**解析思科数据中心虚拟化技术和部署**

数据中心的发展正在经历从整合，虚拟化到自动化的演变，基于云计算的数据中心是未来的更远的目标。整合是基础，虚拟化技术为自动化、云计算数据中心的实现提供支持。

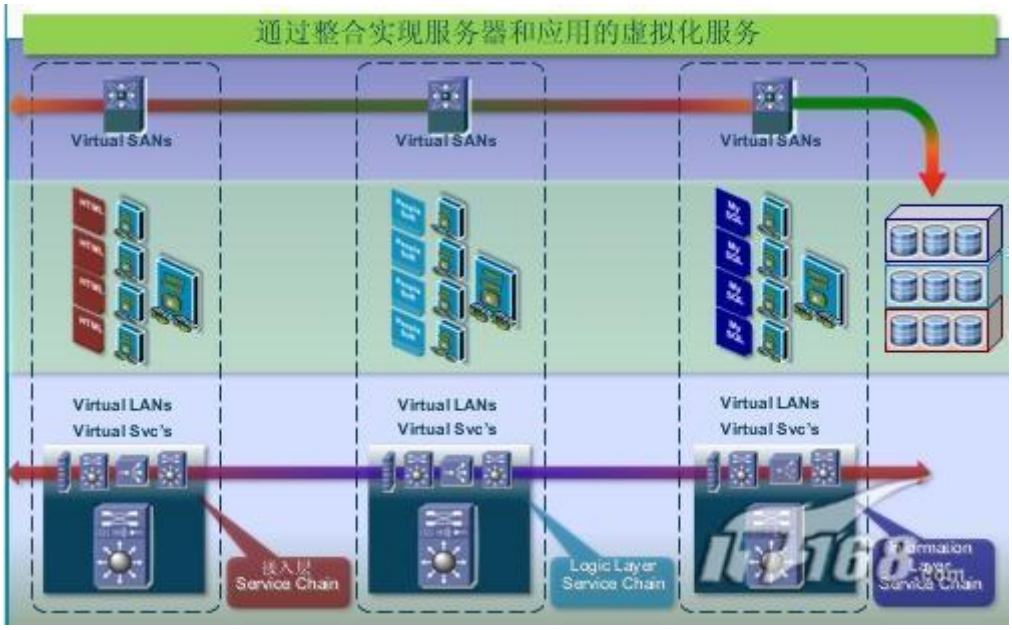
数据中心的虚拟化有很多的技术优点：可以通过整合或者共享物理资产来提高资源利用率，调查公司的结果显示，全球多数的数据中心的资源利用率在15%～20%之间，通过整合、虚拟化技术可以实现 50%～60%的利用率;通过虚拟化技术可以实现节能高效的绿色数据中心，如可以减少物理设备、电缆，空间、电力、制冷等的需求;可以实现对资源的快速部署以及重部署以满足业务发展需求。

数据中心虚拟化的简单示意图。

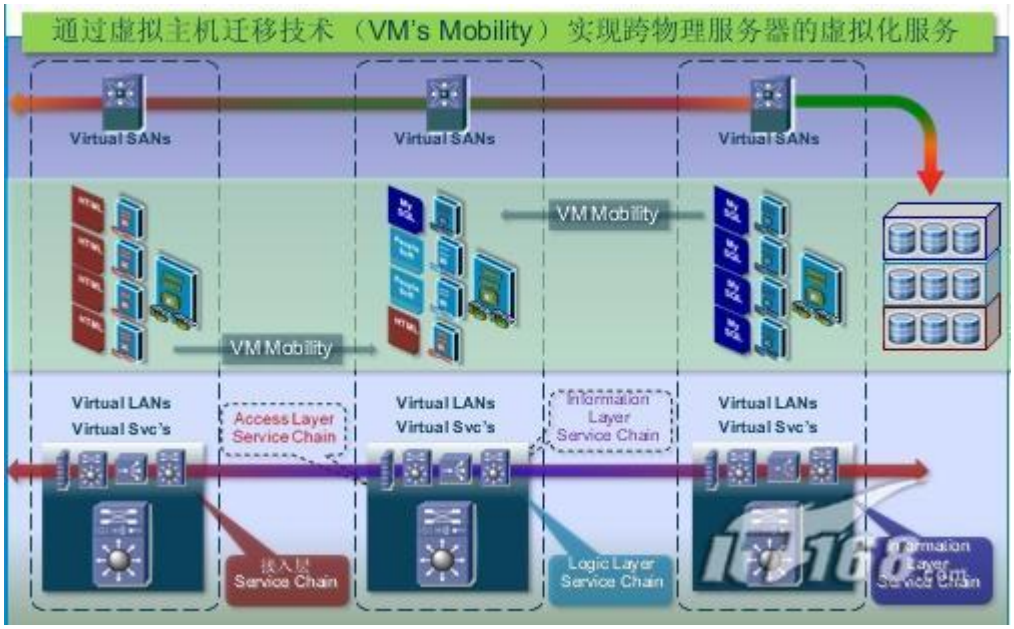

数据中心的资源，包括服务器资源、I/O 资源、存储资源组成一个资源池，通过上层的管理、调度系统在智能的虚拟化的网络结构上实现将资源池中的资源根据应用的需求分配到不同的应用处理系统。虚拟化数据中心可以实现根据应用的需求让数据中心的物理 IT 资源流动起来，更好的为应用提供资源调配与部署。

数据中心虚拟化发展的第一个阶段是通过整合实现服务器和应用的虚拟化服务，这阶段的数据中心也是很多公司已经做的或正要做的。在这一阶段，数据

中心虚拟化实现的是区域内的虚拟化，表现为数据中心的服务如网络服务、安全服务、逻辑服务还是与物理服务器的部署相关联;虚拟机上的 VLAN 与网络交换层上的VLAN对应;存储LUN以类似映射到物理服务器的方式映射到虚拟机。如下图。

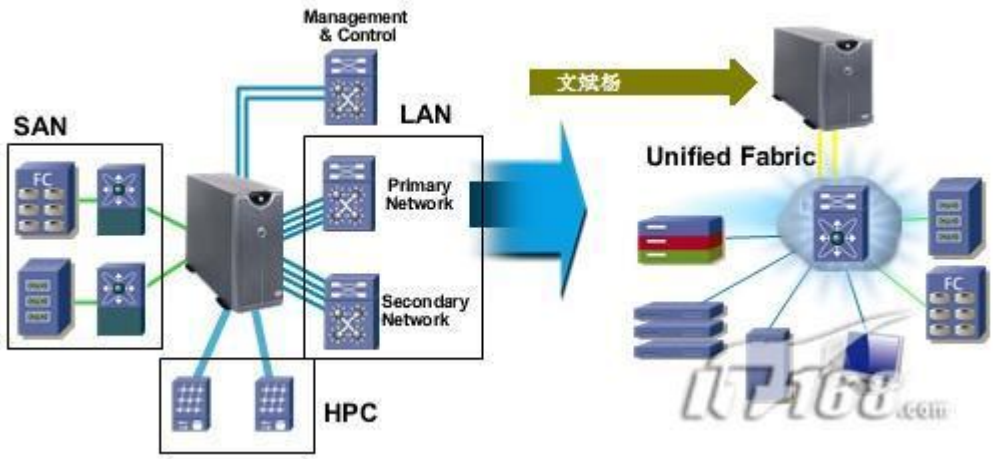



数据中心虚拟化发展的第二个阶段是通过虚拟主机迁移技术(VM's Mobility)实现跨物理服务器的虚拟化服务。如下图。



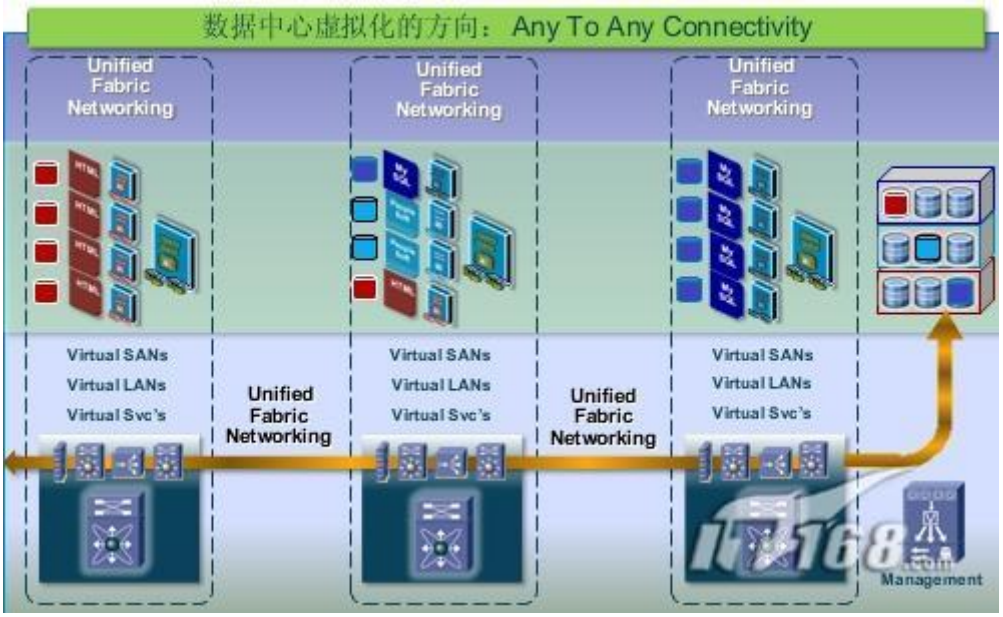

在这个阶段，实现了数据中心内的跨区域虚拟化，虚拟机可以在不同的物理服务器之间切换，但是，为满足虚拟机的应用环境和应用需求，需要网络为应用提供智能服务，同时还需要为虚拟化提供灵活的部署和服务。

思科在下一代的数据中心设计中采用统一交换的以太网架构，思科数据中心统一交换架构的愿景图如下。



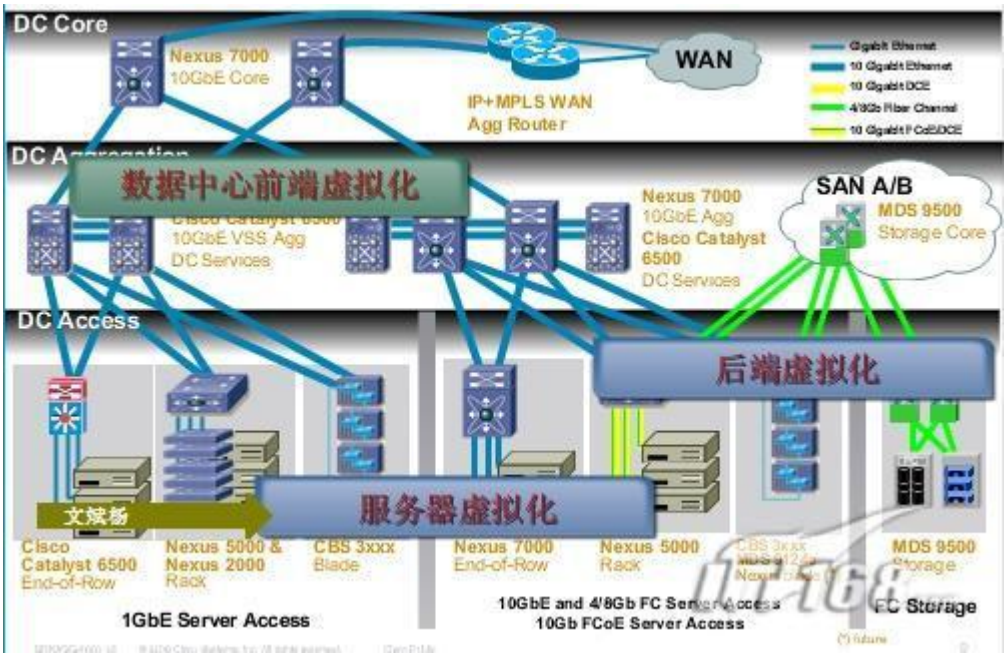

改进之前的数据中心物理上存在几个不同的网络系统，如局域网架构、SAN 网络架构、高层的计算网络架构、管理控制网络架构，各个网络上采用的技术不同，如局域网主要采用以太网技术，SAN 网络主要采用 Fiber Channel 技术。而在思科的下一代统一交换架构数据中心中，数据中心的服务器资源、存储资源、网络服务等都通过统一的交换架构连接在一起，数据中心只有一个物理网络架构，可以实现动态的资源调配，提升效率和简化操作。

统一交换架构下数据中心的虚拟化如下图。




统一交换架构下数据中心虚拟化简化了数据中心的管理和运维，实现了真正的任意 IT 资源之间的灵活连接，实现了统一的 I/O，在统一的 I/O 上可以实现最新的万兆网、无丢失(FCoE)、低延时的数据中心以太网技术。统一交换架构下数据中心虚拟化为未来的进一步的虚拟化和基于云计算的数据中心提供了平台。

数据中心虚拟化架构包括数据中心前端虚拟化、服务器虚拟化、数据中心后端虚拟化。如下图。

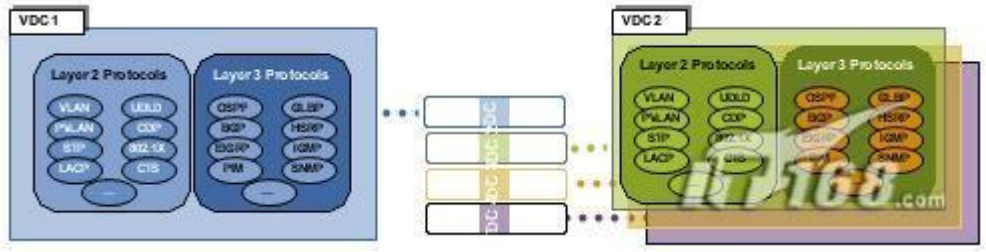



思科设计数据中心时采用分层、分区的设计方式，层次设计包括核心层、汇聚层、接入层，接入层的不同功能的服务器位于不同的区域，服务器经过每个区域的汇聚层连接到核心层。数据中心前端虚拟化是指对服务器网络接口之前的数据中心基于以太网的网络架构的虚拟化。服务器虚拟化指在一台物理服务器上为多个应用需求实现多个虚拟机，并且实现区域内资源的动态调配、迁移，服务器虚拟化技术的实现需要网络的支持配合。数据中心后端虚拟化指通过虚拟化技术将服务器和存储资源更好的调配使用起来。

下面按数据中心层次化设计中的核心层、汇聚层、接入层依次介绍思科在前端虚拟化上的最新的一些技术实现。

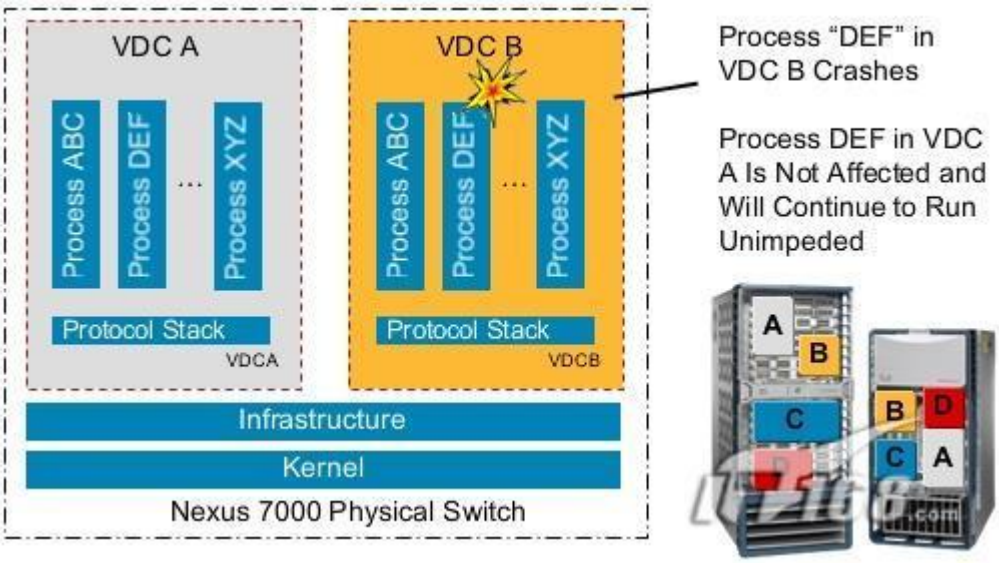
思科数据中心核心层虚拟化技术。

思科为数据中心级和园区骨干网级网络提供了 Nexus 交换机网络技术和 Nexus 系列产品。Nexus 系列产品中采用了 VDC(Virtual Device Content)技术，可以将一台物理交换机逻辑上模拟成多台虚拟交换机，如下图模拟出的两个虚拟交换机 VDC1 和 VDC2。



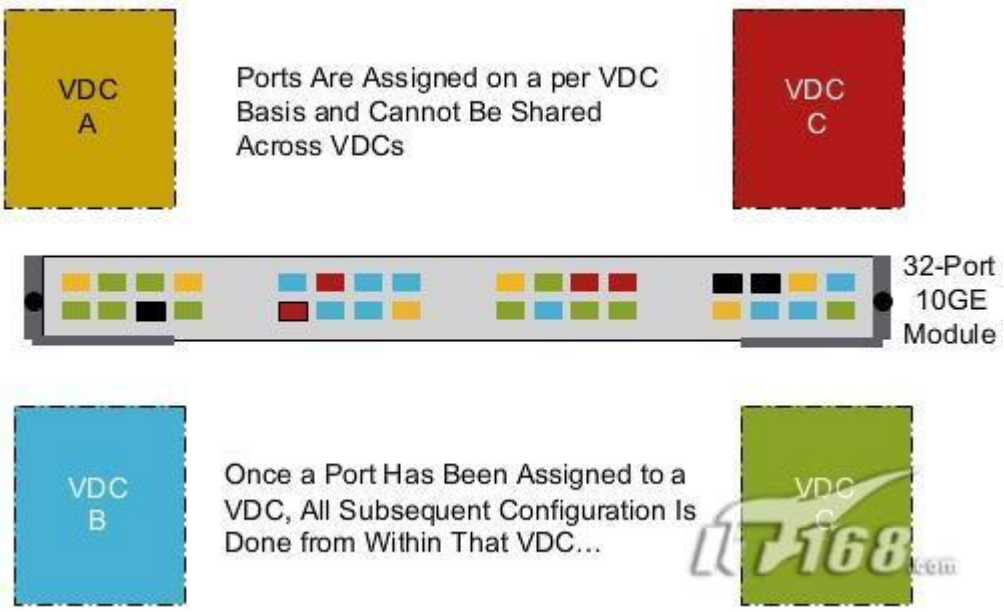
VDC 技术可以实现每个模拟出的 VDC 都拥有它自身的软件进程、专用硬件资源(接口)和独立的管理环境，可以实现独立的安全管理界限划分和故障隔离域。VDC 技术有助于将分立网络整合为一个通用基础设施，保留物理上独立的网络的管理界限划分和故障隔离特性，并提供单一基础设施所拥有的多种运营成本优势。

VDC 可以实现故障域隔离，如下图。



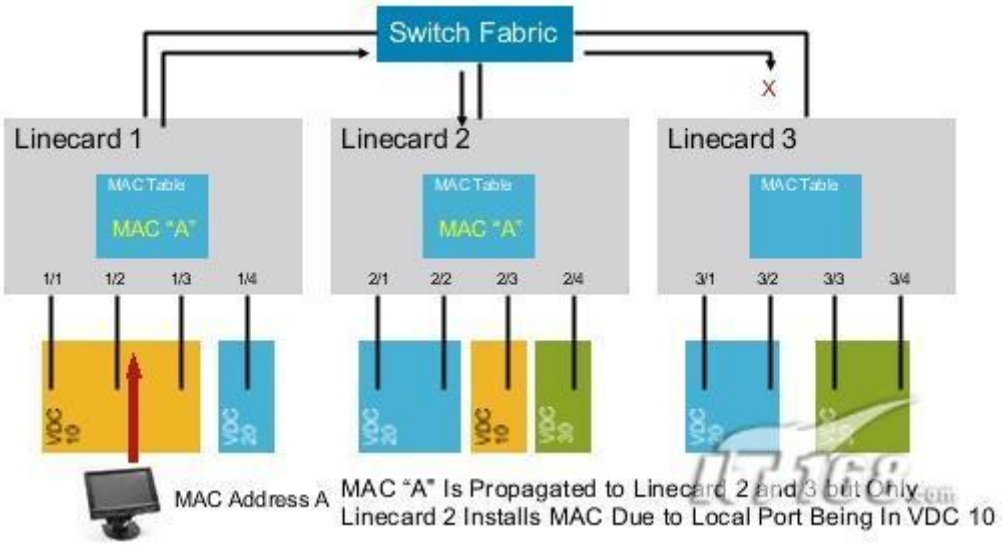

一个 VDC 为所有的运行在它上面的进程建立故障隔离域，如图中 VDC B 中的“DEF”进程发生故障后不会影响到 VDC A 中的“DEF”进程。

VDC 的端口分配如下图。



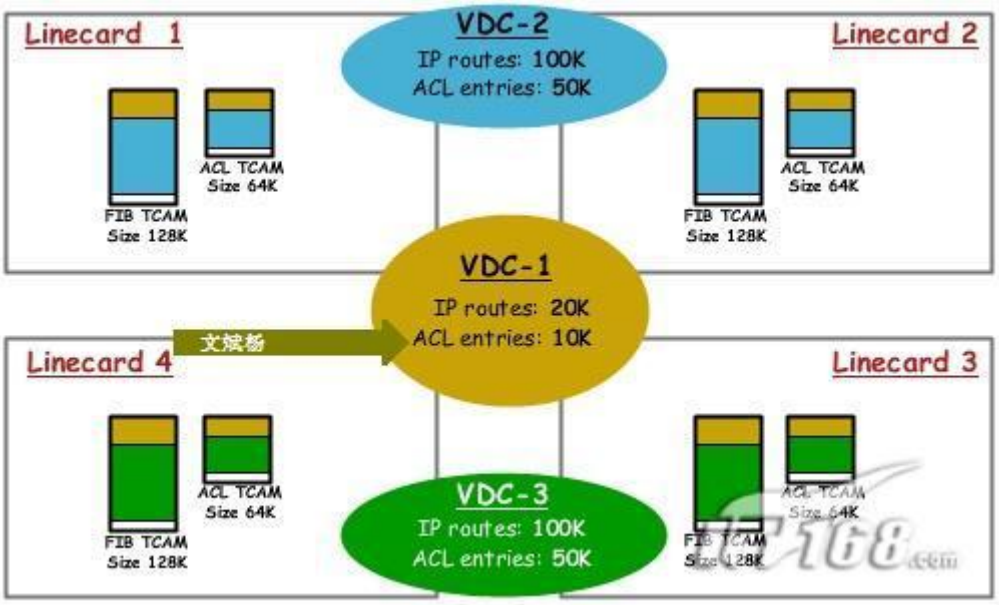
端口分配到各 VDC 后不能在 VDC 之间共享，一旦某一端口分配到一个 VDC，就只能在那个 VDC 中对该端口进行配置。(*注 上图中右下角应为 VDC D)。

VDC 资源使用示意图一如下。



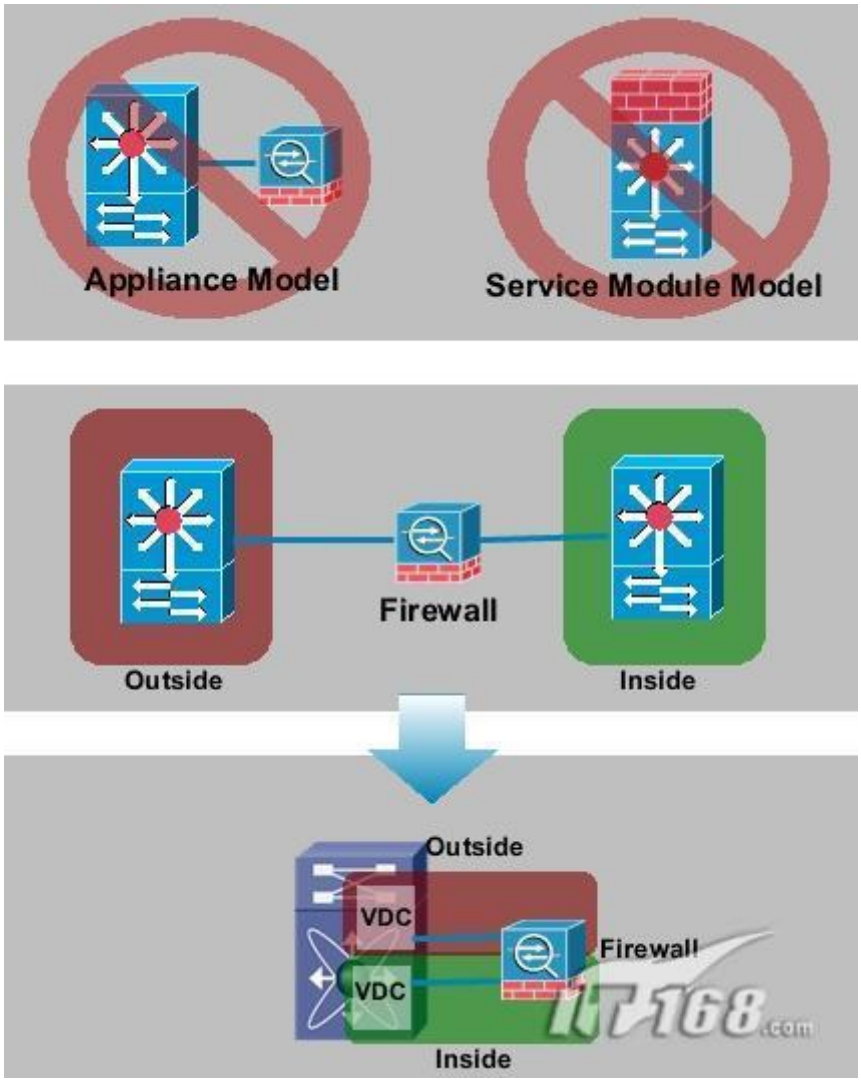

在一个 VDC 中的 MAC 地址只会广播到有接口分配到该 VDC 上的 Linecard 地址表中，如图中的 Linecard 1 和 Linecard 2，它们都有接口分配到 VDC 10。

VDC 资源使用示意图如下。



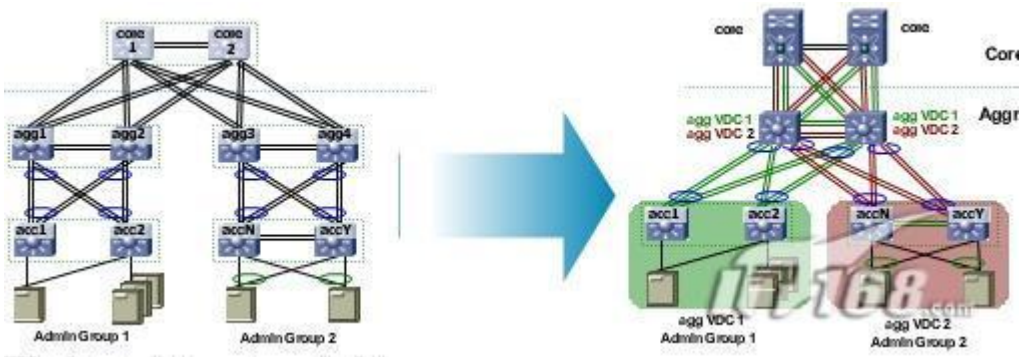

Linecard 1 和 Linecard 2 都给 VDC-2 分配了 100k 的 FIB TCAM 和 50k 的 ACL TCAM 资源。

VDC 安全分区技术应用如下图。



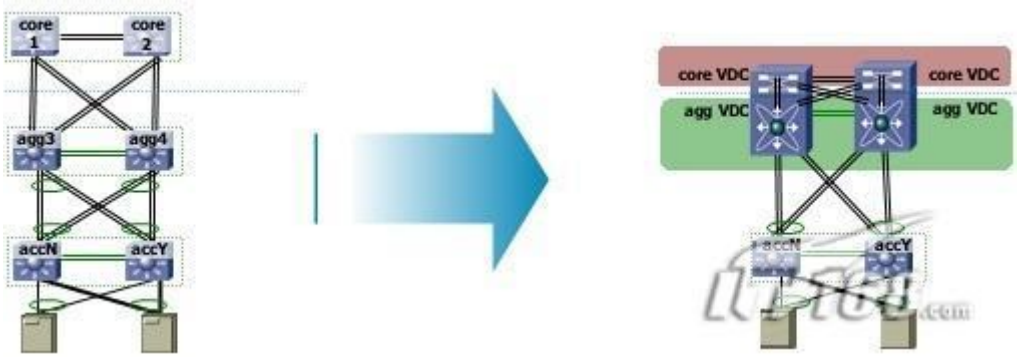

最上面的两个模型是现在应用的比较多的，但用户可能希望能实现中间的内外分明的架构，这可以通过最下面的两个 VDC 的应用来实现，其中一个 VDC 为 Outside，另一个为 Inside，之间通过防火墙连接。每个 VDC 可以运行独立的转发机制、路由机制、管理机制和登录机制。

VDC 可以实现数据中心设计的水平整合，如下图。

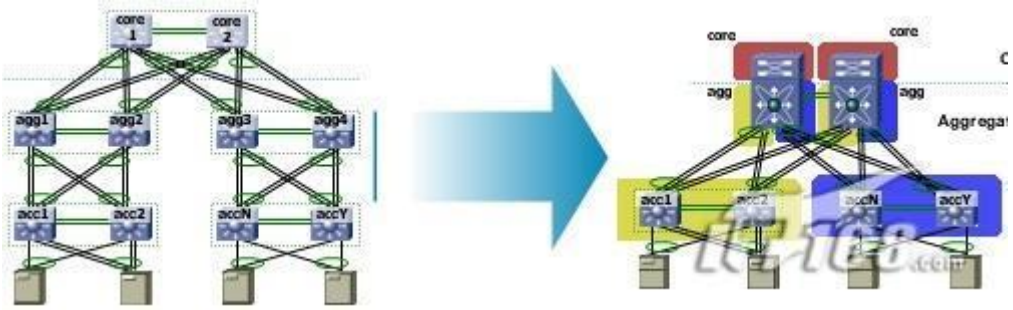


当两个汇聚区不是很大时可以通过将一个 Nexus 设备划分为 VDC 1 和 VDC 2 来分别代替实现两个汇聚区。

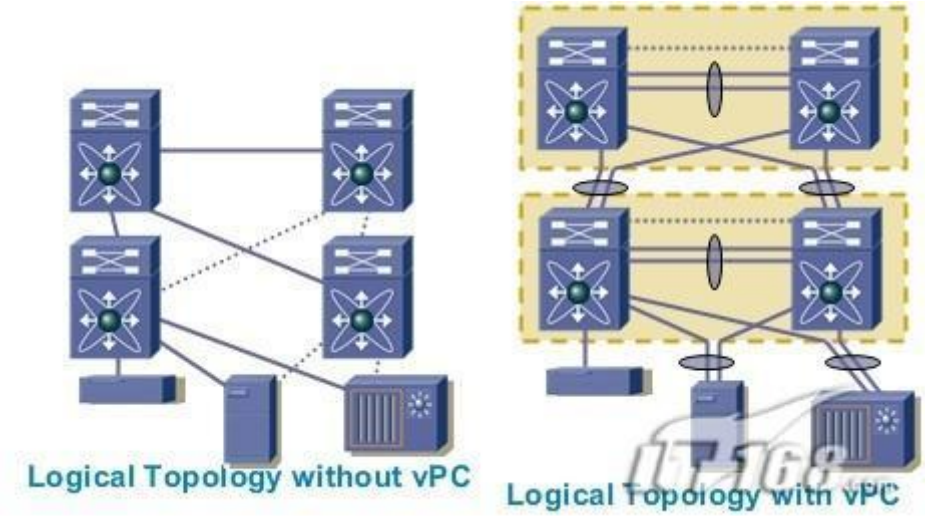
VDC 实现数据中心设计的垂直整合，如下图。一个 Nexus 设备模拟成 Core VDC 和 Agg VDC 分别实现核心层和汇聚层功能。




VDC 实现数据中心设计水平和垂直的综合的整合应用，如下图。



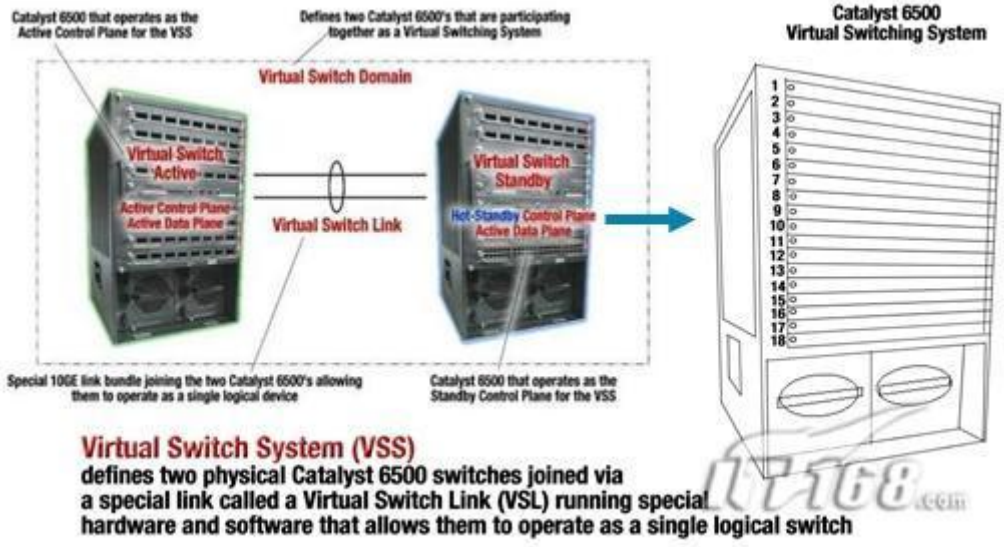

vPC(virtual Port-Channel)技术。下图是传统的和使用了 vPC 技术的交换机互联逻辑拓扑图。vPC 技术可以在 Cisco Nexus 7000 系列产品上实现。



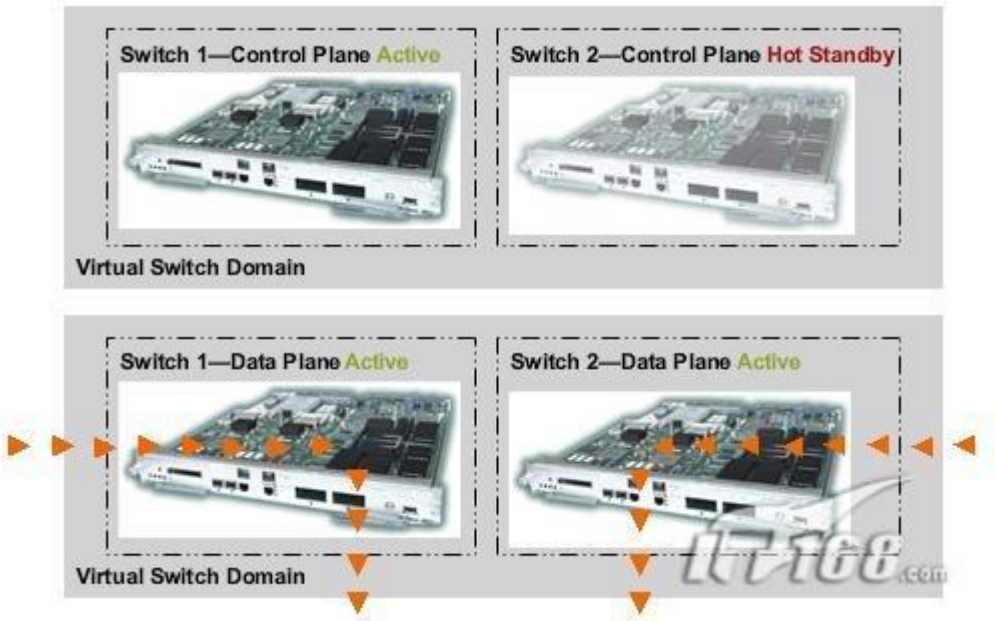
传统的技术实现交换机互联时，如果互联结构中存在环路，则会 block 环路中的部分支路。vPC 技术可以实现在单个设备上使用 port-channel 连接两个上行交换机，完全使用所有上行链路的带宽，并消除 STP blocked ports，在 link/device 失效下提供快速收敛。

VPC 设计和传统设计相比的优势如下图。

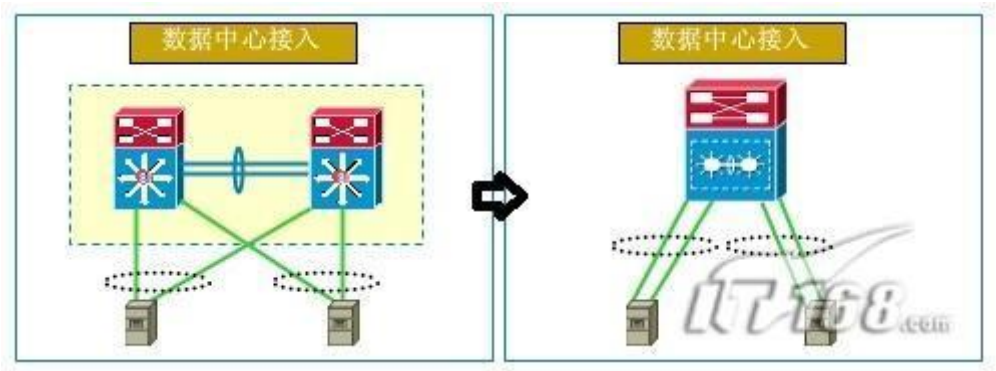

虚拟交换系统 VSS(Virtual Switch System)。两台 Cisco Catalyst 6500 系列交换机通过 VSS 连接后可以实现如同操作单一逻辑交换机的效果。如下图。



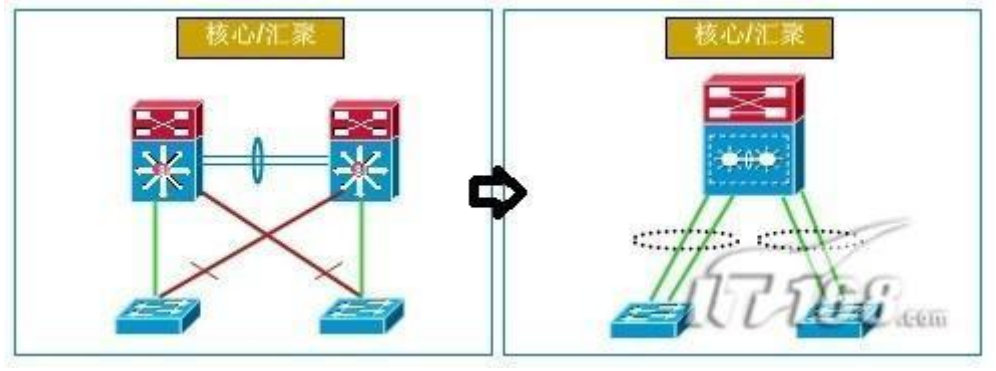

虚拟交换系统 VSS 与 vPC 技术有一些不同的地方，如在控制层面上两个交换机有主次之分，但在数据处理上是双活的。




6500-VSS 应用于数据中心接入：不再需要复杂的、难于诊断的 STP;可以简化管理，实现一个管理点，一个路由和 STP 节点;系统总带宽提升至 1.4Tbps。

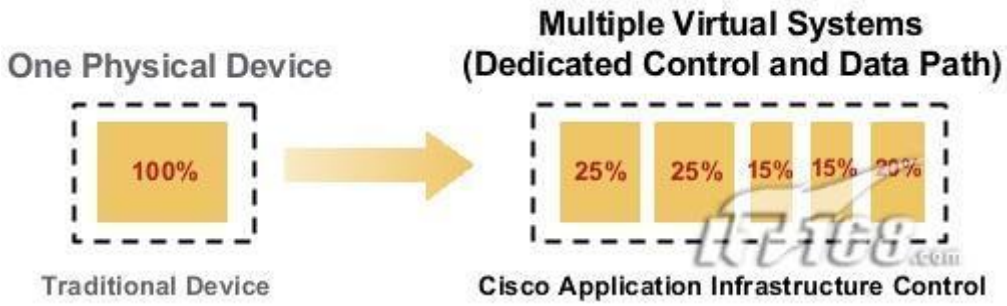



6500-VSS 应用于核心/汇聚层：实现网络系统虚拟化;提供机箱间的状态化切换(SSO)，改进无中断通信，切换时间〈200ms;跨机箱 EtherChannel，优化路径选择。

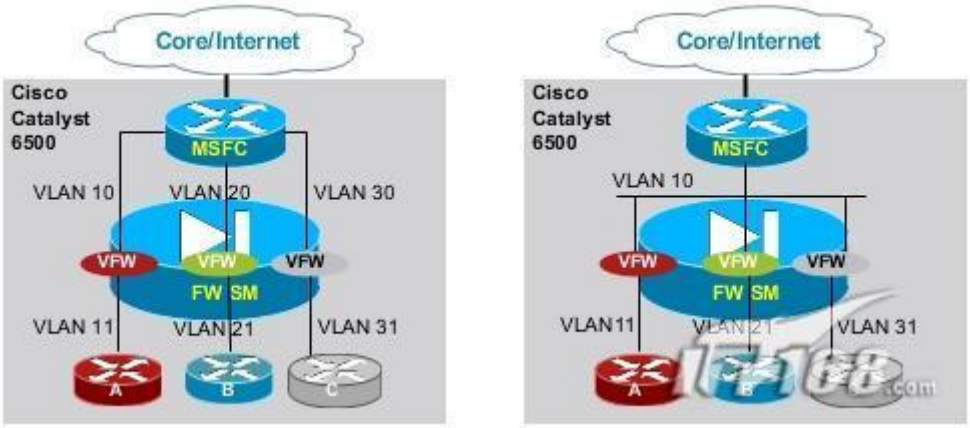



VSS 和 vPC 技术比较如下图。

ACE 模块，实现服务器负载均衡和 SSL，可以将一个物理设备虚拟成不同的功能区域，虚拟的功能区有独立的配置文件、路由表、应用规则设置等。

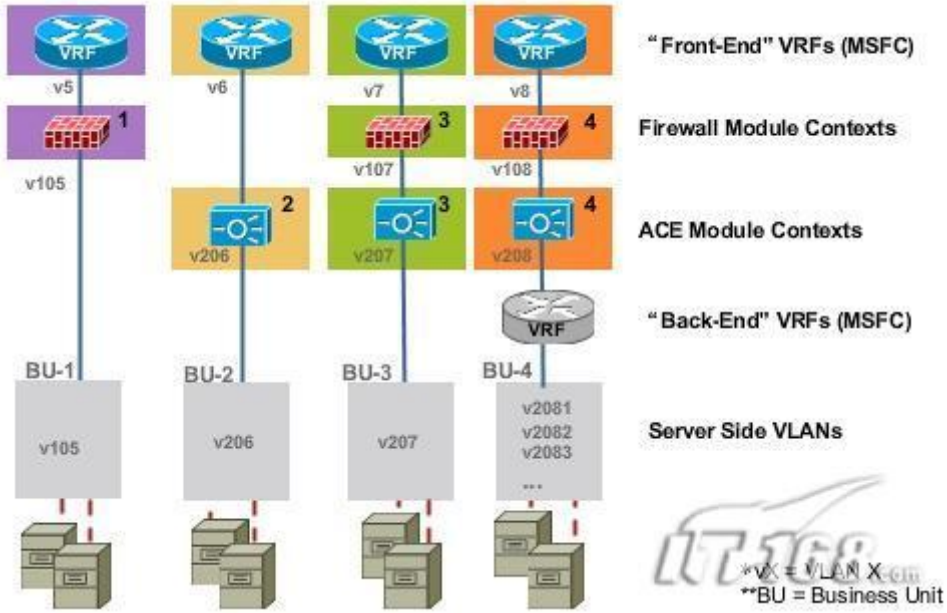



防火墙模块 FWSM，可实现最多 250 个虚拟防火墙。

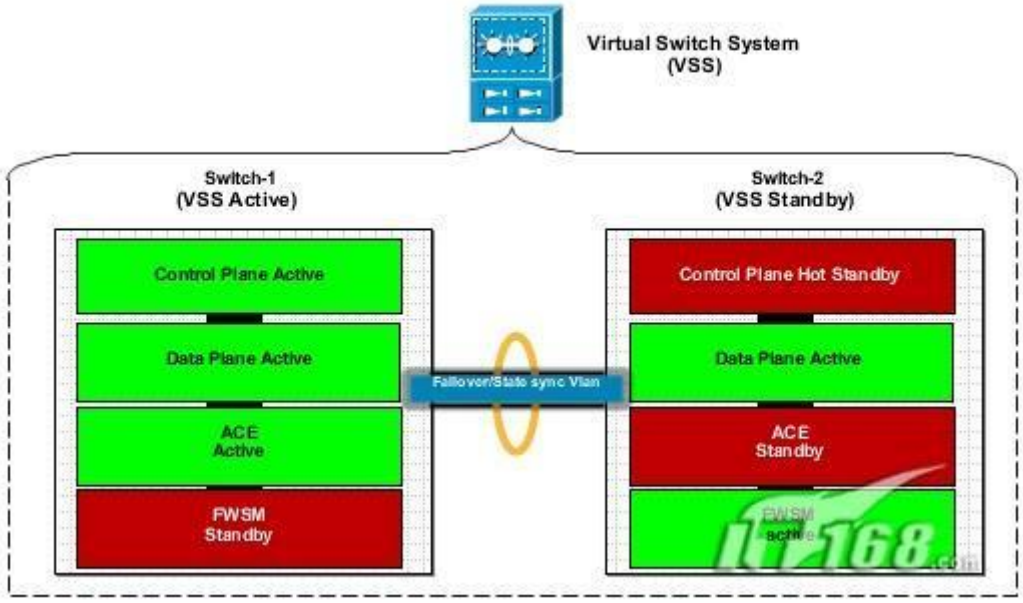



虚拟局域网 VLANs 需要时可以共享，如上图左边的 VLAN 10，各虚拟防火墙可以有各自的策略设置。

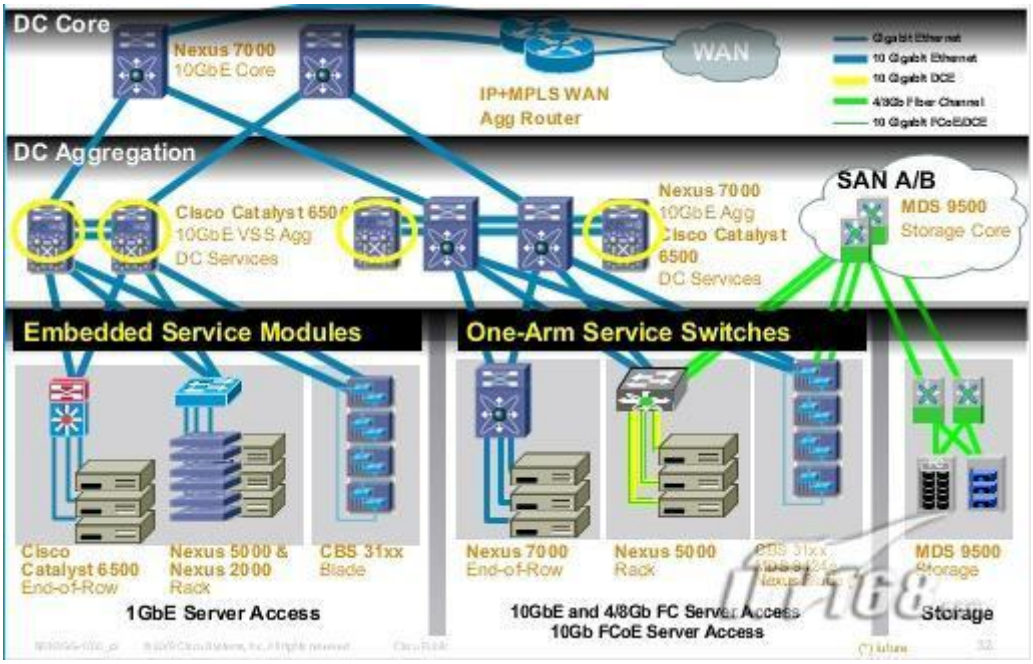
虚拟技术联合应用示例如下图。




VSS 技术框架下 ACE 和 FWSM 模块设计示例如下图。



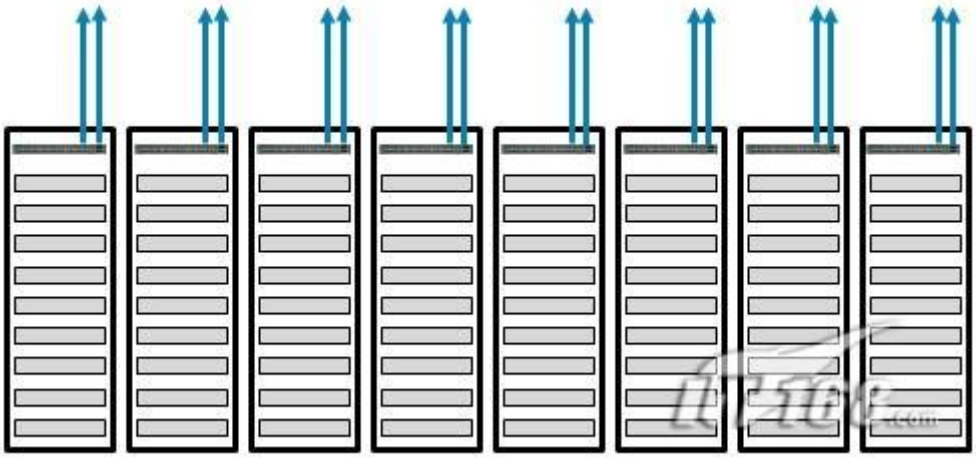

汇聚层网络服务的部署设计。



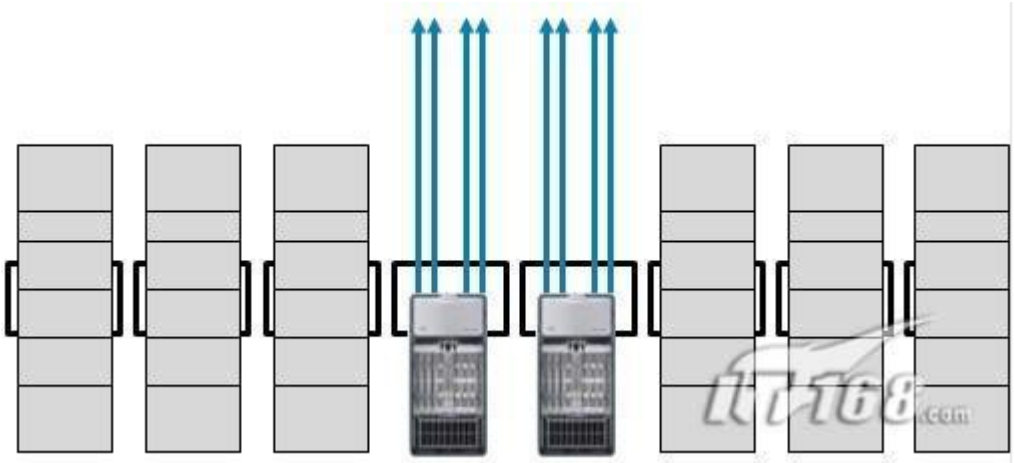

| | | VSS on Catalyst 6500 | vPC |
|---|---|---|---|
| **Availability** | Software | 12.2(33)SXH1 | N7K - 4.1 (Q1CY09)<br>N5k - 4.1 (H2CY09) |
| **General** | Multi-chassis Etherchannel (Active-active) | Yes | Yes |
| | Load-Balancing at L2/L3 | Yes | Yes |
| **Control Plane/ HA** | Control Plane | Unified | Independent |
| | Configuration Files | Unified | Independent |
| | Supervisor Redundancy | Single sup (redundancy across chassis) | Redundant supervisors per chassis |
| **L2** | Link Agg Protocol | LACP, PaGP(+) | LACP |
| | STP Required | No | No |
| **L3** | Single Logical Gateway | Yes (No Need for FHRP) | Yes, active-active HSRP |
| | Routing Instance | Single | Independent |
| | Routing Peers | Reduced | Same as before |

数据中心接入层的可选设计有 Top of Rack(架顶)和 Middle of Row(列头)，如下。

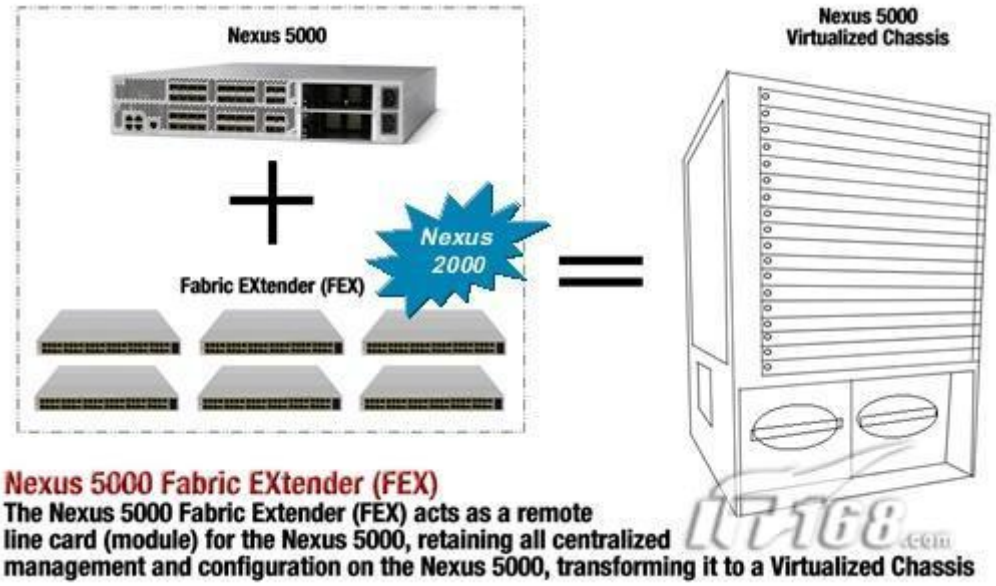
架顶式，一个机柜中一台或两台交换机连接 1-RU(1 机架单元，如 20 台)服务器。



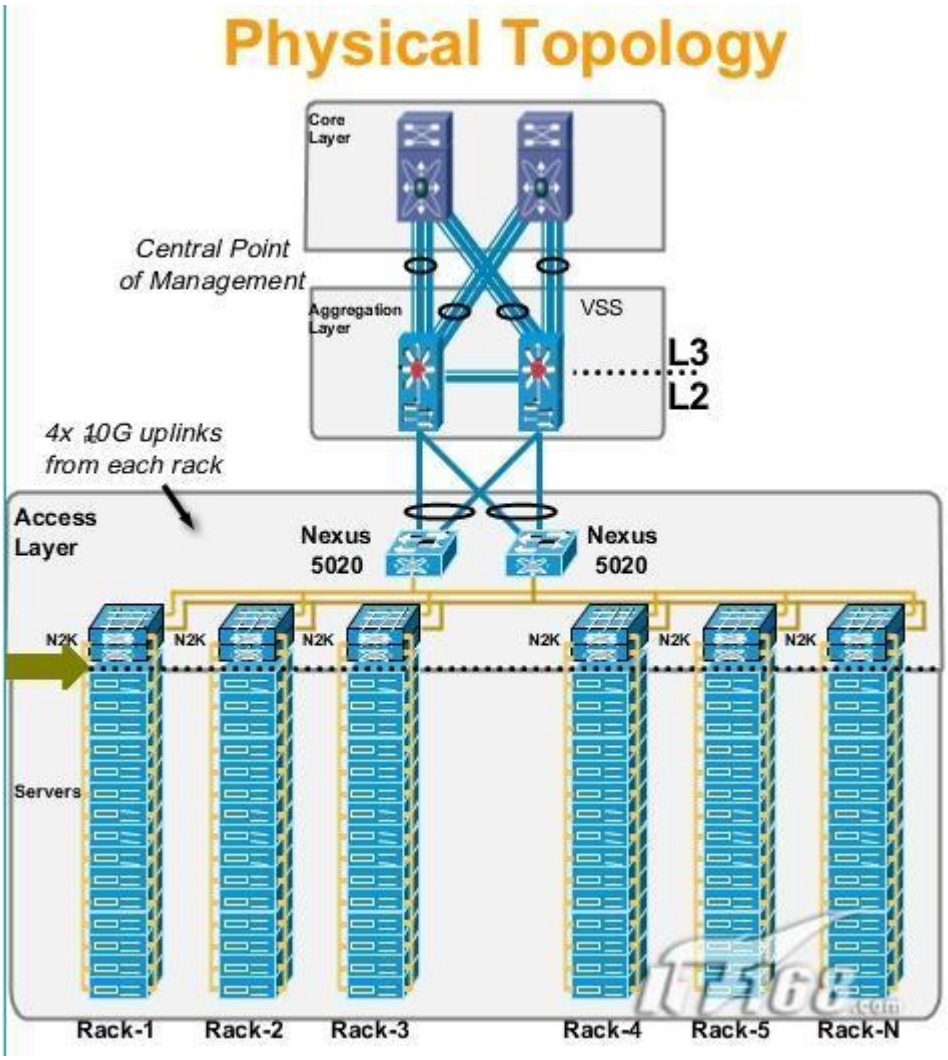
列头式，可用于服务器占用空间较大时，交换机集中放置。



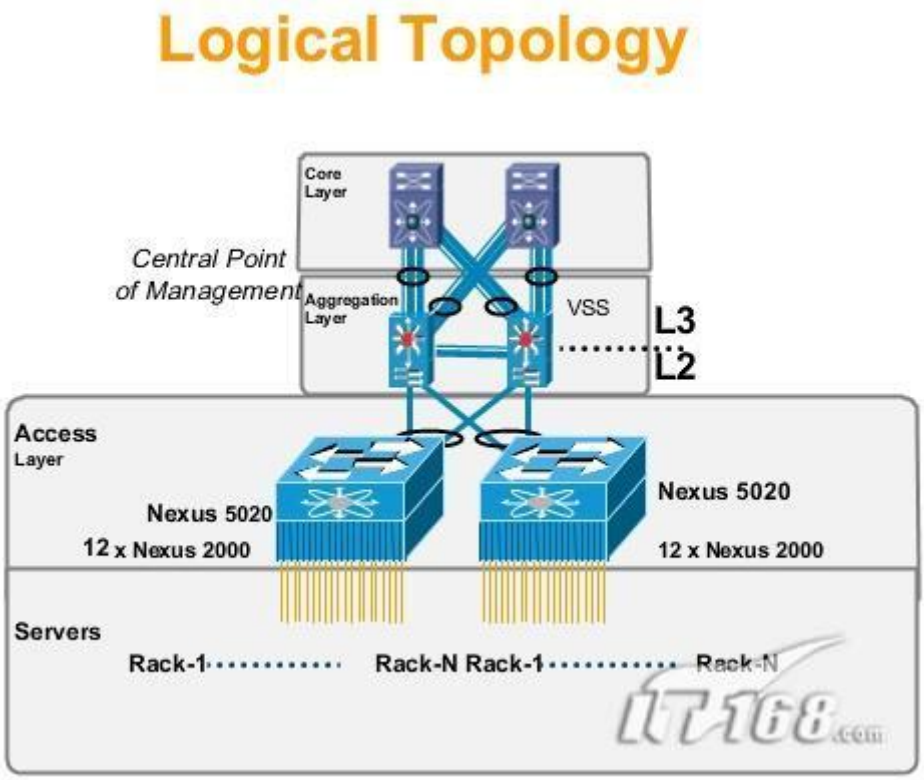
采用 Nexus 5000 做控制中心可控制多台远端 Nexus 2000，联合使用效果图如下。



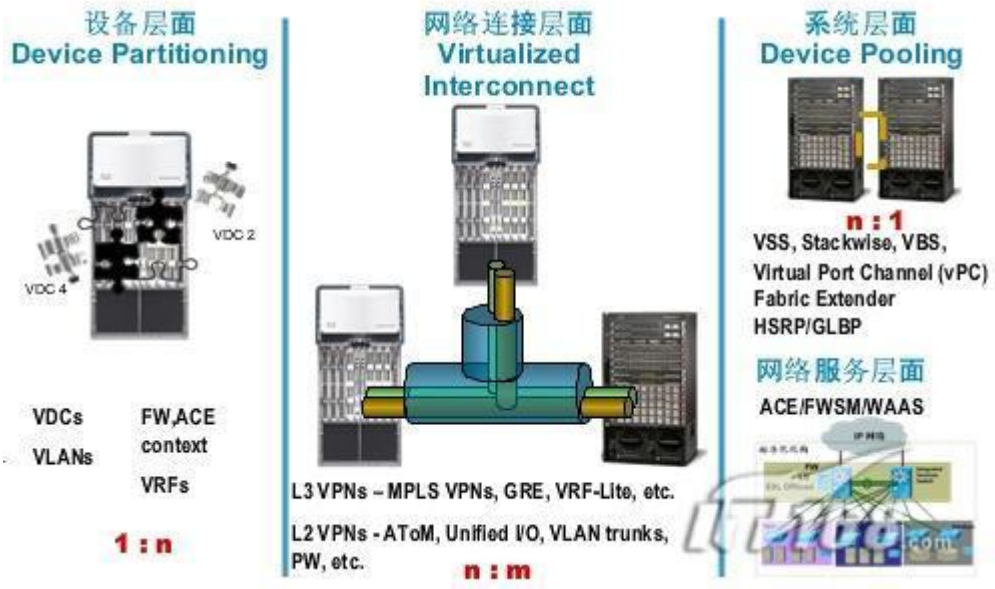

部署 Nexus 2000 的物理拓扑结构。



逻辑拓扑结构。

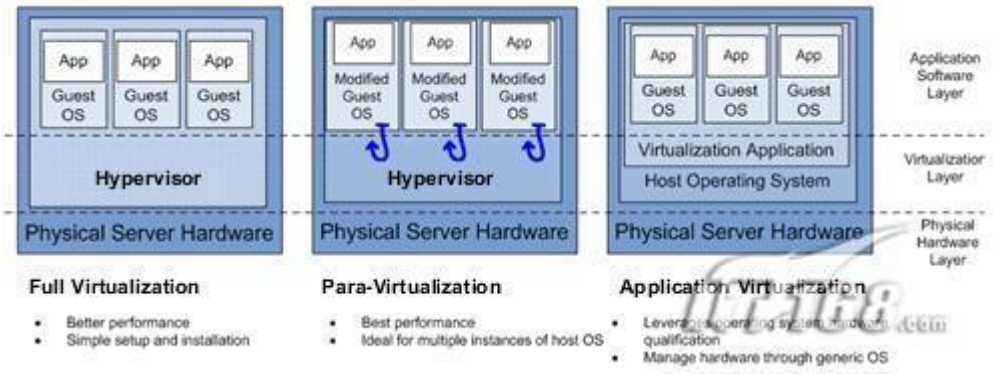



前端网络虚拟化技术各层面的简单汇总如下。




服务器虚拟化

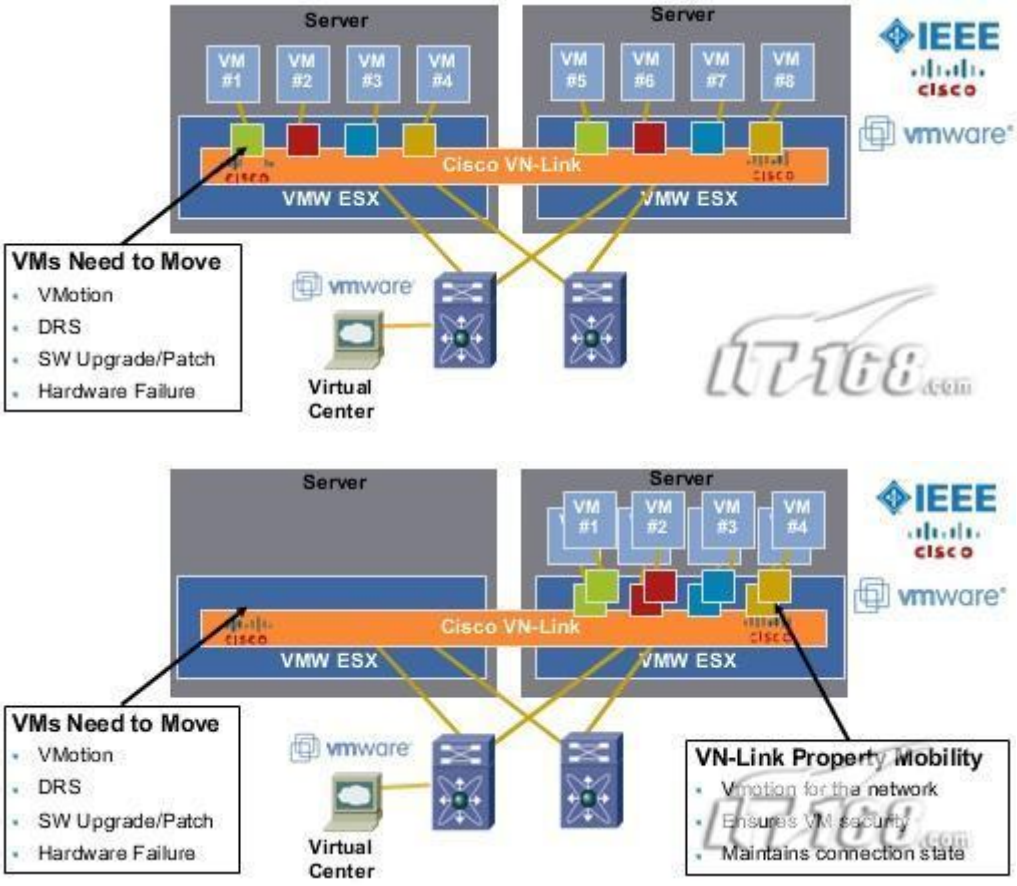
服务器虚拟化有全部虚拟化、部分虚拟化、应用虚拟化三个发展需求，如下图。




在服务器 VMotion(虚拟机迁移)过程中存在以下的问题：VMotion 可以跨网络动态迁移虚拟机，管理策略上如何适应;无法察看本地交换流量和为其设定策略;无法识别一条物理链路上多个虚拟机的流量。

Cisco VN-Link 技术针对这些问题实现：将网络延伸到服务器虚拟机;提供一致的连接服务;协调、统一的管理。VN-Link 将网络交换延伸到服务器虚拟机，在服务器虚拟机的迁移过程中网络信息将伴随迁移。

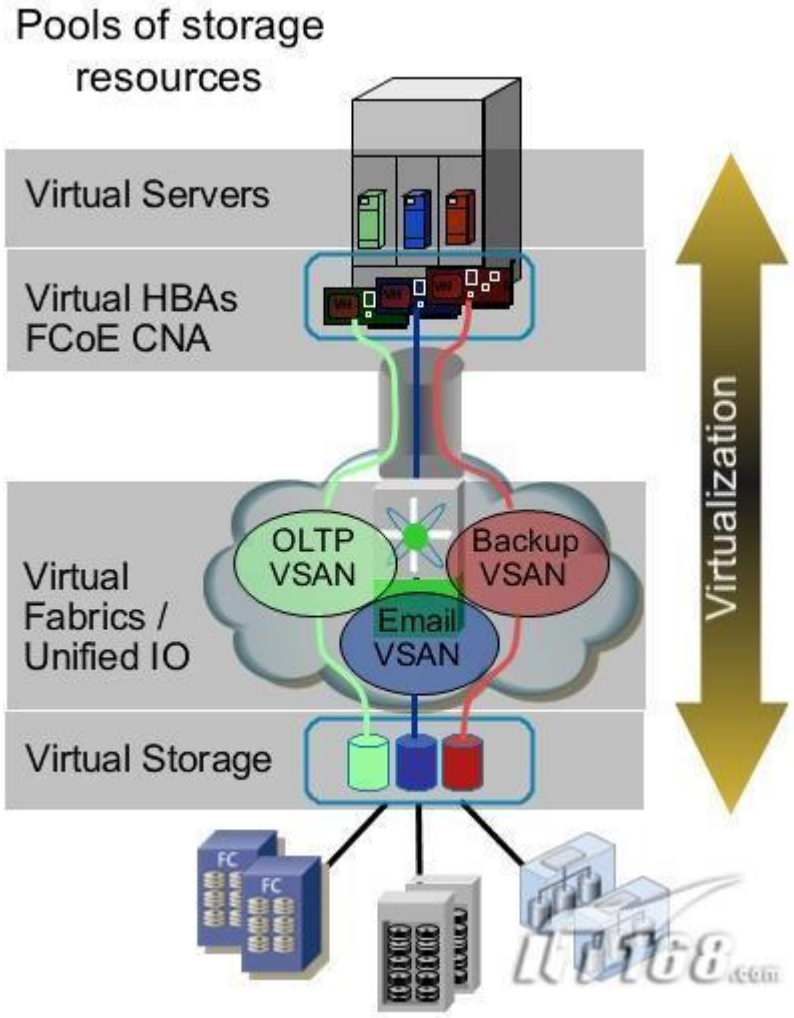
VN-Link 技术实现基于策略的虚拟机连接、网络与安全技术的移动、不间断的运行模式，示意图如下。



在虚拟机资源的迁移中，需要一起移动 DRS，SW 升级文件/补丁，硬件错误记录等。VN-Link 能实现网络的虚拟迁移，虚拟机的安全防护，保留之前的连接状态等。

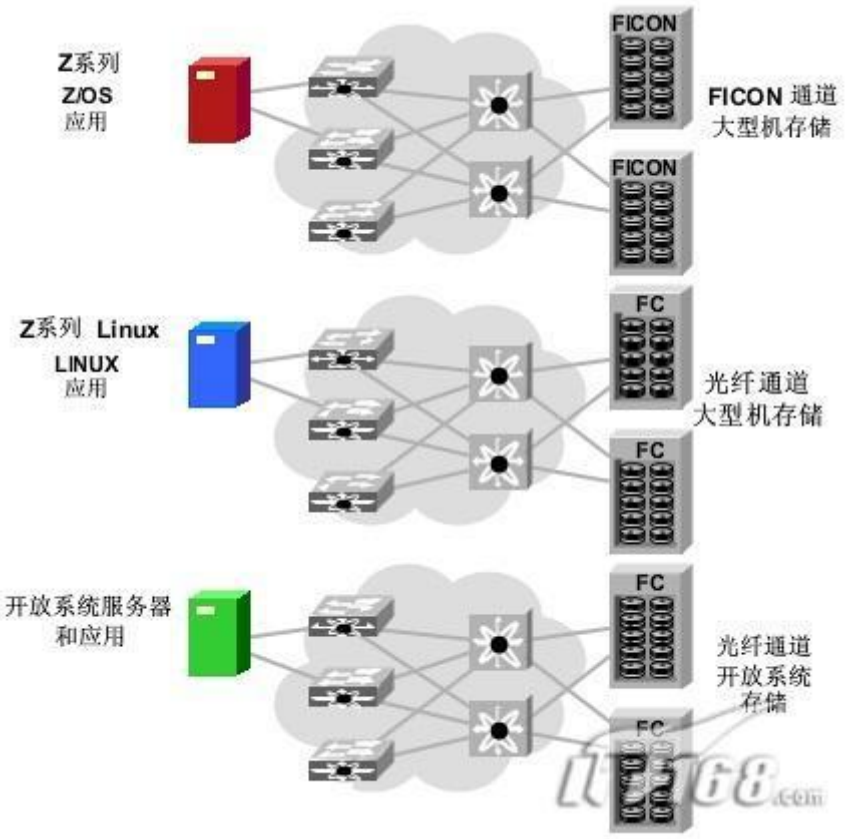
Cisco Nexus 1000V 是思科最新的基于 VN-Link 技术的软件，也是业界首个第三方软件交换机，提供 VN-Link 特性，确保在 VMotion 过程中虚拟机的可视性和连通性。

后端虚拟化的主要内容如下图。包括虚拟化服务器、HBAs，统一输入输出/Fabrics、存储等。



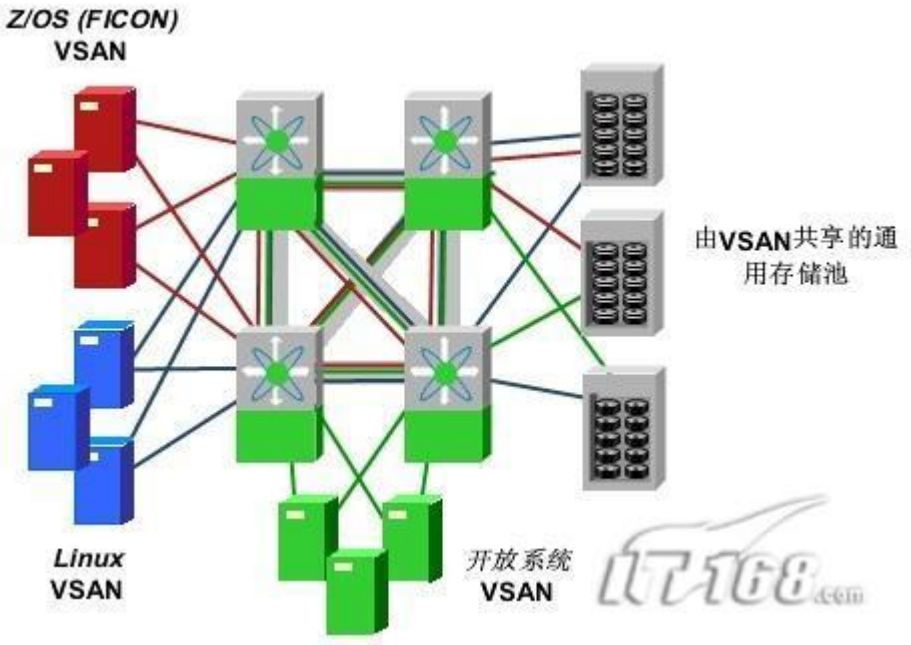

后端虚拟化可以优化资源的使用、增加灵活性和敏捷能力、简化管理、减少 TCO。

传统的基于应用/部门的 SAN 存储网络的架构如下图。



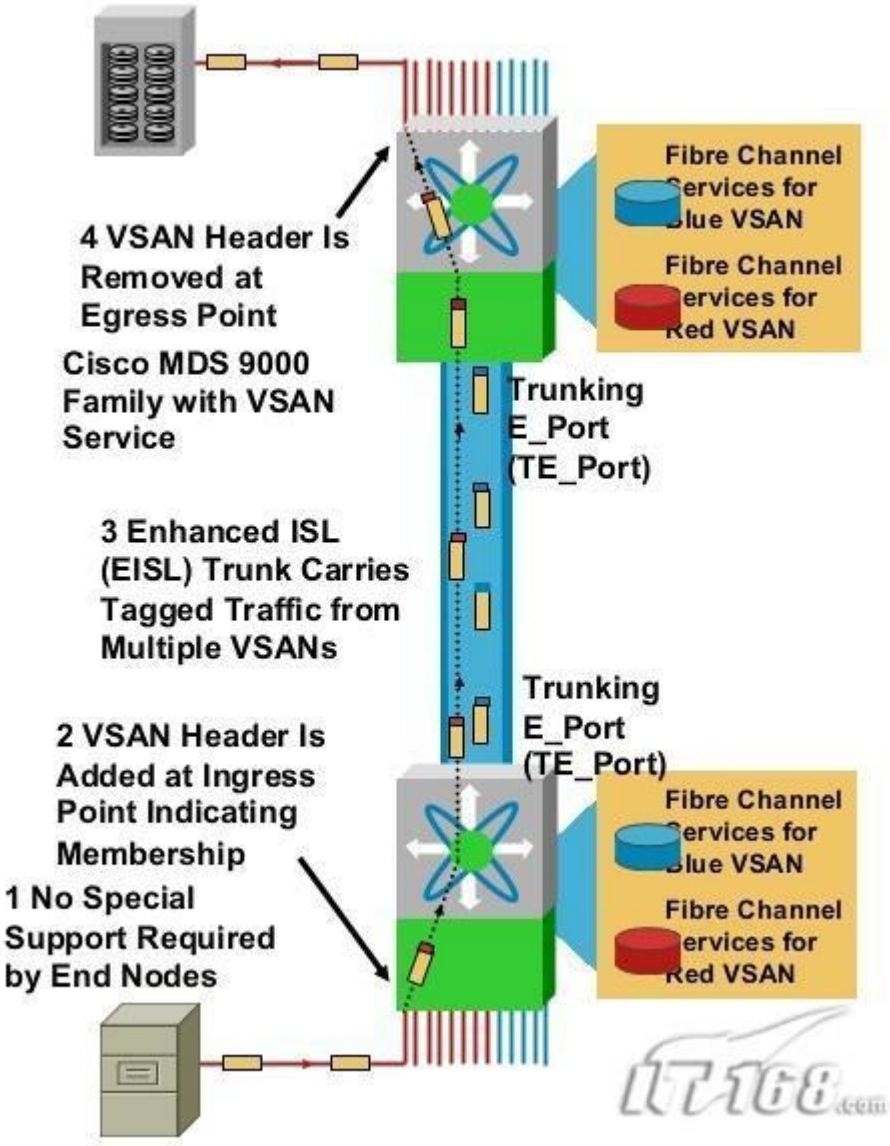

总体上看各个 SAN 网络如一座座的孤岛，每个岛上的端口都过量，需要管理大量的交换机，并且他们的资源无法共享。

整合的 VSAN 存储网络架构如下图。



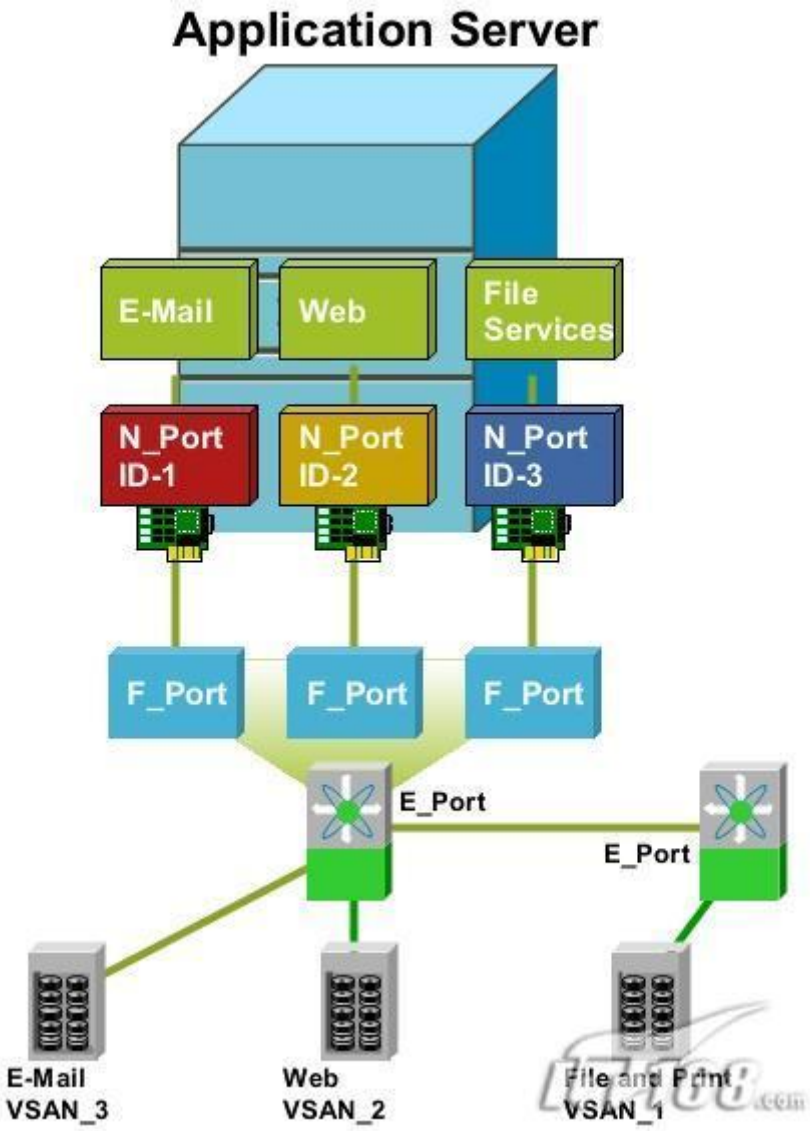
它是一个供所有应用系统使用的公用存储网络，能够做到：最大化端口利用率，无需多余扩展端口;减少整体交换设备数量，降低管理复杂度;更灵活的配置存储资源，提高资源利用率;为未来的存储虚拟化打下基础。

VSAN 技术的工作原理和流程如下。



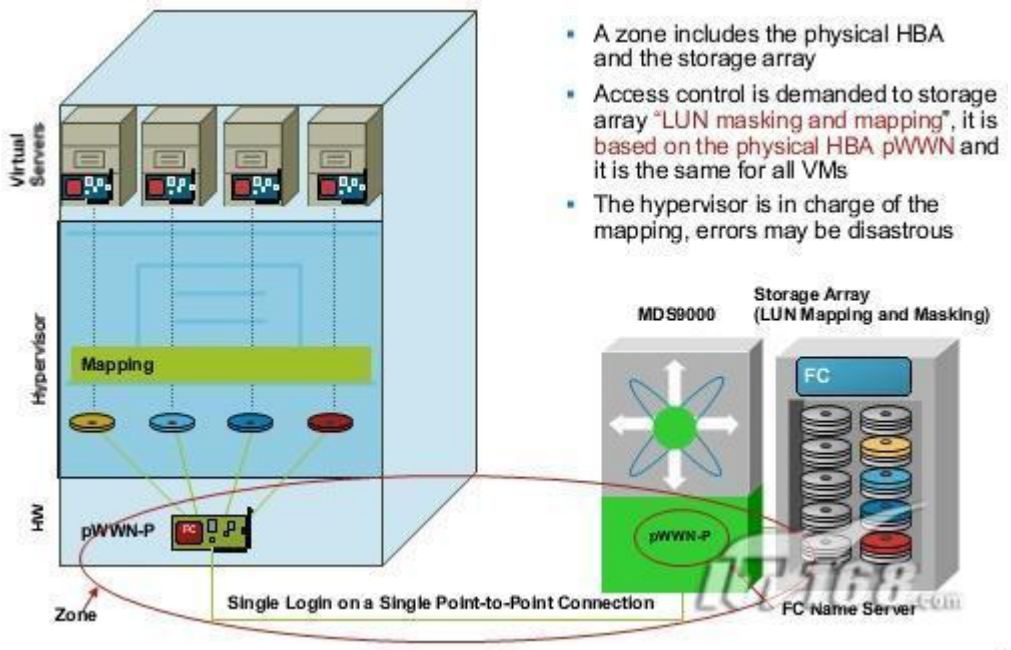

VSAN 实现不需要端设备的特殊支持，SAN 交换机会在输入输出数据时添加和去除表明数据属性的标签，实现基于硬件层面上的不同 VSAN 流量的隔离，SAN 交换机中的 Fibre Channel 为不同的 VSAN 数据提供服务，如图中的蓝色和红色。

N-Port ID 虚拟化技术(NPIV)能将一个 HBA 卡接口虚拟成多个接口，相应的，在交换机上也有多个对应接口，如下图。



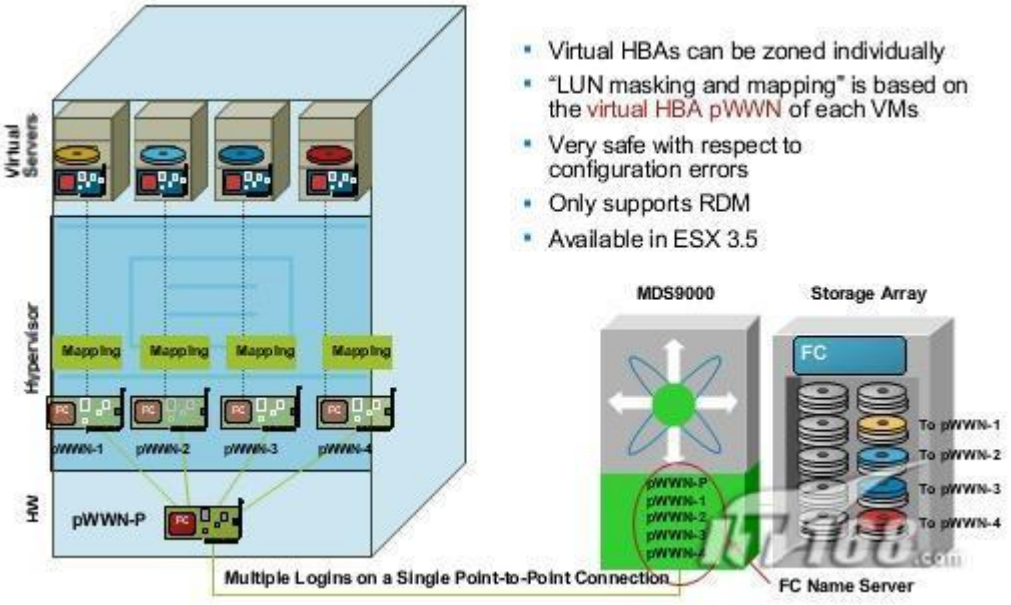

虚拟的 N-Port 能为同一个 VSAN 的不同数据(如 E-Mail,Web)提供独立接口，并且可以在应用层面上实现对各虚拟接口的存取控制、域控制、端口安全控制等。目前，N-Port ID 虚拟化技术是为同一个 VSAN 的需求虚拟出多个 N-Port 接口。

一个未使用 NPIV 技术的由虚拟服务器到存储的网络连接图如下。



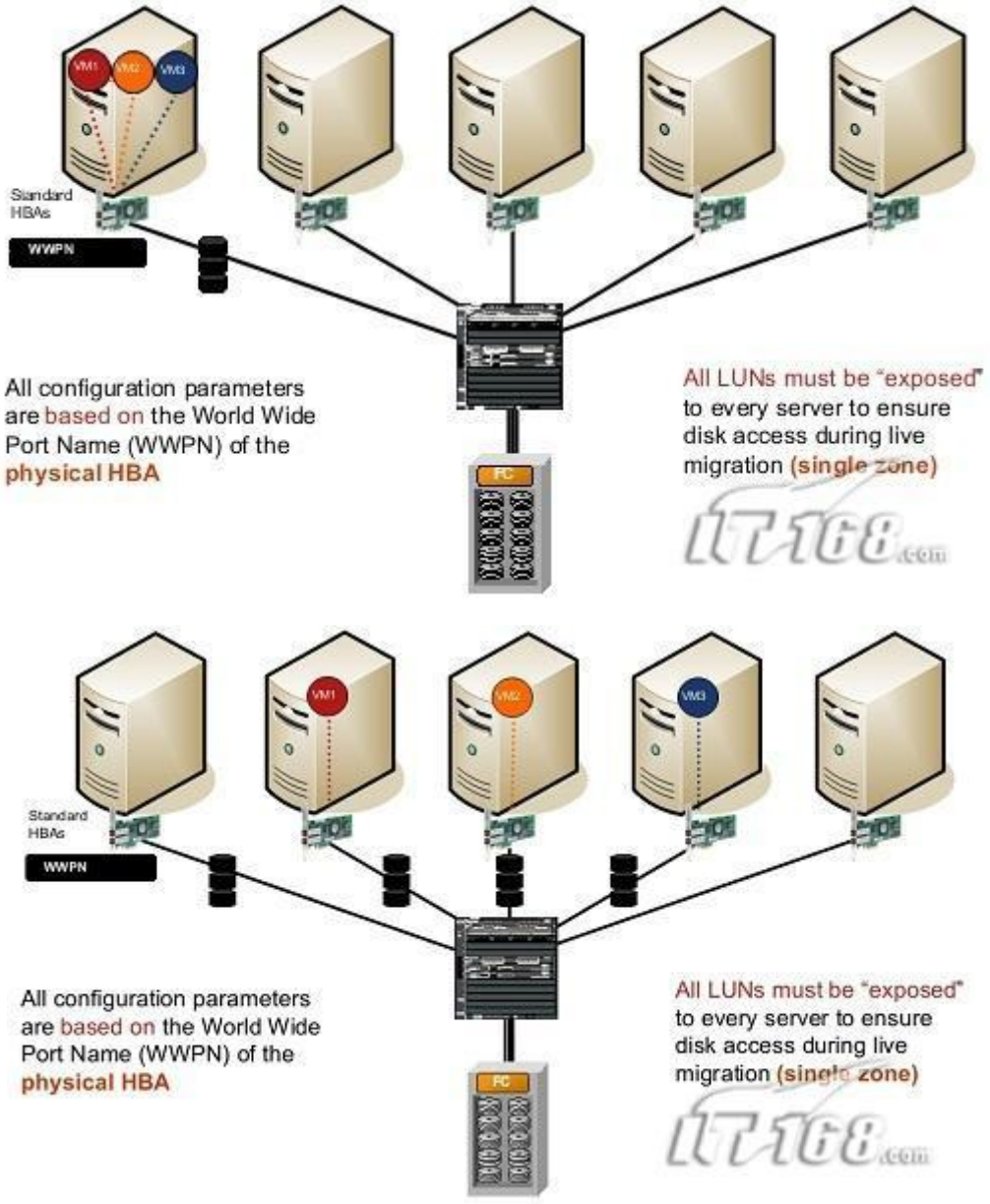

HBA 卡到交换机只有一个点到点连接，MDS S9000 交换机中只有一个 WWPN(World Wide Port Name)，所有的存储卷需要基于同一个物理 HBA 的存取控制。管理程序(hypervisor)负责映射和错误处理。

使用 NPIV 技术后。



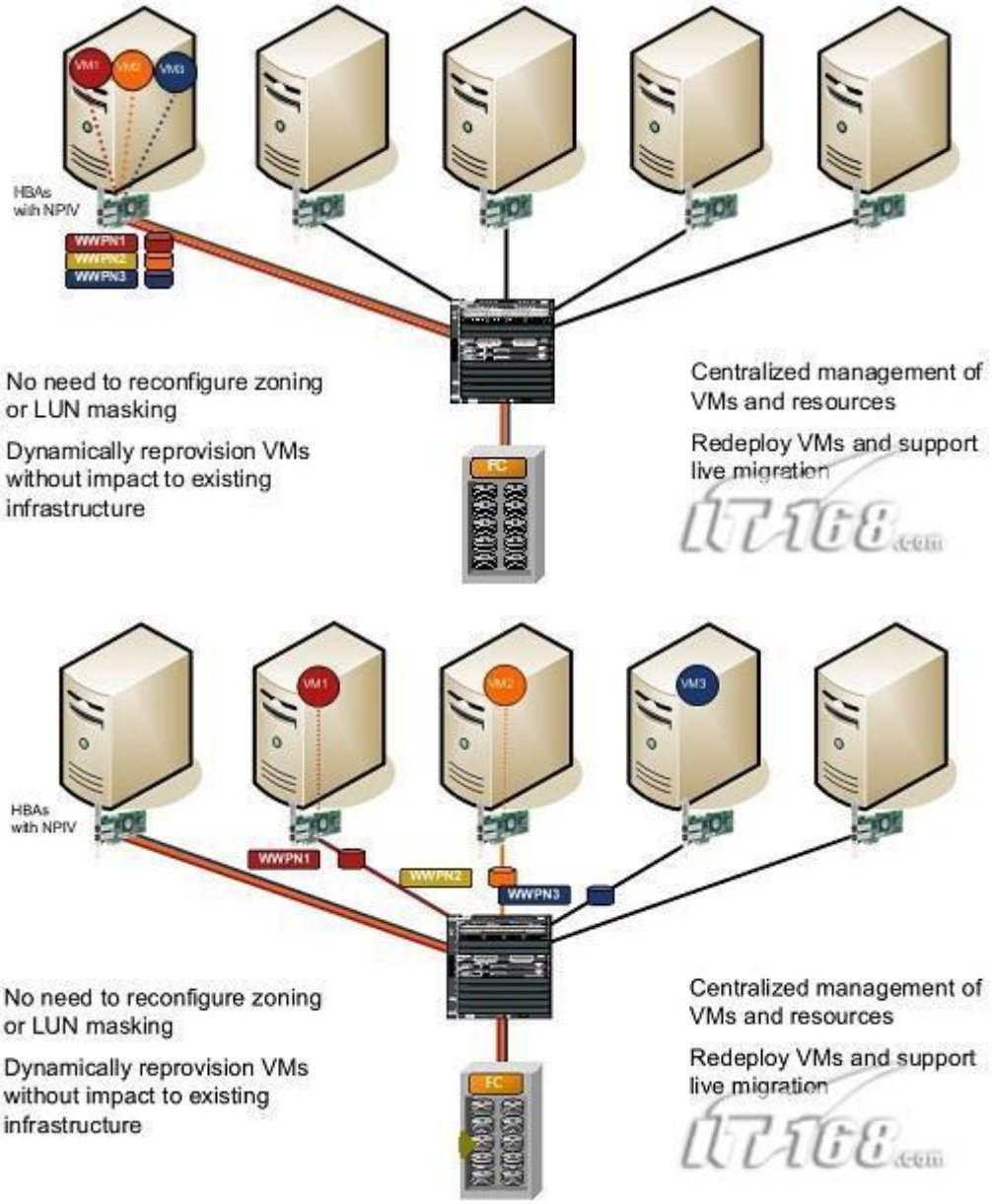

不同虚拟服务器通过不同虚拟 N-Port 连接到交换机，交换机为所有虚拟 N-Port 设置 WWPN，存储数据卷只能通过对应 N-Port 和虚拟服务器来处理。

没有使用 NPIV 技术的 VMotion LUN 迁移图如下。




所有配置参数都是基于一个物理 HBA 的 WWPN，所有虚拟服务器的 LUN 必须对所有服务器可见以确保迁移过程中磁盘数据存取的顺利完成，这种“暴露”会带来安全隐患。

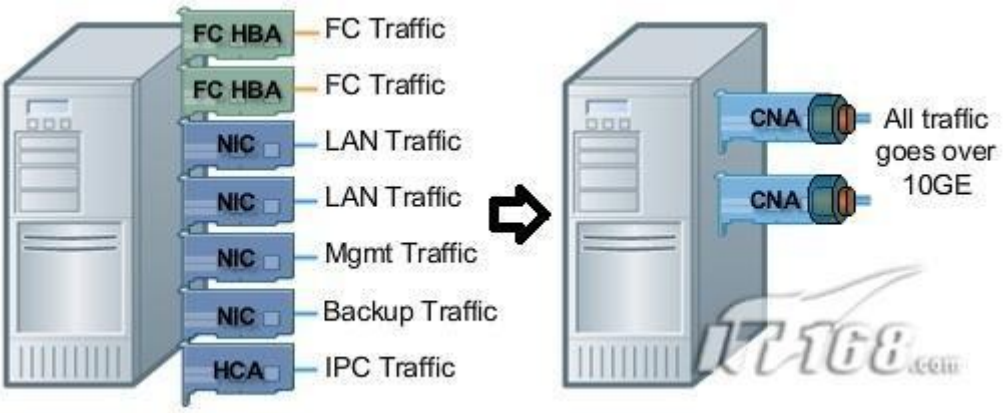
使用 NPIV 技术的 VMotion LUN 迁移。



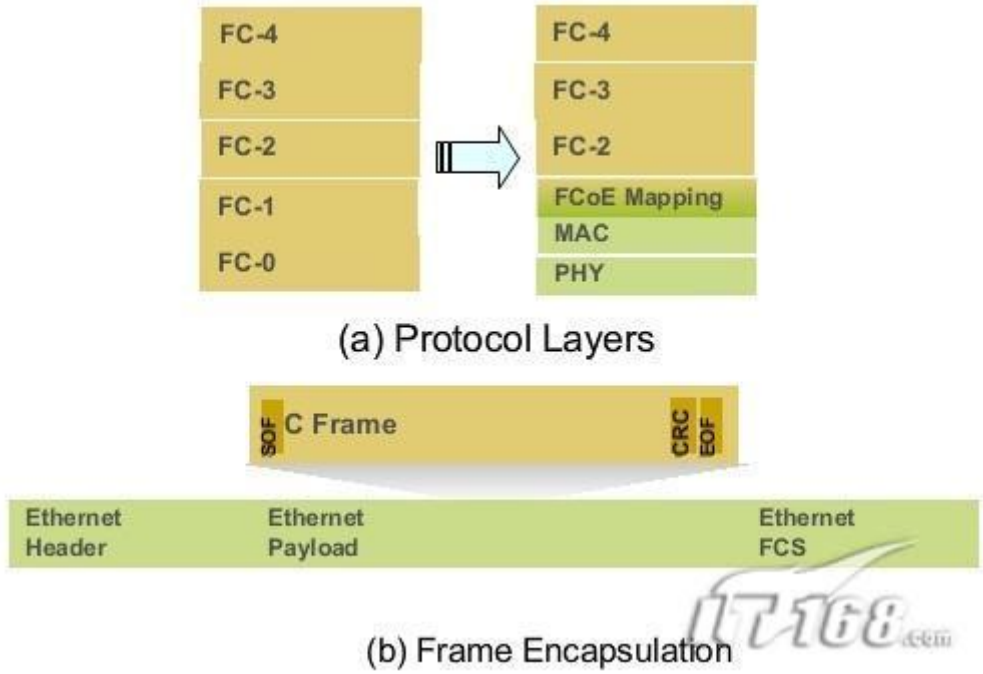

虚拟服务器迁移后，对应的 N-Port 参数配置和与存储的映射关系都会跟随迁移，确保特定存储只能通过相应虚拟服务器访问。使用 NPIV 技术可以简化管理和配置，增加安全性。

数据中心 FCoE(FC over Ethernet)技术实现在以太网架构上映射 FC(Fibre Channel)帧，使得 FC 运行在一个无损的数据中心以太网络上。FCoE 技术有以下的一些优点：光纤存储和以太网共享同一个端口;更少的线缆和适配器;软件配置 I/O;与现有的 SAN 环境可以互操作。

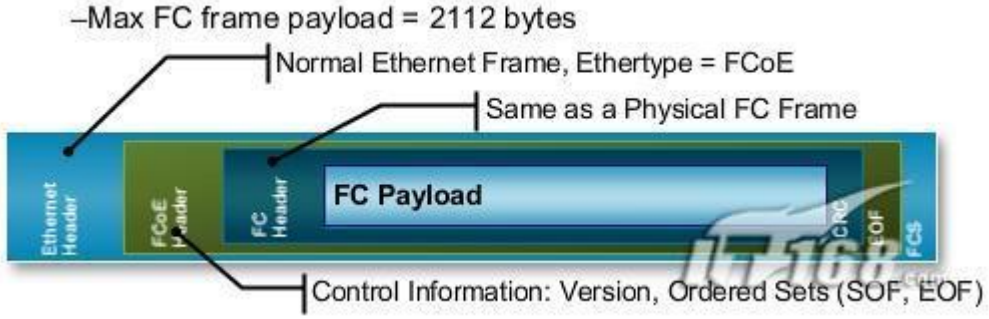
基于 FCoE 技术的数据中心统一 I/O 能够实现用少数的 CNA(Converged Network adapter)代替数量较多的 NIC、HBA、HCA，所有的流量通过 CNA 万兆以太网传输，如下图。



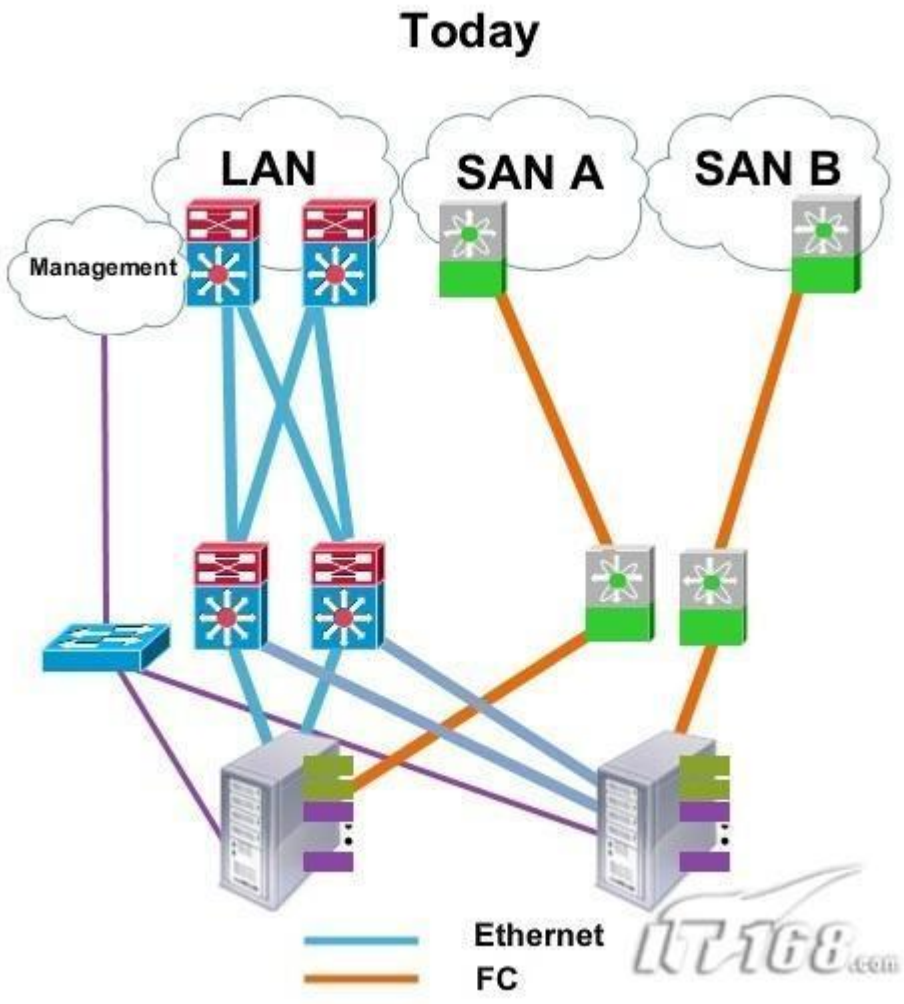

FCoE 下 FC 协议层的转换和 FC 帧、以太网数据帧结构如下图。



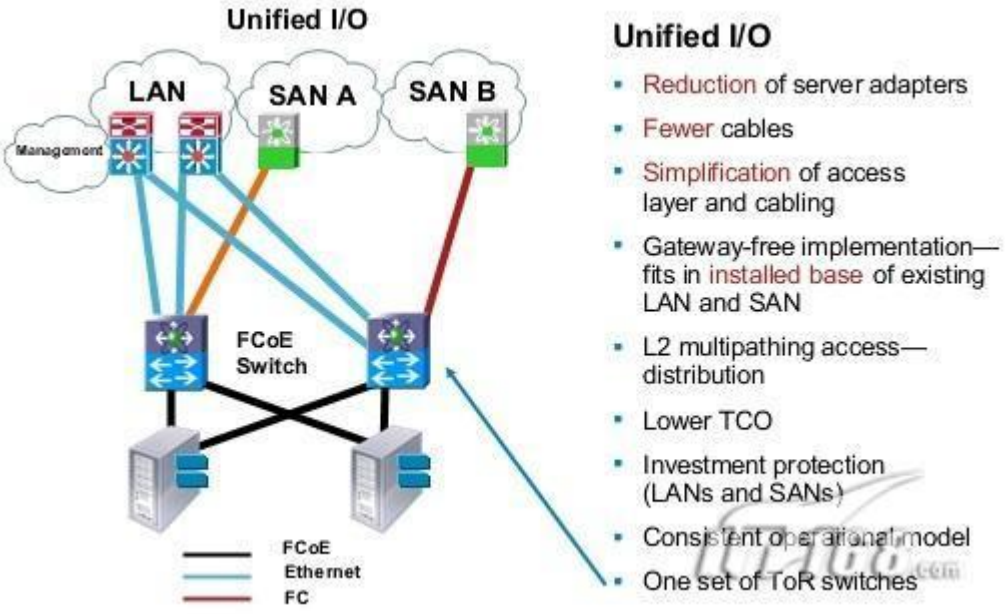

FCoE 在以太网上传输的 FC 数据帧的构成如下图。




FCoE 为 Unified I/O 的实现提供支持，现在的数据中心网络架构如下图，网络结构中存在局域网、SAN 网络等独立的采用不同协议的网络。

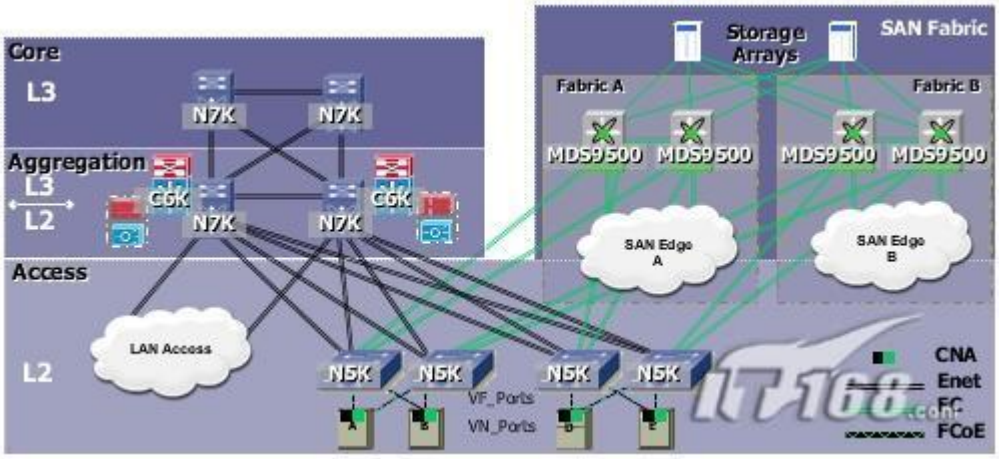



采用了 FCoE 技术实现的 Unified I/O 网络结构如下图。

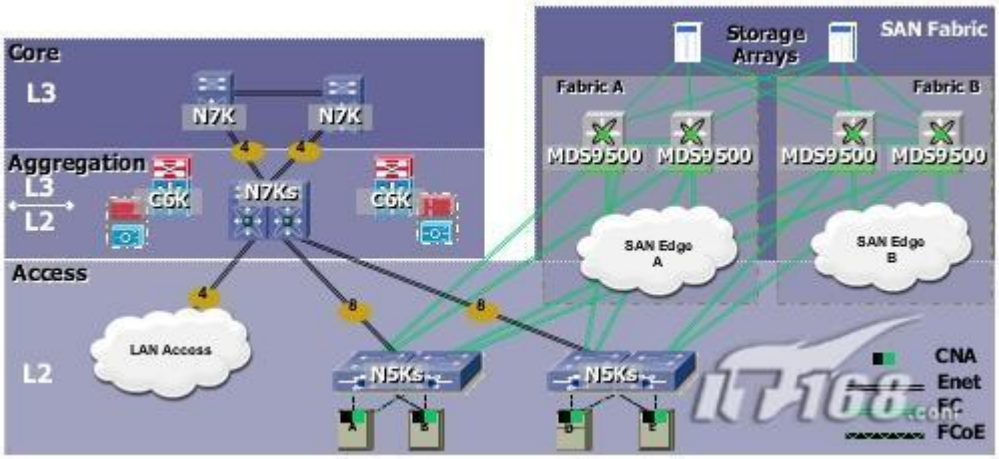


Unified I/O 在数据中心网络架构中的部署。

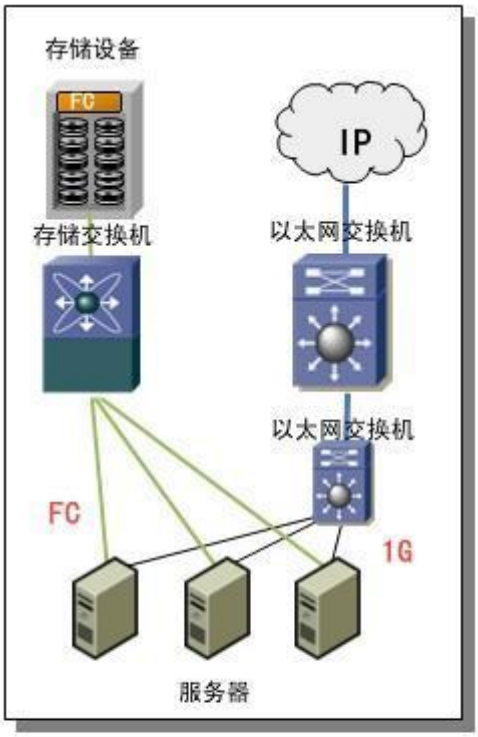
部署之前。




部署之后，实现汇聚层、服务器层的整合。

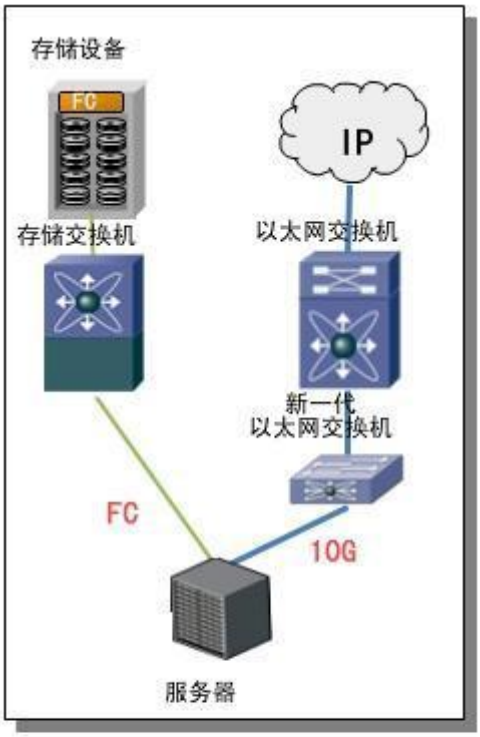



企业合理的数据中心网络虚拟化技术的运用路线如下。



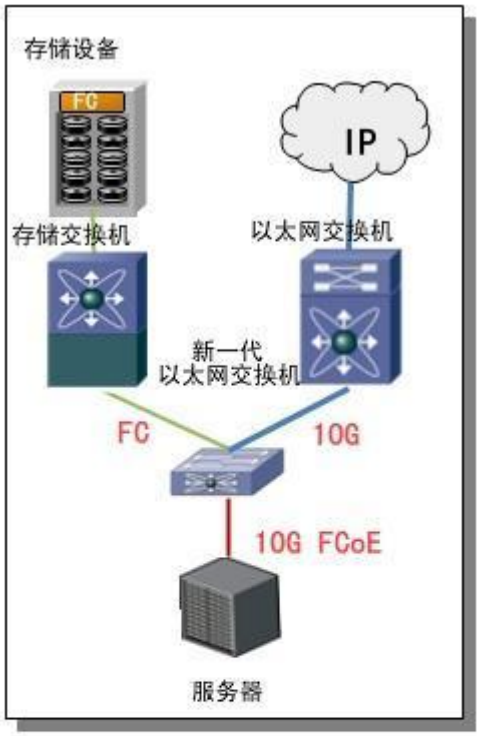

## 现阶段

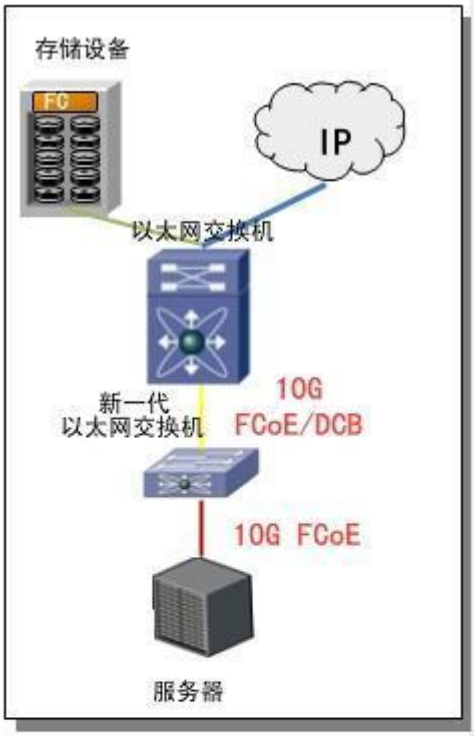
- 服务器利用效率较低
- 千兆以太网上联
- 存储和IP分离




## 第一步：服务器虚拟化

- 服务器整合虚拟化，提高利用率
- 要求万兆以太网上联
- 存储和IP分离




## 第二步：接入交换机整合

•服务器上联融合，降低运维成本

•要求万兆FCoE网络上联

•网络支持DCB无丢帧技术

•存储和IP部分融合

## 第三阶段：IP存储完全融合

•IP和存储完全融合，降低运维成本

•要求万兆FCoE网络上联

•数据中心DCB技术全面部署

•存储和IP全部融合