，淘汰下来的老设备再利用也成问题，未来的升级仍然无法避免同类事情的发生。构建一套高性价比的视频存储系统想必是在线学习平台运维人员欢迎和期盼的。本文根据云南省干部在线学习平台视频存储系统实际经验，提出一套高性能、低成本、易运维、稳定可靠的分布式课件视频存储系统的构建方案。

2　分布式视频存储系统设计

视频存储系统的设计原则是能够响应大并发访问，视频点播快，延迟低，稳定流畅，同时成本低，易于维护和未来扩展，避免重复投资。因此本文提出利用基于服务器 PCIe 接口固态硬盘与 DFS 分布式文件系统构建视频点播存储系统的方案。

存储设备不采用统一存储系统，而是将视频服务器本身作为存储系统。存储介质使用 PCIe 接口的固态硬盘。目前的主流服务器固态硬盘接口分为有 SATA 和 PCIe 两个类型。SATA SSD 作为替代机械硬盘最初形式，用的还是机械硬盘的接口管理方式，通过 SATA 协议，使用和 SATA 机械硬盘同样的接口，传输数据。但是由于 SATA 本身的原因，成为了性能瓶颈，因为 SSD NAND 芯片提供的性能远远超过 SATA 接口的传输能力。作为替代 SATA SSD 的技术，PCIe SSD 解决了 SATA SSD 的性能瓶颈，不通过阵列卡或 HBA 卡进行硬盘管理，而是通过 NVMe 协议直接连接到 PCIe 总线上。消除了 SATA 接口的瓶颈。从生产成本上来说，SATA SSD 低于 PCIe SSD。但是性能上读写速度，PCIe SSD 的效果远远高于 SATA SSD 的。不过经过几年的发展，PCIe 接口 SSD 目前已经普及并得到主流服务器厂商的支持，PCIe SSD 的生产成本也得到了下降，对于在线学习平台来说，视频系统最大性能瓶颈就是 IO，所以 PCIe SSD 是一个很好的解决 O 瓶颈的选择，它可以极大的提升视频点播的响应速度和流畅度。PCIe SSD 性能卓越但是除了成本比 SATA SSD 高以外还有一个问题，它无法实现 Raid，也就不能像机械硬盘或者 SATA SSD 那样设置阵列从而进行硬盘层面的数据保护。单方面增加安装有 PCIe SSD 视频服务器的数量，用负载均衡技术进行响应分摊不能够解决数据唯一性、容灾等核心问题。

利用 DFS 分布式文件系统技术，可以实现服务器级的数据容灾、并且能够实现多台服务器数据同步，与 PCIe SSD 组合能够构建一套响应快、稳定且扩展性强的视频存储系统。PCIe SSD 加 DFS 分布式文件系统可谓一个完美的组合，实现视频快速响应与流畅播放，同时又能弥补数据保护、数据唯一性、灾备等一系列 PCIe SSD 的短板，实现两种技术的互补。DFS 是微软 Windowsserver 上的分布式文件服务，全称 Distributed file systems 及分布式文件系统。通过 DFS，可以让我们通过单一路径访问到所有分散的单一服务器的共享文件夹的内容，即 DFS 命名空间服务。同时通过 DFS 服务，还可以让多个单一服务器的数据进行同步，从而提供文件服务器负载均衡和容错能力，即 DFS 复制。利用 DFS 复制我们就能够在多台视频服务器间建立集群，将多台视频服务器的文件数据进行实时同步与复制，从而实现视频服务器间的热备与冗余工作。无论前端访问压力有多大我们只需要根据实际情况增加视频服务器并加入到 DFS 复制组中就能完成整个视频存储系统的性能提升，实现在线实时的业务横向扩展。通过 VPN 技术我们还可以构架跨物理区域的分布式视频

存储系统。

对于存储系统来说，重复数据删除也是必须具备的功能。本方案中我们利用 Windows Server 的重复数据删除功能就能够优化我们构建的分布式视频存储系统的 PCIe SSD 存储上的存储资源。重复数据删除是 Windows Server 的一项功能，可帮助降低冗余数据对存储成本的影响。启用后，重复数据删除会检查卷上的数据（检查是否存在重复分区），优化卷上的可用空间。卷数据集的重复分区只存储一次，并可以压缩，节省更多空间。重复数据删除可优化冗余，而不会损坏数据保真度或完整性。通过重复数据删除可以节省的空间，重复率很高的数据集的优化率最高可达 95%，存储使用率最高降低 20 倍。

每台用于承载视频业务的服务器配置 PCIe 固体硬盘，安装 Windows Server 操作系统，将所有视频服务器加入到同一个域环境中。每台视频服务器启用 DFS 复制功能，并在 PCIe SSD 盘符上都建立名称相同的文件夹放置视频数据。配置 DFS 群组，同步方式设置为交错复制，这样配置是让群组中的每一台服务器（简称：节点），都可以将数据同步给其它节点，一旦任何一个节点下线或者故障都不会影响其它节点的业务运行，实现节点间的热备与冗余。完成 DFS 群组配置后，每个节点启用重复数据删除功能，目录指向给放置视频数据的文件夹。最后将视频数据拷贝到群组中任意一个节点 DFS 文件夹下，一段时间后数据就会同步到群组中的其它节点的 DFS 文件夹里，同步时间与数据量大小，网络带宽，服务器性能有关。DFS 会实时监控同步目录中的文件状态，对新的文件改动将会实时同步到所有节点上，保证同步目录的数据完全一致。重复数据删除则会对目录下的重复数据进行删除，优化存储空间。至此一套高性能分布式视频存储系统构建完毕。

3 与独立存储系统的对比

在线学习平台的核心业务是向学员提供视频服务，视频服务压力较大，这就要求视频存储系统能够有足够性能完成支撑。普通的阵列或者部门级的存储系统往往无法承载，需要采用专业级高端的 NAS 系统或者 SAN 系统。若采用 NAS 方式，成本很高，系统复杂，需要专业人员进行维护。由于 NAS 的本身特性，NAS 控制器是 NAS 系统的瓶颈所在，一旦视频业务量增加，NAS 系统无法实现快速扩展，需中断业务才能进行扩容。并且控制器扩容也受 NAS 系统型号本身限制，并不能无限扩容，若扩充控制器也无法满足业务需求，就需要采购更加高端的 NAS，投资成本不容小觑，根本用说后期运维开销了。若采用 SAN 方式，成本依旧很高，系统复杂，而且数据传输会收到 HBA 光纤通道的限制，成为瓶颈，且无法进行传输通道扩容，SAN 系统固态硬盘价格不菲，维护困难。对于大并发响应的视频服务，SAN 系统还需要解决数据唯一性的问题，这就需要引入数据同步技术，进一步增加运维难度。本文提出的分布式存储系统，结构简单，部署容易，性能优异运维人员只需管理服务器的本地硬盘即可，无论业务如何扩展，由于采用分布式架构，只需要扩展服务器，且无瓶颈限制。同时节省投资成本和运维成本。

4 结语

本文通过对主流存储系统的分析，结构干部在线学习平台的特性，介绍了一种低成本、高性能、易管理维护的分布式视频存储系统的构建方案。

作者简介：刘骞（1980.11—），男，云南昆明人，研究生，讲师，研究方向：计算机网络技术，数据中心运维，网络安全，在线教育学习平台开发，网络架构设计，UI 设计，软件开发。

【参考文献】

[1]冯雪，李靖. 微软分布式文件系统在实践中的应用[J]. 计算机与网络，2019(4).
[2]顾武雄. 用 DFS 技术进行备份部署[J]. 网络安全和信息化，2018(7).
[3]ALONG，SSD 才是王道——M.2 PCIe 固态硬盘导购[J]. 电脑知识与技术（经验技巧），2019(8).
[4]姜微微，陈乃阔，耿士华. NVMe/PCIe SSD 闪存控制技术在服务器中的应用分析[J]. 信息技术与信息化，2015(5)：118-120.