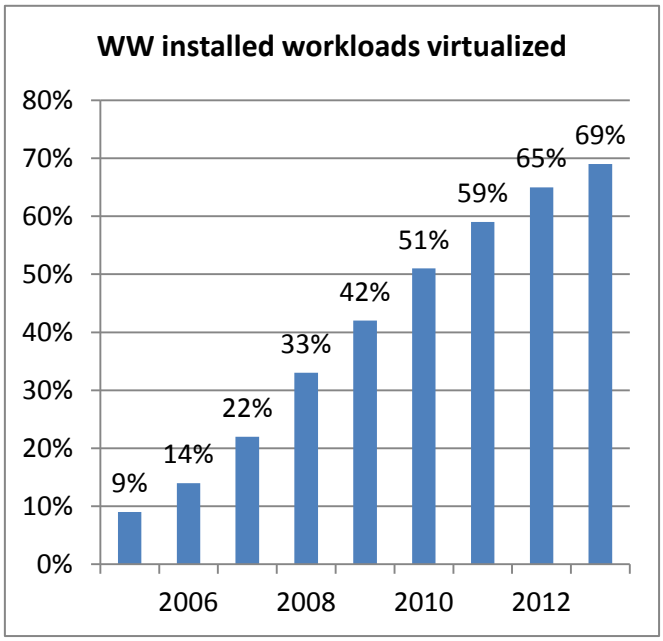
虚拟化数据中心的网络“指挥塔”

文/子康

随着技术的成熟和应用的推动，很多企业的IT系统已经迈出了走向云计算的第一步，这一步就是虚拟机的应用。在一台物理服务器上虚拟出多个服务器，这种方式给IT带来了实实在在的效益：服务器的采购数量减少了；用虚拟机为原来没有HA的业务增加上了HA，减少了业务中断的投诉和抱怨；硬件不再只发挥出10%的能力，有了更强劲的用武之地，……

下图是IDC 2011年的报告，报告显示，2010年已经有51%的负荷是由虚拟机在承担；到2013年，将有69%的负荷将由虚拟机来承担。

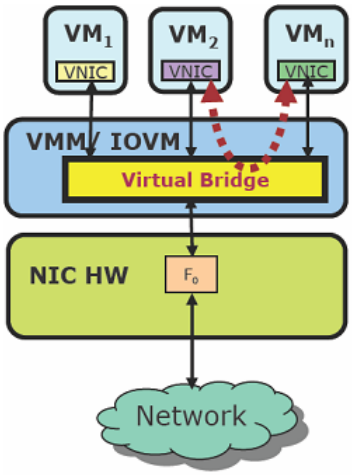
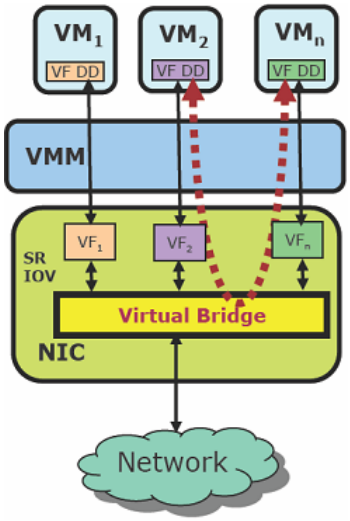
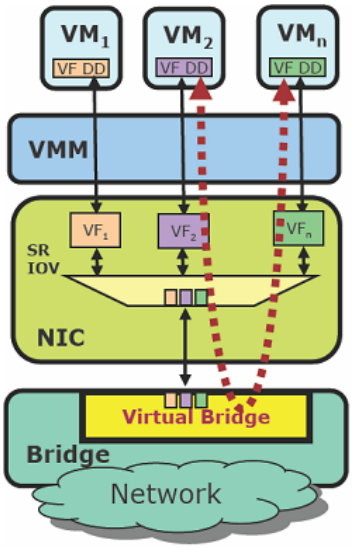


和虚拟机的应用与技术发展相比，网络对虚拟机感知的发展明显滞后了，这给虚拟机的应用带来了诸多困扰。服务器中需要有虚拟交换机来实现虚拟机间的通信；虚拟交换机的管理需要服务器管理员掌握网络知识，或者网络管理员将手伸到服务器领域；虚拟机网络故障的问题，是在虚拟交换机，还是在物理网络，故障的定位更加困难；随着虚拟机技术的发展，虚拟机迁移、资源池调度的应用越来越普及，在虚拟机迁移的目的地，网络需要提前就绪，包括配置、动态表项等。

就像飞机起降前机场的准备就绪需要由机场指挥塔来调度一样，虚拟机起降前网络的准备就绪也需要网络“指挥塔”来调度解决。

虚拟机的网络环境分析

IEEE标准802.1Qbg[1]中对虚拟机接入网络的各种方案做了全面的总结。主要有以下三种：

软件实现的虚拟交换机	由智能网卡实现交换机功能	接入交换机实现环回交换
		
优点：产品成熟，各虚拟机平台都提供；接入交换机采用普通交换机即可。 缺点：占用服务器资源；性能较低；实现的网络功能有限；数据流量管控困难。	优点：性能较高；接入交换机采用普通交换机即可。 缺点：虚拟机的实时迁移难以实现；需要采用具有该功能的网卡；数据流量管控困难。	优点：性能高；便于网络统一管理；数据流量管控容易。 缺点：需要采用支持该功能的接入交换机

“软件实现的虚拟交换机”是最原始、最基本的方式，VMWare ESX、微软 Hyper-V 等虚拟机平台都作为基本功能模块提供。“由智能网卡实现交换”是网卡厂商主导的硬件加速方案，虚拟机平台对这种方式也已经逐步支持起来。这两种方式在数据流量管控上都存在较大的困难，例如，要实现流量采集，需要在物理服务器中创建一个虚拟机，专门用来运行探针。“接入交换机环回转发”的方式性能最佳，流量管控能力最强，但需要接入交换机支持该功能，由于部分旧的交换机需要被更换，适合在新建数据中心部署。

总之，无论哪种方式，网络“指挥塔”都需要能够很好的支持。

[1]截至2012年4月15日，802.1Qbg尚未成为标准

“指挥塔”要纵观全局

网络“指挥塔”要能够看到“虚拟交换机”，看到虚拟机和虚拟交换机的连接关系，看到“虚拟交换机”和物理交换机的连接关系，这是对虚拟机网络进行指挥调度的基础。

本文以华为公司的相应产品为蓝本，介绍业界在虚拟机网络管理方面的技术应用。

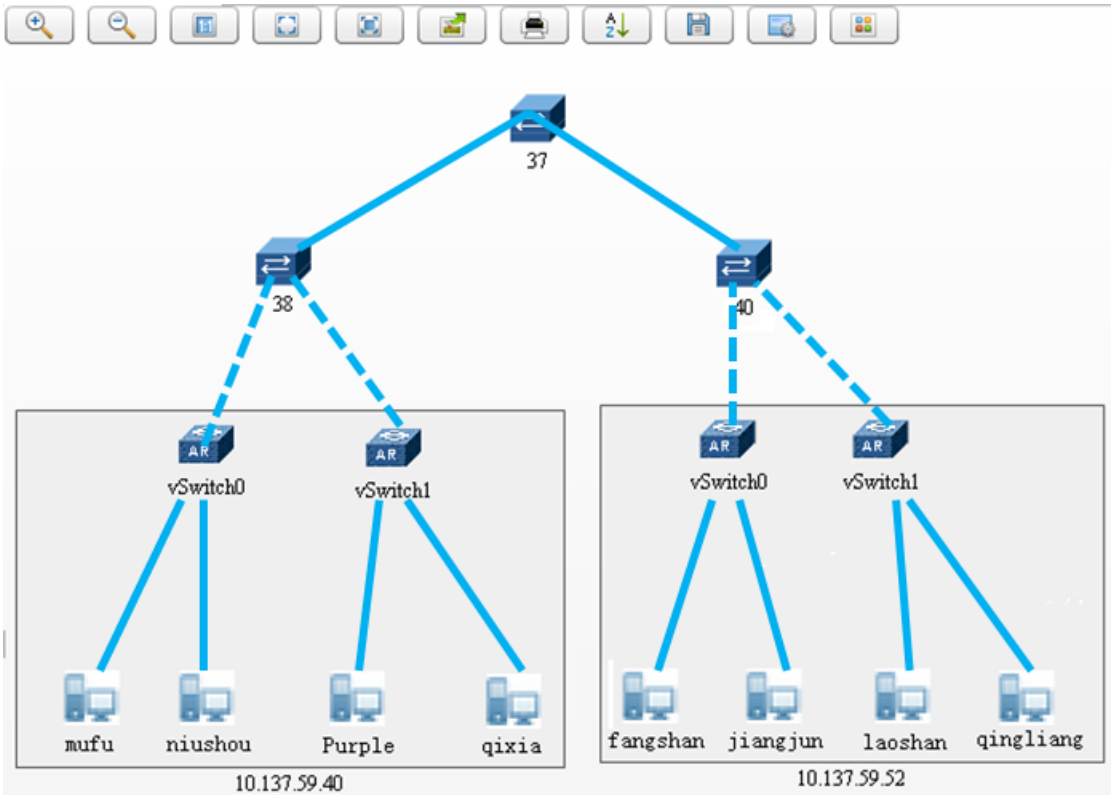
华为公司的虚拟机网络“指挥塔”nCenter（Network Center），和VM的管理器vCenter是兄弟。

虚拟机网络管理一般包括虚拟资源管理和虚拟机迁移管理，其中虚拟资源管理是指物理和虚拟资源的信息采集和拓扑管理，物理和虚拟资源包含虚拟机、虚拟交换机、物理服务器、物理交换机。

nCenter通过网管标准协议发现TOR，通过vCenter的开放接口从vCenter获取虚拟机信息（包括虚拟机和虚拟交换机的连接关系）。

TOR通过LLDP、CDP等设备发现协议发现虚拟交换机，明确虚拟交换机和TOR的拓扑关系。

综合上述信息，nCenter能够绘制出完整的物理、虚拟资源及拓扑。下图是nCenter上查看虚拟机网络拓扑的一个实例。



图中38、40是TOR，下面的框分别是两台物理服务器，内部各有几台虚拟交换机和几个虚拟机。图中清晰简洁地呈现出了物理节点、虚拟节点，物理、虚拟的拓扑连接关系也清楚明了。故障定位有图可循，能够极大的提高管理维护效率，降低管理维护成本。

拓扑管理还提供搜索功能，在大规模的网络中，也能方便快捷地搜索出虚拟机。

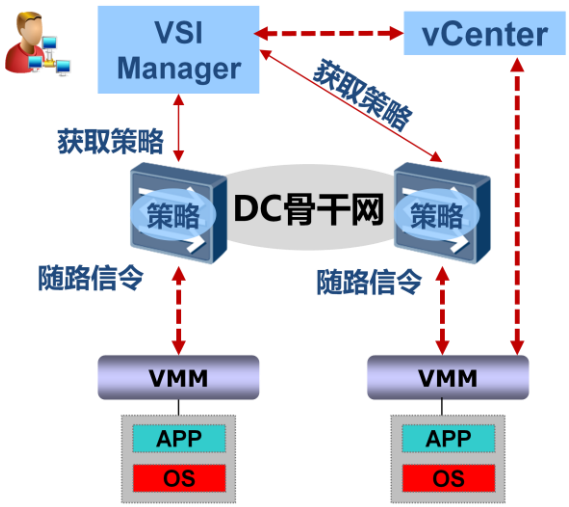
虚拟机“起降”的指挥调度

仅仅实现拓扑管理，还不能成为“指挥塔”。“指挥塔”还必须能够管理虚拟机的“起降”（迁移），在虚拟机“起降”时，网络必须能够按需配置、动态调整、及时就绪。

每个虚拟机，根据其部署的具体业务，需要规划网络配置，包括：QoS、ACL等。在虚拟机部署前，需要先在nCenter上创建策略模版，策略模版被统一管理起来，虚拟机“起降”时，参数配置就可以从策略模版中获取。

开放支持各种虚拟机“起降”

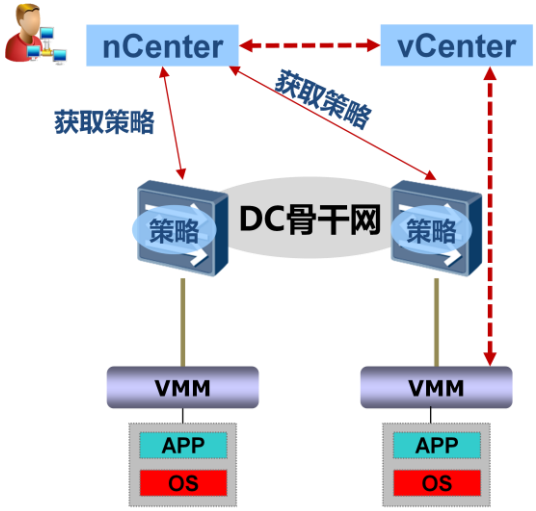
IEEE的标准802.1Qbg有两种方案，一种是“带内管理”方案。如下图所示：



其中VSI Manager是管理VSI(Virtual Station Interface)的配置信息，即上面说的虚拟机的策略模版。随路信令在802.1Qbg中定义了，包括：ECP(Edge Control Protocol)用于封装VDP协议；VDP(VSI Discovery and Configuration Protocol)用于VSI的发现与配置，基于ECP；和CDCP(S-Channel Discovery and Configuration Protocol)用于配置、创建和删除S-Channel（非必选）。

服务器中的虚拟机创建、删除，通过VDP协议通告给TOR，TOR向VSI Manager获取网络策略，完成网络属性配置。因为VDP协议和虚拟机的网络链路是同一条，因此，称为“带内管理”。

另一种方案是“带外管理”方案。如下图所示：



虚拟机的创建、删除、迁移，vCenter是控制发起者，通过vCenter的开放接口通知到nCenter，nCenter向相关的网络设备下发相关的网络策略的配置。

“带内管理”方案，由于目前协议标准尚未确立，各虚拟机平台厂商均未推出相关产品，协议未具体规定和vCenter的接口，VSI Manager需要针对各虚拟机平台进行适配，因此，目前还难以实际应用。

“带外管理”方案，各虚拟机平台厂商均提供开放接口，nCenter按照开放接口适配各虚拟机平台，是开放合作的方案。

采用“带外管理”方案，不用等待虚拟机平台支持802.1Qbg标准的VDP，现在就可以用虚拟机平台的开放接口，实现虚拟机的网络感知。

因此，华为nCenter采用了“带外管理”方案，开放支持VMware、Citrix Xen，以及微软Hyper-V等虚拟化平台。

繁忙的机场需要高效的调度技术

nCenter向网络设备下发策略配置，可以采用命令行、SNMP或NETCONF等方式，但在原型测试中，发现性能只能达到每秒10～20个虚拟机上线；而采用RADIUS协议，原型测试结果能够达到每秒200个虚拟机上线，这个性能能够满足多大规模的虚拟机“起降”呢?

让我们来计算一下，假设有N个物理服务器，其中1/2忙，每个忙的服务器需要将4个VM（测试数据，受限于带宽和CPU能力）迁移出去，每个虚拟机的迁移时间为3分钟（180秒）。则每秒处理虚拟机迁移数量为：N/2×4/180。

如果是1万台物理服务器，则每秒处理虚拟机迁移数量为：10000/2×4/180=111。

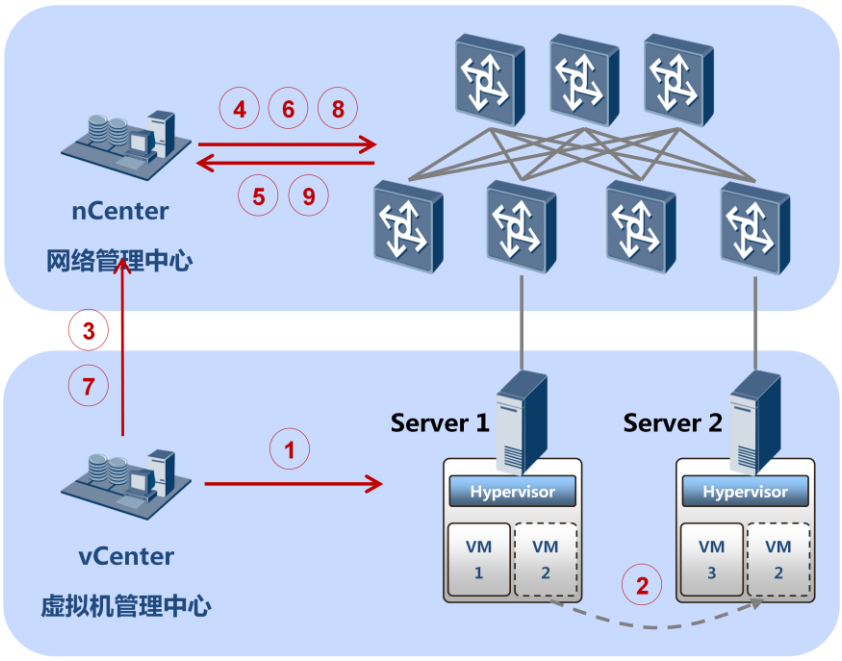
200个虚拟机每秒的处理性能，可以满足：200×180/4*2＝18000台物理服务器的云计算环境。

nCenter采用RADIUS协议，能够满足近2万台物理服务器的云计算环境虚拟机突发大规模“起降”。

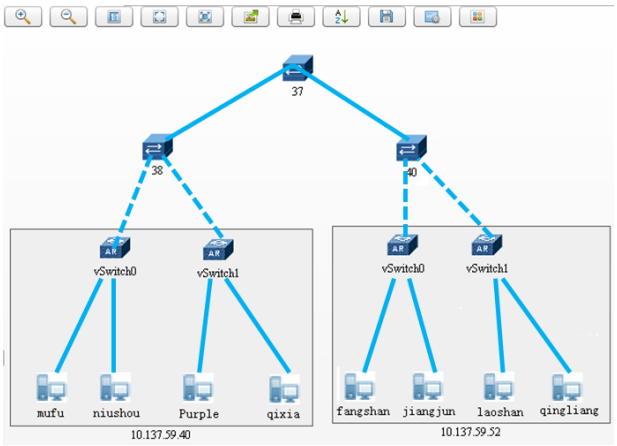
虚拟机“起降”

在虚拟机“起降”的过程中，nCenter负责网络策略的迁移，和虚拟机平台vCenter配合，保证了流程处理的及时、准确和自动化。

下图是虚拟机迁移的流程：

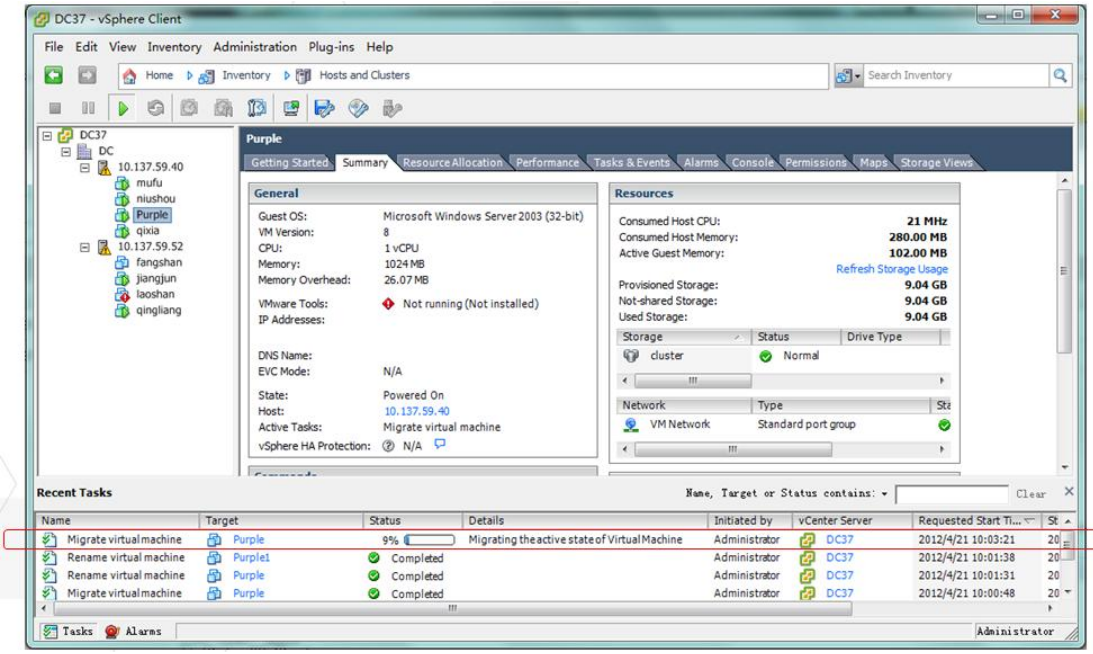


迁移前的拓扑：准备将“Purple”这个VM迁移到服务器.52上



① vCenter启动VM迁移。

② 进行VM迁移。



③ vCenter通过开放接口通知nCenter迁移开始。

④ nCenter通知目的TOR交换机VM上线，上线信息中包含VM的身份信息：VM的MAC地址、VLAN信息、应用的策略模板ID。

⑤ 目的TOR通过RADIUS协议向nCenter申请VM策略（ACL、QoS、DHCP Snooping绑定表）。

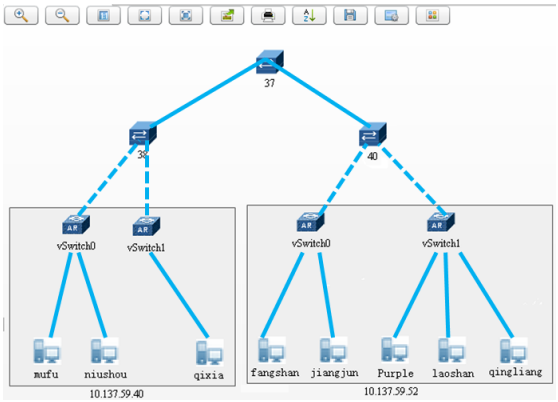
⑥ nCenter内置的RADIUS服务器响应目的TOR的申请，将VM绑定策略应答给目的TOR，目的TOR收到后，解析出VM的策略，完成转发配置。

⑦ vCenter通过开放接口通知nCenter迁移完成。

⑧ nCenter通知源TOR交换机VM下线。

⑨ 源TOR接收到下线通知，删除本地策略的同时，通过RADIUS的用户下线接口通知RADIUS服务器更新用户在线状态。

迁移后的拓扑：虚拟机“Purple”已经成功迁移到服务器.52上



华为虚拟感知解决方案总结

华为虚拟感知解决方案以nCenter为核心，和各种虚拟化平台vCenter开放兼容，在接入交换机上及时下发静态配置、动态表项，实现了对服务器虚拟化环境的全面支持，通过开放高效的网络“指挥塔”，真正建立起供虚拟机自由“起降”的“云机场”。

让我们回顾一下前面分析的网络和虚拟机配合的问题：

管理界面：系统管理员只需要管理服务器、虚拟机，网络管理员负责管理虚拟交换机、物理交换机、虚拟机的网络属性，管理界面清晰。

可视运维：nCenter提供虚拟机、虚拟交换机、物理服务器、物理交换机的一体化拓扑视图，方便故障定位。

虚拟感知：虚拟机创建、迁移，都能够得到及时感知、处理，网络开通速度快，和虚拟化平台、服务器的兼容性好。

结束语

华为虚拟感知解决方案，为虚拟机的应用打通了网络之“脉”，助力数据中心进一步扩大虚拟机的应用，降低IT成本，提高IT效率。相信在不久的将来，一定能够建设起来完全虚拟化、自动化、高效的云计算基础设施，以IaaS的方式支持大容量、多样化的业务应用系统，更好地满足支撑业务运营创新发展的需要。