

高校数据中心业务与存储性能适配的解决路径

林凡瑞 叶衍统*

温州医科大学

DOI:10.32629/acair.v4i1.19360

[摘要] 在教育数字化转型背景下,高校数据中心正面临资源需求动态化和数据流量洪峰化等挑战。传统集中式SAN存储架构受IOPS瓶颈和横向扩展能力不足的技术局限已无法满足需求,而超融合架构虽在分布式计算场景表现优异,但在部分关键业务系统的兼容性仍不如传统架构。为了适配不同业务的存储性能需求,本文结合了两类架构的优势,在利旧基础上,提出了一种基于超融合的混合存储虚拟化架构的解决路径。在温州医科大学的实践中表明该路径的可行性,对同类院校的虚拟化数据中心具有参考价值。

[关键词] 超融合; SAN; 混合存储架构; 数据中心

中图分类号: C37 文献标识码: A

Pathways for Adapting Storage Performance to Heterogeneous Services in University Data Centers

Fanrui Lin Yantong Ye*

Wenzhou Medical University

[Abstract] In the context of digital transformation in education, university data centers are confronted with challenges such as dynamic resource demand and peak data traffic. The traditional centralized SAN storage architecture, constrained by IOPS bottlenecks and limited horizontal scalability, can no longer meet business requirements. Although the hyper-converged architecture performs well in distributed computing scenarios, its compatibility with some key business systems still lags behind that of traditional architectures. This article combines the advantages of two architectures and proposes a construction scheme for a hybrid storage architecture based on hyper-converged technology, building upon legacy system. In the practice of Wenzhou Medical University has demonstrated the feasibility of this solution, providing valuable reference for virtualized data center in similar institutions.

[Key words] Hyper Converged Infrastructure; SAN; hybrid storage; data center

引言

随着教育数字化转型的发展,在线教学平台、选课系统等高并发业务场景呈现爆发式增长。这类业务普遍具有明显的潮汐特征:在日常运行中需要应对突发流量冲击,而在开学选课、在线考试等高峰期则面临指数级增长的流量洪峰挑战。与此同时,适合分布式运行的大数据实时分析等新型应用场景快速普及,教育数据总量呈现指数级攀升趋势。在此背景下,传统集中式SAN存储架构的局限性日益凸显——其纵向扩展能力不足,硬件扩容成本高昂,部署流程复杂且运维难度大^[1]。这些问题导致传统架构既难以满足高校数字化业务对弹性伸缩的核心需求,也无法支撑AI训练所需的分布式并行计算和大数据分析的高吞吐要求。面对这些挑战,构建具备弹性扩容能力、快速部署特性和智能运维等特征的新一代数据中心架构已成为当务之急。这种

架构转型不仅能有效应对教育数字化场景的动态负载变化,更能为高校科研创新提供坚实的技术底座。

1 高校数据中心新业务需求分析

高校数据中心传统的核心系统包括:门户网站、一卡通系统、财务系统、邮件服务等多种应用^[2]。在教育数字化转型加速深化的背景下,高校数据中心面临多重挑战,如业务多元化(在线选课系统、在线教学平台、大数据分析等);在线选课系统需应对瞬时流量激增;教学科研资源配比呈现季节性波动。根据高校常见业务特点分为稳态业务和敏态业务两类。稳态业务如财务、人事、一卡通、考易、邮件等核心数据库,要求强一致性、延迟敏感、高可用、长期稳定等。敏态业务主要有以下几类业务:在线选课系统、在线教学平台、门户网站、考易系统等,要求高并发、突发流量承载能力强、高峰期流量洪峰;邮件

附件等要求海量非结构化数据, 高峰期流量洪峰; 大数据分析、AI训练等要求多节点并行计算加速处理、高吞吐; 敏捷业务、测试系统等要求快速迭代、动态资源调度; 视频等备份数据要求成本敏感、存储空间弹性扩展、延时容忍。

2 传统集中式SAN存储架构在虚拟化数据中心的现状及不足

虚拟化数据中心的本质是构建软件定义数据中心(Software-Defined Data Center, SDDC), 其核心目标是通过抽象化、池化和自动化技术, 使客户能够以更低的成本获得敏捷性、弹性扩展能力和智能化管理体验。数据中心的虚拟化主要包括服务器虚拟化、网络虚拟化、存储虚拟化、安全虚拟化等^[3]。目前国内部分高校虚拟化数据中心仍采用集中式SAN(Storage Area Network)存储架构, 通过独立的光纤通道(Fibre Channel Storage Area Network, FC-SAN)或iSCSI协议(IP-SAN)构建专用网络, 隔离业务流量与存储数据流, 保障传输带宽与低延迟, 以及其高性能、稳定性为关键业务和重要数据提供块级存储服务^[4]。结合虚拟化技术, 系统的部署和管理效率提高了数倍以上。

但随着教育数字化新业务的不断发展, 传统的集中式SAN存储架构已无法支撑和满足所有业务。如AI训练、大数据服务等服务只有在分布式存储的环境中才能充分发挥其作用; 在线教学平台、选课系统等高峰期流量洪峰时易崩溃, 因传统架构增加了烟囱式系统的数量, 随着虚拟机的快速增长和存储阵列的负载提升会导致存储读写瓶颈。同时扩展能力受限、纵向升级成本高(如高端控制器、SAN交换机)、安装部署复杂周期长, 硬件来自多个厂商, 型号配置不一, 运维管理界面不一, 维护难度大等不足亟需解决^[5]。

3 超融合基础架构的优势

超融合基础架构(Hyper-Converged Infrastructure, HCI)凭借硬件资源池化架构与统一智能管理平台, 逐步成为教育行业数字化转型的核心技术。作为下一代云数据中心解决方案, HCI通过软件定义技术实现计算、存储、网络及安全服务深度集成。以标准x86/ARM服务器为硬件载体, 通过虚拟化引擎将物理计算资源(CPU/内存)、存储空间及网络功能深度融合, 解决了传统IT架构中“计算-存储-网络”三层分离导致的资源孤岛难题。支持计算/存储资源的动态池化与按需分配, 结合自动化部署工具, 可实现业务系统的分钟级交付。模块化节点支持横向扩展(Scale-Out)与纵向性能提升。数据安全集成快照/克隆、分布式RAID保护、多租户隔离^[6]。

性能对比分析得出, 稳态业务要求的一致性强度(强一致性和读写时延(延迟敏感)的性能), 集中式SAN存储架构更加出色; 而其他敏态业务所要求的性能中, 如IO性能(高并发)、扩展性(易扩容, 解决高峰期流量洪峰等)、总拥有成本TCO(成本敏感)等, 超融合架构更加出色。表明不同的业务适合不同的架构, 稳态业务更合适部署在集中式SAN存储架构中, 敏态业务更合适部署在超融合架构中。

4 基于超融合的混合存储架构的解决路径

温州医科大学数据中心的建设目标是建立一套高并发、稳

定、易扩展、易管理和可持续发展的全面虚拟化的基础架构, 以满足数字化教育长期持续的扩展能力。根据分析的结果, 设计了在利旧现有集中式SAN存储架构上, 基于超融合的混合存储架构, 利用虚拟化技术对数据中心的计算池、存储池、网络池等重新进行了全面整合。采用分层部署策略: 稳态业务(财务、一卡通等核心数据库): 利旧在集中式SAN存储架构中, 保障数据的一致性及低延迟。敏态业务(在线选课、大数据分析等业务): 迁移至超融合架构中, 利用其弹性扩展能力应对在线选课系统高峰期流量洪峰负载。该设计方案既同时保证了传统业务和数字化业务的性能, 又最大限度的节约了改造成本。

4.1 混合存储的虚拟化架构设计

此混合架构为了实现更好地兼容与管理, 在利旧现有存储、网络架构下, 按业务需求搭建两套资源池详见图1, 一套用于稳态业务的利旧现有集中式SAN架构资源池, 一套用于敏态业务的超融合架构资源池。将两套资源池中的服务器都采用VMware vSphere进行统一的服务器虚拟化, 计算节点部署为ESXi主机, 构建服务器的CPU、内存、存储和网络资源池, 通过vCenter Server进行统一管理, 实现整理硬件资源按需分配。

计算池利旧部分配置了2台HP C7000服务器刀箱(内插6个BL460c G9半高和12个BL660c G9全高的刀片服务器)共18个计算节点、2台华为E9000刀箱(内插16个CH121 V3半高和4个CH242 V3全高的刀片服务器)共20个计算节点, 总共38个计算节点; 新增部分配置了CPU为9*2*Intel Gold 6330和内存为9*512G。并安装了ESXi操作系统用以构建CPU、内存的资源池。

存储池利旧部分配置了2台EMC VNX 5400磁盘阵列(双活)和2台华为2600 OceanStor V3存储(双活), 均采用双冗余控制器架构。用EMC VPLEX(分布式、全冗余的存储硬件产品)将EMC与华为存储整合为统一存储池, 进行跨品牌的存储资源管理, 可不停机的进行异构存储间的迁移。再配合Oracle数据库构建了双活架构的高可用方案; 新增部分配置了9*2*3.2TB NVMe SSD的缓存层和9*4*16TB NL SAS的容量层组合成的混闪池, 搭建软件定义的分布式存储系统的虚拟存储区域网(virtual storage area network, VSAN)。

网络池利旧部分配置2台华为OceanStor SNS2124光纤交换机互为热备, 服务器集群通过光纤网卡(FC HBA)卡连接到光纤交换机接入存储网络, 组成一个FC-SAN架构的网络存储, 通过8 Gbps的光纤通道进行虚拟机与存储设备的通信。HP C7000服务器刀箱中配置了38个千兆物理网卡、华为E9000刀箱配置了100个万兆物理网卡; 新增部分配置了36张万兆物理网卡。配置kernel端口用于ESXi管理网络、IP存储、迁移等。可配置成标准vSwitch(Virtual Switch)(VSS)或分布式vSwitch(VDS)虚拟交换机, 再划分多个端口组(PortGroup), 每个端口组可独立配置虚拟局域网(Virtual Local Area Network, VLAN)。虚拟机通过虚拟网卡(vNIC)接入vSwitch端口, vSwitch通过服务器的物理网卡作为上行链路连接外部网络, 实现了虚拟机访问存储以及与校园网的通信。

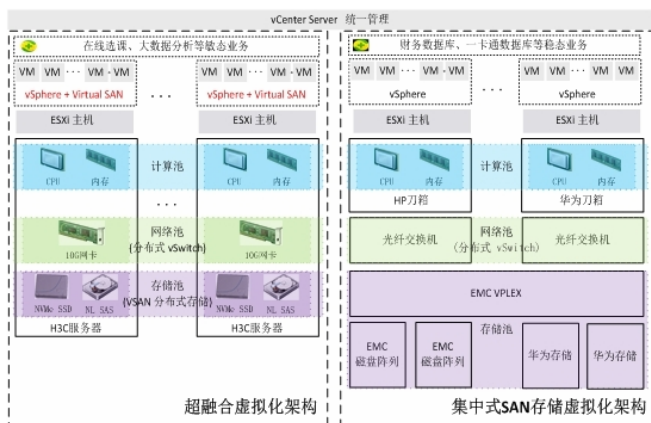


图1 基于超融合的混合存储的虚拟化架构图

5 建设效果

在一年的试运行中,利旧的集中式SAN架构中除了保留核心数据库等稳态业务,其余历史敏态业务迁移到超融合架构中。之前困扰的在线选课系统高峰期流量洪峰时易崩溃的问题,通过临时增加资源得到了缓解,以及适合分布式运行的AI训练等业务也得到更好的发挥。超融合架构减少了存储设备,简化了运维管理,大幅提升了运维人员解决问题的效率。未来随着规模不断发展,超融合架构占比也不断地提升,其绿色节能的优势也将随之凸显,将逐步转型为绿色环保型数据中心。

6 结论

基于超融合的混合存储架构的解决路径,在满足业务性能要求的基础上,最大限度的节约了成本,降低了改造风险,破解

了高校数据中心在教育数字化转型中,因不同业务中对存储性能需求各异难以满足的困境,取得了显著成效。对同类院校的虚拟化数据中心建设具有参考价值。

[参考文献]

[1]李艳,吕鹏,李琰.高等院校高性能数据中心虚拟化系统的建设与研究——以中南民族大学及其图书馆为例[J].图书馆理论与实践,2016,(02):74-79.

[2]叶建华,邓世昆,杨勇.高校数据中心存储系统的规划与构建[J].中山大学学报(自然科学版),2009,48(S1):229-231.

[3]孟凡立,孙荣,徐明.高校虚拟化数据中心建设探究[J].实验室研究与探索,2012,31(12):62-66,71.

[4]朱光菊.高校数字图书馆存储网络建设研究与实践[J].现代情报,2010,30(12):77-80.

[5]刘乃嘉,王茜,程志锐,等.基于分布式文件系统的高校存储平台探索与实践[J].计算机科学,2012,39(S2):54-56.

[6]朱康林.分布式虚拟化存储在公安专科类学校中的应用[J].计算机科学,2016,43(S1):571-576.

作者简介:

林凡瑞(1990--),男,汉族,浙江省温州市人,硕士研究生,研究方向为大数据分析、存储优化。

*通讯作者:

叶衍统(1997--),男,汉族,浙江省温州市人,硕士研究生,研究方向为深度学习、自然语言处理。