

图书馆数据中心虚拟化网络设计研究

许卓斌¹ 林俊伟²

¹(厦门大学信息与网络中心 厦门 361005)

²(厦门大学图书馆 厦门 361005)

【摘要】针对厦门大学图书馆虚拟数据中心的网络设计问题,对底层虚拟网络进行分析和调优,提出数据隔离、链路冗余、带宽共用的虚拟网络设计原则,以期实现一个兼顾性能和可靠性的虚拟化数据中心网络设计方案。

【关键词】数据中心 虚拟化 网络设计

【分类号】TP393

Study on Planning of Library Data Center Virtualization Network

Xu Zhuobin¹ Lin Junwei²

¹(Information & Network Center , Xiamen University , Xiamen 361005 , China)

²(Xiamen University Library , Xiamen 361005 , China)

【Abstract】This paper targets on the network design problems in virtual data center of Xiamen University Library , including analysis and tuning of virtualized infrastructural network. It proposes the design principles of data segregation , links redundancy , and bandwidth sharing in order to implement a virtualized data center network with balance between performance and reliability.

【Keywords】Data center Virtualization Network planning

1 引言

数据中心(Data Center)也被形象地称为服务器农场(Server Farm),指集中部署、运维服务器系统及相关设备的设施。随着厦门大学图书馆信息化建设的发展,各类信息化应用运行所需的服务器数量日益增多。厦门大学图书馆数据中心的主要业务是为厦门大学图书馆各类信息应用服务以及为各类文献论文数据库提供服务器硬件支撑平台。近年来,厦门大学图书馆的信息应用迅速扩展,目前数据中心承载运行的业务系统如表1所示。

厦门大学图书馆数据中心目前承载了56个不同业务系统的运行,并且开发人员还提出了更多用于开发测试和备份的服务器的运行承载需求。在传统数据中心模式下,为了业务运行与管理的安全与可靠性,通常一台服务器只运行一项专有业务系统,关键业务采用双机HA架构提高可靠性。随着业务系统的逐步增加,传统数据中心存在设备数量过多、管理操作繁琐、设备能耗严重、故障多发等问题,因此厦门大学图书馆启动了数据中心虚拟化改造项目。虚拟化技术是可为与数据中心相关的各种IT实体资源创建对应虚拟对象进行操作与应用的抽象方法,可为同类资源提供通用的抽象接口集,隐藏不同实体之间属性和操作之间的差异,进而提高服务器使用效率,降低硬件成本,解决传统单一物理服务器部署的弊端,有效地实现数字图书馆信息化应用系统的硬件利用率与高可用性(HA)^[1]。而数据中心的虚拟化升级中,虚拟化网络规划设计是一个主要难点。厦门大学图书馆承载的大量信息服务平台不仅依赖网络提供正常服务,还依赖网络提供集中存储,如果虚拟数据中心网络规划不当,将造

收稿日期: 2013-05-13

收修改稿日期: 2013-06-04

成虚拟网络拥塞、数据传输效率低下甚至服务异常中断等问题。

表 1 厦门大学图书馆数据中心业务系统清单

业务 IP	业务系统	业务 IP	业务系统
210.34.4.2	SQL 数据库系统	210.34.4.50	VPN 远程接入系统
210.34.4.3	e-Book 系统	210.34.4.52	多媒体主页测试
210.34.4.4	个人主页系统	210.34.4.53	超星图书 1 系统
210.34.4.5	CNKI 系统	210.34.4.54	超星图书 2 系统
210.34.4.6	DSpace 系统	210.34.4.56	OPAC 测试系统
210.34.4.7	校外入口地址转换系统	210.34.4.57	检索数据库测试系统
210.34.4.8	ILAS II 备份系统	210.34.4.60	阅报机系统
210.34.4.9	内部办公程序系统	210.34.4.61	图书外采系统
210.34.4.10	ILAS 系统	210.34.4.62	特藏中转系统
210.34.4.11	CD-NET 光盘数据库系统	210.34.4.67	汇文系统备份系统
210.34.4.12	馆际互借系统	210.34.4.75	万方信息发布系统
210.34.4.13	内部文件系统	210.34.4.91	超星元数据 1 系统
210.34.4.14	FTP 系统	210.34.4.92	超星元数据 2 系统
210.34.4.16	磁带库管理系统	210.34.4.93	元数据搜索系统
210.34.4.17	网络图苑系统	210.34.4.96	员工文件备份系统
210.34.4.21	多媒体中心系统	210.34.4.100	一卡通对接系统
210.34.4.23	ADKS & 高清媒体系统	210.34.4.102	无线网络管理系统
210.34.4.26	重点学科导航网站系统	210.34.4.103	无线网络出口控制器
210.34.4.30	万方数据库镜像系统	210.34.4.107	DSpace 测试系统
210.34.4.32	开发组文件备份系统	210.34.4.109	馆际互借 SAAS 版
210.34.4.33	汇文系统测试系统	210.34.4.111	GitLab 版本系统
210.34.4.34	汇文系统生产系统	210.34.4.112	Moodle@ Sakai 系统
210.34.4.35	维普镜像系统	210.34.4.115	书目检索系统
210.34.4.36	CALIS 系统	210.34.4.117	自助借还系统
210.34.4.45	RFID 系统	210.34.4.118	阅览室终端系统
210.34.4.46	汇文 5.0 客户端测试	210.34.4.122	相图知识库系统
210.34.4.47	OPAC 生产系统	210.34.4.124	万方搜索系统
210.34.4.49	生物论文测试系统	210.34.4.172	开发测试系统

2 虚拟化数据中心网络设计需求及技术思路

数据中心的虚拟化网络是服务器处理业务流量,完成内部存储及管理数据交换的必要途径,其规划设计直接影响整个数据中心的运行效果。传统数据中心架构的设计是为了解决“客户端-服务器”架构模式的应用程序自上而下的大流量数据传输问题,虚拟化数据中心网络则从根本上改变了数据中心网络架构的需求,最重要的原因是虚拟化引入了虚拟机热迁移技术,从而要求网络支持大范围的二层网络区域,这种需求从根本上改变了传统三层网络为主要考虑因素来构建数据中心网络的方案。

厦门大学图书馆数据中心虚拟化项目中使用 VMware4.1 作为虚拟化的平台系统,在网络设计上将虚拟环境中的网络划分为三种主要类型来区分各网络在不同方面的需求,以便规划。

(1) 控制网络:用于连接 vCenter、vSphere Client 和 vCenter Server 的网络,实现对 ESXi Server 的监控与管理;

(2) 核心网络:用于连接 iSCSI 和 NFS/NAS Net-

works 等存储设备,同时提供 VMware vMotion 等主机热迁移^[2]相关功能的数据链路;

(3) 业务网络:用于为运行在虚拟化平台上的信息系统提供服务的网络。

这三种类型的网络,依据各自在虚拟数据中心网络功能分工的不同,对网络带宽性能和可靠性有不同的需求,其功能与需求的对比情况如表 2 所示:

表 2 构成虚拟网络的三类网络功能需求对比

vNetwork	主要功能	带宽需求	独占性	可靠性
控制网络	连接 vCenter、vSphere Client 和 vCenter Server,在 HA 和集群模式下传递心跳、设备负载状态等信息。	低	中	高
核心网络	连接 iSCSI 和 NFS/NAS Networks 等存储设备,同时承载 VMware vMotion 等数据迁移相关功能。	高	低	中
业务网络	虚拟机网络连接,用于提供虚拟机服务器业务系统业务访问的网络端口。	高	低	中

厦门大学图书馆数据中心虚拟化网络设计应保持上述三个功能网络的彼此隔离,并尽可能充分地利用物理网络带宽和现有服务器资源。

国外虚拟化数据中心的实际应用中,网络设计通常要求数据中心的服务器具备针对虚拟化网络需求特性的配置。在每台用于虚拟化的物理主机上配置 6 个以上独立网络接口,为不同功能的虚拟网络配置独立并冗余的物理网卡,独立 VLAN,独立 IP 段,并通过与物理交换机端口的交叉互联以提供物理链路的冗余度。若以这种思路实施厦门大学图书馆数据中心的虚拟化改造,意味着数据中心原有的服务器资源将因为不具备足够的网络接口而无法使用,造成现有资产不能利用,而全部使用新购服务器,这将导致建设成本过高而不具可行性。

国内的小型虚拟化数据中心,则常由于成本考量,没有充分考虑虚拟化数据中心对主机网络接口数量的需求,简单使用端口汇聚(Trunk)技术实现物理链路的冗余和带宽扩容,在实际应用中会经常因为网络隔离度不够而影响整个虚拟化主机系统的网络性能和高可用性。若以这种思路实施厦门大学图书馆数据中心的虚拟化改造,虽然可降低成本,提高硬件利用率,但只部分实现了服务器虚拟化功能,数据中心重要的服务器热迁移(vMotion)能力、服务器高可用性(HA)能力均不能实现,集中存储架构也因为网络性能不足会出现性能瓶颈,最终在功能和性能上不能满足厦门大学图书馆的应用需求。

在厦门大学图书馆数据中心虚拟化工作中,笔者通过对不同功能虚拟网络的需求分析,总结网络性能调优经验,提出了数据隔离、链路冗余、带宽共用的虚拟网络设计思路,综合使用 SR-IOV 虚拟通道技术、VM 虚拟交换机、虚拟局域网(VLAN)、链路聚合(Trunk)等多种技术,针对厦门大学图书馆数据中心设备和业务现状量身定制了一个兼顾成本、性能与可靠性的虚拟化数据中心网络设计方案。

3 虚拟化数据中心网络设计实现方案

厦门大学图书馆数据中心主要用于虚拟化的服务器如表 3 所示:

表 3 数据中心虚拟化服务器物理主机配置清单

品牌型号	CPU 配置	内存配置	存储模式	网络接口数	数量
IBM X3850 X5	4 × 6 核 Xeon	196GB	iSCSI	6	4
IBM X3650 M4	2 × 6 核 Xeon	128GB	iSCSI	4	12
IBM HS22	2 × 4 核 Xeon	64GB	iSCSI	2	4

表 3 中的服务器均通过 iSCSI 模式使用一台 IBM Storwize v7000 作为虚拟化数据中心的集中存储设备。针对表 3 中不同类型的服务器,根据虚拟化系统的功能需求规划基于 ESXi4.1 主机的多个虚拟网络,配置对应的虚拟交换机及其对应 VLAN 及 IP 范围。功能网络及对应虚拟交换机的命名如表 4 所示:

表 4 虚拟化数据中心功能网络及虚拟交换机命名

功能网络	虚拟交换机
控制网络	VM_console_Switch
核心网络	VM_Kernel_Switch
业务网络	VM_port_Switch

以此定义设计表 3 中不同硬件配置的服务器虚拟化数据中心网络的规划方案,在这批服务器上实现 VMware ESXi 主机的部署。

3.1 6 网口服务器虚拟网络规划

厦门大学图书馆数据中心最新购置的 IBM X3850 X5 型服务器,在配置方案上已对服务器虚拟化做了针对性的设计。该服务器配置了 6 个独立的网络链路接口,作为 ESXi 宿主机可以为三类主要功能网络分别分配独立并冗余的网络链路接口、独立 VLAN、独立 IP 段。数据中心接入交换机配置上需要 2 台物理交换机,分别提供 3 个物理端口,通过交叉互联以实现网络物理链路层面的冗余度。同时在网络逻辑规划上,为三个功能网络分别规划独立的 Vlan 和 IP 地址段。虚

拟交换机使用 802.1Q 协议与上层物理交换机连接,物理交换机需要创建与 vSwitch 相同的 VLAN ID,所有物理交换机发布 VLAN ID 到相应端口来保证业务通讯。对 6 网口 ESXi 主机上的虚拟网络的配置规划如表 5 所示:

表 5 6 物理网络端口的 ESXi 主机虚拟网络配置

ESXiHost	vSwitch	VMNIC	Uplink	VLAN	IP
ESXiHost01	VM_port_Switch	VMNIC1	Switch1/1	Vlan1	192.168.1.0/24
		VMNIC2	Switch2/1	Vlan1	192.168.1.0/24
	VM_Kernel_Switch	VMNIC3	Switch1/2	Vlan2	192.168.2.0/24
		VMNIC4	Switch2/2	Vlan2	192.168.2.0/24
	VM_console_Switch	VMNIC5	Switch1/3	Vlan3	192.168.3.0/24
		VMNIC6	Switch2/3	Vlan3	192.168.3.0/24

表 5 给出的网络规划方案中,业务网络、核心网络和控制网络均有独立并冗余的底层物理网络链路、VLAN、地址段支持。在物理服务器出现故障的时候,独立的控制网络可以传输设备状态信息及控制指令。vMotion、iSCSI 等大数据量的传输操作则在核心网络里进行,互不干扰,系统具备极高的可靠性。然而这种方案的网络规划需要有充足的硬件网络资源与逻辑网络资源,每台 ESXi 主机需要配置 6 个以上的网络链路接口,对于厦门大学图书馆数据中心较早购置的大量 4 网口服务器和二网口刀片服务器,这种规划方案并不适用。

3.2 4 网口服务器虚拟网络规划

厦门大学图书馆数据中心的主流服务器 IBM X3650 M4 配置了 4 个独立的网络链路接口,由于网络链路接口数量的不足,虚拟数据中心的网络规划必须有所取舍。业务网络、核心网络和控制网络除了通过物理链路隔离,在逻辑上还可以通过 VLAN 进行隔离(二层网络),也可以通过路由/IP 分段隔离(三层网络),尝试虚拟网络配置方案如表 6 所示:

表 6 4 物理网络端口的 ESXi 主机虚拟网络配置一

ESXiHost	vSwitch	VMNIC	Uplink	VLAN	IP
ESXiHost01	VM_port_Switch	VMNIC1	Switch1/1	Vlan1	192.168.1.0/24
		VMNIC2	Switch2/1	Vlan1	192.168.1.0/24
	VM_Kernel_Switch	VMNIC3	Switch1/2	Vlan2	192.168.2.0/24
		VMNIC4	Switch2/2	Vlan3	192.168.3.0/24

其中业务网络要求占用独立 IP 网段资源,固定占用两个物理链路,并规划在一个虚拟交换机 VM_port_Switch 里,实现链路层的性能负载均衡和冗余能力。而核心网络和控制网络,分别使用一个独立的网络物理链路,这虽然满足了各个网络互不干扰独立运行的要求,但核心网络和控制网络一方面没有了链路冗余

性,另一方面性能也得不到保障。由于服务器依赖于 iSCSI 实现存储,核心网络除了处理 vMotion 这类突发性的带宽应用之外,还需要承载 iSCSI 存储服务流量,只靠单条网络链路支撑会成为虚拟服务器的性能瓶颈,而控制网络带宽需求很小,使用独立的一条物理链路造成了带宽的浪费。在服务器仅有 4 个网络链路接口、2 个网口固定分配给业务网络的前提下,让核心网络和控制网络共用一个虚拟网络,以节约物理链路资源。根据这个思路重新设计虚拟网络配置方案,如表 7 所示:

表 7 4 物理网络端口的 ESXi 主机虚拟网络配置二

ESXiHost	vSwitch	VMNIC	Uplink	VLAN	IP
ESXiHost01	VM_port_Switch	VMNIC1	Switch1/1	Vlan1	192.168.1.0/24
		VMNIC2	Switch2/1	Vlan1	192.168.1.0/24
	VM_Kernel_Switch/ VM_console_Switch	VMNIC3	Switch1/2	Vlan1	192.168.1.0/24
		VMNIC4	Switch2/2	Vlan2	192.168.2.0/24

在该配置方案中,给业务网络提供了三条独立的物理链路做带宽聚合,核心网络和控制网络则使用同一条物理链路和相同的逻辑网络资源,网络地址可以互访。但在实际测试中,在该配置下进行 vMotion(虚拟机热迁移)操作的时候,会发生 vCenter 与 vSphere Client 断开的情况,从而导致 vMotion 失败。在该配置中,ESXiHost01 的核心网络端口 IP 设置为 192.168.2.10(00:50:56:a6:4a:86),控制网络端口 IP 设置为 192.168.2.20(00:50:56:a6:2a:06)。在网络 Vlan2 中,对 vMotion 操作发生的网络数据流进行抓包分析,结果如图 1 所示:

```
01. 00:21:24.742815 e4:1f:13:38:83:10 > 00:50:56:a6:4a:86, ethertype IPv4 (0x0800), length 1514: IP 192.168.3.10.54202 > 192.168.2.10.8000: .45982380:45983828(1448) ack 1 win 4163 <nop,nop,timestamp 524300752 4355672>
02. 00:21:24.742819 e4:1f:13:38:83:10 > 00:50:56:a6:4a:86, ethertype IPv4 (0x0800), length 1514: IP 192.168.3.10.54202 > 192.168.2.10.8000: .45983828:45985276(1448) ack 1 win 4163 <nop,nop,timestamp 524300752 4355672>
```

图 1 4 物理网络端口的 ESXi 主机虚拟网络配置二抓包截图

在观测 TCP ACK 包的时候应该得到一个 MAC 为 00:50:56:a6:4a:86 的包,然而抓包的结果如图 2 所示。

该 ACK 包的源 MAC 现在成了 00:50:56:a6:2a:06,即控制网络端口的 MAC 地址。造成这种情况的原因是虚拟交换机的 MAC 地址表中包含 vMotion 和 Management 两个 MAC 表,当用户连接时使用 Manage-

```
01. 00:28:24.303925 00:50:56:a6:2a:06 > e4:1f:13:38:83:10, ethertype IPv4 (0x0800), length 66: 192.168.2.10.irdmi > 192.168.3.10.66572: .ack 403289767 win 34276 <nop,nop,timestamp 4365775 524300938>
02. 00:28:24.303926 00:50:56:a6:2a:06 > e4:1f:13:38:83:10, ethertype IPv4 (0x0800), length 66: 192.168.2.10.irdmi > 192.168.3.10.66572: .ack 403289593 win 34102 <nop,nop,timestamp 4365775 524300938>
```

图 2 4 物理网络端口的 ESXi 主机虚拟网络配置二 TCP ACK 抓包截图

ment 的 IP 地址,而进行 vMotion 时是另外一个 IP 地址,交换机就会匹配另一个 MAC 表而导致原有的 Client Session 断开连接,MAC 的混乱进而造成该网络里的单播泛洪^[3](Unicast Flooding)。这样的配置无法正常完成 vMotion 操作,不能满足图书馆虚拟化数据中心的高可用性设计需求,因此再尝试虚拟网络配置方案如表 8 所示:

表 8 4 物理网络端口的 ESXi 主机虚拟网络配置三

ESXiHost	vSwitch	VMNIC	Uplink	VLAN	IP
ESXiHost01	VM_port_Switch	VMNIC1	Switch1/1	Vlan1	192.168.1.0/24
		VMNIC2	Switch2/1	Vlan1	192.168.1.0/24
	VM_Kernel_Switch/ VM_console_Switch	VMNIC3	Switch1/2	Vlan2	192.168.2.0/24
		VMNIC4	Switch2/2	Vlan2	192.168.2.0/24

与配置方案二相比,配置方案三为核心网络和控制网络增加了物理网络接口,避免在二层形成单播泛洪,同时还实现了物理链路的冗余,但实际操作中,如果 VMNIC3 或者 VMNIC4 有一个网卡出现故障之后,在这个网络中仍会同配置方案二的情况一样,出现单播泛洪,使得该配置下的两张物理网卡没有起到应有的冗余作用。在该方案下测试 vMotion 操作,没有出现之前的 vCenter 与 vSphere Client 断开连接的情况,但是在人为断开网线制造 VMNIC3 或者 VMNIC4 网卡故障之后,再次进行 vMotion 操作,出现了 Client Session 断开连接的情况,抓包分析显示该网络中再度出现了单播泛洪。说明该配置方案虽然建立了控制网络和核心网络共用的虚拟交换机,也配置了两个物理网络接口,但由于网络没有实现隔离,仍会在一条物理链路出现故障的时候出现单播泛洪,从而导致网络堵塞,系统仍缺乏冗余能力。

为避免在一条物理链路发生故障的情况下出现单播泛洪,需要对核心网络和控制网络进行必须的隔离。以此网络隔离的思路,尝试虚拟网络配置方案如表 9 所示。

表 9 4 物理网络端口的 ESXi 主机虚拟网络配置四

ESXiHost	vSwitch	VMNIC	Uplink	VLAN	IP
ESXiHost01	VM_port_Switch	VMNIC1	Switch1/1	Vlan1	192.168.1.0/24
	VM_Kernel_Switch/	VMNIC2	Switch2/1	Vlan1	192.168.1.0/24
	VM_console_Switch	VMNIC3	Switch1/2	Vlan2	192.168.2.0/24
		VMNIC4	Switch2/2	Vlan3	192.168.3.0/24

在配置方案四中,核心网络和控制网络虽然仍共用一个虚拟交换机,但进行了 VLAN 划分,从二层逻辑上隔离了 vMotion 和 Management 网络,即使 VMNIC3 和 VMNIC4 中有一张网卡出现了故障,由于区隔了 VLAN,交换机在自己的 MAC Map 中也不会认为它们是同一个源,故该设计能够在使用双网卡冗余的同时避免单播泛洪的出现。在该方案下实际测试 vMotion 操作,没有出现 Client Session 断开连接的现象。然后在分别断开 VMNIC3 与 VMNIC4 的网线的情况下进行 vMotion 操作测试,除了 vMotion 操作的损耗时间略有延迟之外,Client Session 仍然保持稳定连接,没有出现断开的情况,网络抓包分析也未见单播泛洪迹象。该网络配置方案可以实现图书馆虚拟化数据中心网络规划的基本需求,最终被用作本数据中心 4 网口服务器的虚拟网络规划方案。

3.3 二网口服务器虚拟网络规划

厦门大学图书馆数据中心还配置有部分 IBM HS22 刀片服务器,刀片服务器具有集成度高、外部线缆少、管理维护简单、占用空间小、能耗低等优势,但由于刀片服务器在体积上的设计,只允许配置 2 个以下物理网络接口,如果使用原有的传统网卡,则 2 个网络链路接口无法完整实现虚拟数据中心的各功能虚拟网络。为了使刀片服务器适应虚拟化网络规划的需求,数据中心购置了支持 SR-IOV 虚拟通道技术^[4]的万兆网卡对刀片服务器的原有网卡进行替换升级。该网卡可将上层软件虚拟化为多个独立的虚拟网卡设备。每一个虚拟网卡对于虚拟化软件平台而言,等同于一块普通的物理网卡。通过给每台服务器上安装 2 个单端口万兆 SR-IOV 网卡,在每个端口上各配置 3 个虚拟网卡,在虚拟化平台层得到了 6 个网络链接接口,从而实现虚拟化数据中心网络的配置。

二网口服务器采用虚拟通道网卡技术,可规划虚拟网络配置方案如表 10 所示。

以上方案利用在刀片服务器上配置 2 个虚拟通道网卡获得了 6 个虚拟网卡,同时把网络流量负载分布在 2 个实体的网络链路通道之上,兼顾数据中心的设

表 10 2 物理网络端口的 ESXi 刀片主机虚拟网络配置

ESXiHost	vSwitch	FNIC	VMNIC	Uplink	VLAN	IP
ESXiHost01	VM_port_Switch	FNIC1	VMNIC1	Switch1/1	Vlan1	192.168.1.0/24
		FNIC2	VMNIC2	Switch2/1	Vlan1	192.168.1.0/24
	VM_Kernel_Switch	FNIC1	VMNIC3	Switch1/1	Vlan2	192.168.2.0/24
		FNIC2	VMNIC4	Switch2/1	Vlan2	192.168.2.0/24
	VM_console_Switch	FNIC1	VMNIC5	Switch1/1	Vlan3	192.168.3.0/24
		FNIC2	VMNIC6	Switch2/1	Vlan3	192.168.3.0/24

备与空间成本,较好地实现了虚拟化网络规划。

笔者针对不同配置条件的服务器制定了上述不同的网络规划,在实践部署中将 IP 地址替换成厦门大学校园网实际 IP,最终完成了厦门大学图书馆虚拟化数据中心的网络配置方案。

4 应用效果评估

厦门大学图书馆数据中心按照以上网络设计方案完成了不同配置的 20 台服务器的虚拟化整合。虚拟数据中心的高可用性(HA)能力高度依赖于虚拟网络的带宽与性能,通过模拟故障测试一台 ESXi 主机故障后的宿主机热迁移(vMotion)过程耗时的长短,可以直观反映出虚拟网络的运行性能。测试中选择一台 4 网口 ESXi 主机作为故障模拟对象,手动将该台主机置于维护模式,观察位于该主机上虚拟机热迁移过程和服务中断时间。测试虚拟主机如表 11 所示:

表 11 实际运行的某 4 物理网络端口的 ESXi 主机虚拟主机清单

ESXiHost	控制网络 IP	虚拟主机 ID	业务网络 IP
ESXiHost01	121.192.190.188	jwc01	210.34.4.23
		Info	210.34.4.54
		Bioinfo-oracle	210.34.4.49
		HTTestBBS2012 64bit	210.34.4.32
		erlang 1	210.34.4.40
		Ryu1	210.34.4.172

该 ESXi 主机上原运行了 6 个虚拟主机,在手动进入维护模式以后,虚拟主机成功地进行了自动热迁移,vCenter 上的事件日志如图 3 所示:

名称	目标	请求开始时间	开始时间	完成时间
退出维护...	121.192.190.188	2013/5/23 10:44:07	2013/5/23 10:44:07	2013/5/23 10:44:09
迁移虚拟机	Info	2013/5/23 10:31:48	2013/5/23 10:31:48	2013/5/23 10:33:54
迁移虚拟机	Info	2013/5/23 10:31:48	2013/5/23 10:31:48	2013/5/23 10:34:19
迁移虚拟机	Bioinfo-oracle	2013/5/23 10:31:48	2013/5/23 10:31:48	2013/5/23 10:33:32
迁移虚拟机	HTTestBBS2012...	2013/5/23 10:31:48	2013/5/23 10:31:48	2013/5/23 10:33:12
迁移虚拟机	Ryu1	2013/5/23 10:31:48	2013/5/23 10:31:48	2013/5/23 10:33:38
迁移虚拟机	erlang 1	2013/5/23 10:31:48	2013/5/23 10:31:48	2013/5/23 10:33:52
进入维护...	121.192.190.188	2013/5/23 10:31:47	2013/5/23 10:31:48	2013/5/23 10:34:24

图 3 4 物理网络端口的 ESXi 主机 vMotion 事件日志

该主机的 vMotion 事件日志显示 6 个虚拟主机在 1 至 3 分钟的短暂时间里成功地完成了迁移。同时在

vMotion 期间对虚拟主机 Ryu1 (210.34.4.172) 进行持续的 Ping 测试,结果如图 4 所示:

```

来自 210.34.4.172 的回复: 字节=32 时间<1ms TTL=128
来自 210.34.4.172 的回复: 字节=32 时间<1ms TTL=128
请求超时。
来自 210.34.4.173 的回复: 无法访问目标主机。
来自 210.34.4.173 的回复: 无法访问目标主机。
来自 210.34.4.173 的回复: 无法访问目标主机。
来自 210.34.4.173 的回复: 无法访问目标主机。

```

图 4 对虚拟主机 Ryu1 vMotion 过程中的 Ping 测试记录

在 Ping 测试中可见虚拟主机 Ryu1 在宿主机发生模拟故障以后,仅短暂停止响应了 6 个 Ping 测试包后就恢复了正常工作。

以上对厦门大学图书馆虚拟数据中心实际运行的 ESXi 主机进行模拟故障的测试中,vMotion 的事件日志和对发生迁移虚拟主机的 Ping 测试记录充分表明了数据中心虚拟网络的性能和功能达到了设计需求。

厦门大学图书馆数据中心在虚拟化改造前后的主要相关指标对比如表 12 所示:

表 12 数据中心虚拟化改造前后主要相关指标对比

主要相关指标	虚拟化改造前	虚拟化改造后
实体服务器总量(不含存储设备)	56 台	20 台
实测平均运行功率(含空调系统)	64KW	38KW
最大独立业务系统承载数	56	>200
主机跨硬件热迁移能力	无	有
主机高可用性能力	无	有

从表 12 可以看出,数据中心通过此次虚拟化改造降低了约一半的用电量和物理实体主机维护工作量,而业务系统承载能力扩大到了原来的 4 倍左右。同时虚拟化网络改造后,数据中心还具备了原先没有的主机跨硬件热迁移和高可用性能力,极大地减少了因硬件故障导致业务系统服务中断的可能性,提升了数据中心的品质。数据中心的整体硬件资源利用率、系统可靠性、可维护性比以前都有了较大提升,取得了良好的实际应用效果,得到了用户的广泛认可。

5 结 语

基于对厦门大学图书馆虚拟化数据中心网络规划的实践与研究,笔者对虚拟化数据中心的网络设计总结如下:

(1) 数据隔离:在逻辑网络架构设计上,对于虚拟化数据中心的控制网络、业务网络、核心网络,都必须

在二层上利用 VLAN 加以隔离,或在三层上利用 IP 地址分段加以隔离。

(2) 链路冗余:业务网络和核心网络的流量应使用分别独立的虚拟交换机。需要冗余的虚拟交换机至少应分配两个物理适配器端口,避免单条物理链路失效导致业务中断。

(3) 带宽共用:控制网络上传输的主要是主机状态信息和操作指令,虽然对网络稳定性要求高,但实际占用的带宽并不大,可同核心网络共用带宽资源,以保障如 vMotion 这种需要传递大量数据应用的效率。

虚拟化数据中心的实际运行性能在很大程度上依赖于底层功能网络的设计与优化,厦门大学图书馆虚拟数据中心通过本文规划的虚拟化网络方案,实现了 ESXi 虚拟主机集群的 HA 和 vMotion 功能。但在数据中心的实际运行过程中,随着虚拟主机数量规模的扩大,故障迁移过程中仍会导致一定时长的服务中断,未来仍需深入研究数据中心物理交换机同虚拟交换机的协同优化方法,进一步提高虚拟化数据中心的虚拟主机热迁移效率,最大程度缩短业务故障中止响应时间。

参考文献:

- [1] 朱海涛. VMware 虚拟化系统在高校图书馆中的构建与应用[J]. 现代图书情报技术, 2012(1): 68-72. (Zhu Haitao. Construction and Application of Virtualization System Using VMware in University Library[J]. *New Technology of Library and Information Service*, 2012(1): 68-72.)
- [2] 张雅彬,李伯虎,柴旭东,等. 基于虚拟化技术的云仿真资源迁移技术研究[J]. 系统仿真学报, 2011, 23(6): 1268-1272. (Zhang Yabin, Li Bohu, Chai Xudong, et al. Research on the Virtualization-based Cloud Simulation Resource Migration Technology[J]. *Journal of System Simulation*, 2011, 23(6): 1268-1272.)
- [3] 厉晓华,郭晔. 校园网内单播泛洪机制分析及其对策[J]. 科技创新导报, 2009(12): 20. (Li Xiaohua, Guo Ye. Analysis and Countermeasures of Unicast Flooding Mechanism in Campus Network[J]. *Science and Technology Innovation Herald*, 2009(12): 20.)
- [4] 李超,董青,戴华东. 基于 SR-IOV 的 IO 虚拟化技术[J]. 电脑与信息技术, 2010, 18(5): 1-5. (Li Chao, Dong Qing, Dai Huadong. IO Virtualization Technology Based on SR-IOV[J]. *Computer and Information Technology*, 2010, 18(5): 1-5.)

(作者 E-mail: zbxu@xmu.edu.cn)