SmartCenter 技术白皮书

版本：SmartCenter v1.0

北京智网科技股份有限公司

目录

1. 前言

随着数据集中在企业信息化领域的展开以及云计算在企业信息中

心逐渐落地，新一代的企业级信息和数据中心的云计算基础设施及服务建设成为当前行业信息化的新热点。

传统的数据中心建设过程中，随着应用的不断扩展开和用户量的增加，服务器、存储、网络在数据中心内的不断增长、集中，引起较多的问题。如 数据中心有限空间内物理设备数量不断增长，面临巨大的布线、空间压力，而持续增长的高密 IT 设备功耗、通风、制冷也不断对能耗提出更高要求。服务器、网络、存储等 IT 设备的性能与容量不断增强，但是总体系统利用率低下，统计显示当前服务器平均利用率为 15%，存储利用率在 30%-40%。而企业 IT 的投入仍在不断增加。 因此，对数据中心的资源进行整合、进而虚拟化，以提高数据中心的能效、资源利用率、降低总体运营费用，成为当前 IT 业内最为令人关注的技术领域。同时，虚拟化对 IT 基础设施进行简化、优化。它可以简化对资源以及对资源管理的访问，为新的应用提供更好的支撑，基于虚拟化技术的 IAAS 云计算技术已经在各信息中心逐步落地。

无论是公有云还是私有云全都是指同一个概念:更高效地利用计算资源以支持当下的应用程序和将来的工作负载。云计算的目标是实现高度的灵活性,以提高资源利用率、控制成本并实现内部 IT 基础架构与基于网络的 IT 基础架构之间的轻松集成。智网科技做为国内云计算的驱动者和实践者，推出了 SmartCenter 帮助企业实现云计算平台。

本书描述了智网科技 SmartCenter v1.0 版本的主要技术参数和关

键技术模型。

1.1. 什么是 SmartCenter

SmartCenter 是由北京智网科技股份有限公司推出的一款云计算平台软件,主要服务于各种规模、各种类型的数据中心和信息中心，提供基础设施(IAAS)服务.

2. 系统要求

2.1. 部署方案

SmartCenter 支持高可用集群、多机部署方案。其各个组件相互独立，即可多个组件部署在一台服务器上，也可分别部署在不同的服务器上，但请注意根据服务负载情况适当分配组件的部署位置，以达到最佳效果。

2.2. 硬件要求

为了满足不同用户的需求，本书主要定义了两种硬件需求：最低配置和推荐配置。

2.2.1. 服务器

最低配置：支持 Intel 或者 AMD 虚拟化技术的 64 位 CPU 架构

推荐配置：

服务器类型	CPU	内存	硬盘	网卡
控制节点	64-bit x86 双路四核	12 GB	30 GB (SATA, SAS or SSD)	2 块千兆以太网卡
网络节点	64-bit x86 双路四核	8GB	30 GB(SATA, SAS or SSD)	3 块千兆以太网卡
计算节点	支持 Inel VT/AMD-V 虚拟化技术 的 64-bit x86 双路四核	16GB 以上， 越多越好 可以添加外 部内存扩展 槽的设备	可选择高性 能固态硬盘 技术 2TB (SATA)	2 块千兆以太网卡 可选择万兆以太网网卡
对象存储网关	双路四核	4GB	30 GB(SATA, SAS or SSD)	2 块千兆网卡
对象存储节点	双路四核	8GB	无需使用 RAID 24X2T SATA	2 块千兆网卡

对象存储账户节点	双路四核	8GB	30 GB(SATA, SAS or SSD)	2 块千兆网卡

2.2.2. 存储

为了支持高可用性及动态迁移，你可以配置存储设备。

SmartCenter 的存储解决方案对象存储、块存储。

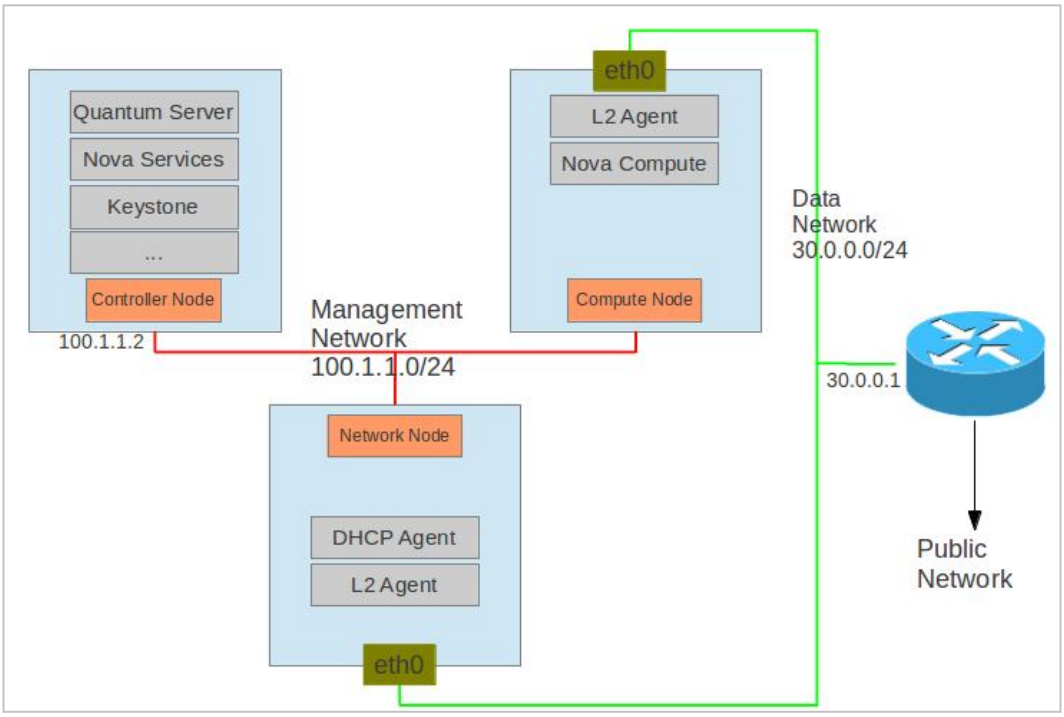
SmartCenter 对象存储设备采用软件冗余及负载均衡设计，避免了单点故障，可以使用普通服务器和家用硬盘等低廉的硬件来提供支持。

块存储可以采用通用的 SAN 解决方案

2.2.3. 网络

SmartCenter 采用 SDN(软件定义网络)构建网络模型.

在 SmartCenter v1.0 中，支持单一扁平网络部署模型:

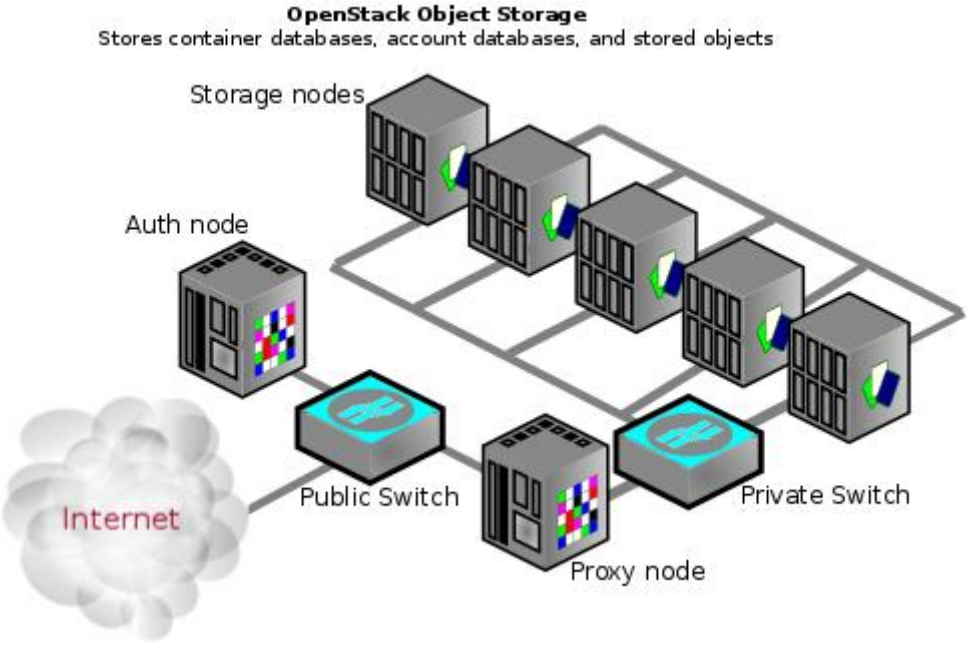


管理网络应至少为千兆以上的快速以太网,万兆光纤网络最佳.

SmartCenter 计算网络模型中共定义了五个子网络：

1. 管理网络(Management Network) ,使用 RFC1918 IP 范围, 无公网路由): 管理网络用来支持云平台基础架构的内部通信,推荐大小: 255 IPs (CIDR /24)
2. 公用网络(Public Network) ，使用公网ＩＰ地址: 公网网络用来提供外部用户通过 API 接口地址访问云平台基础架构. 最小尺寸: 8 IPs (CIDR /29)
3. 虚拟机网络(VM Network),使用 RFC1918 IP 范围, 无公网路由): 虚拟机网络为云平台中的虚拟机实例提供 IP 地址. 推荐尺寸: 1024 IPs (CIDR /22)\
4. 浮动 IP 网络(Floating IP network)，使用公网 IP 地址: 浮动ＩＰ用来连接公用网络到虚拟机实例. 最小尺寸: 16 IPs (CIDR /28)

5. 存储网络(Storage Network),使用 RFC1918 IP 范围,存储网络用于对象存储节点之间的内部通信。推荐尺寸：255IPS (CIDR/24)，存储网络的部署架构如下图所示：



2.3. 软件和许可要求

2.3.1. SmartCenter

SmartCenter 推荐运行在最新的 Linux 内核上(2.6.16 或更高)，我们对 Ubuntu12.04 和 RedHat/CentOS 6.4 进行了充分测试确保可以运行。

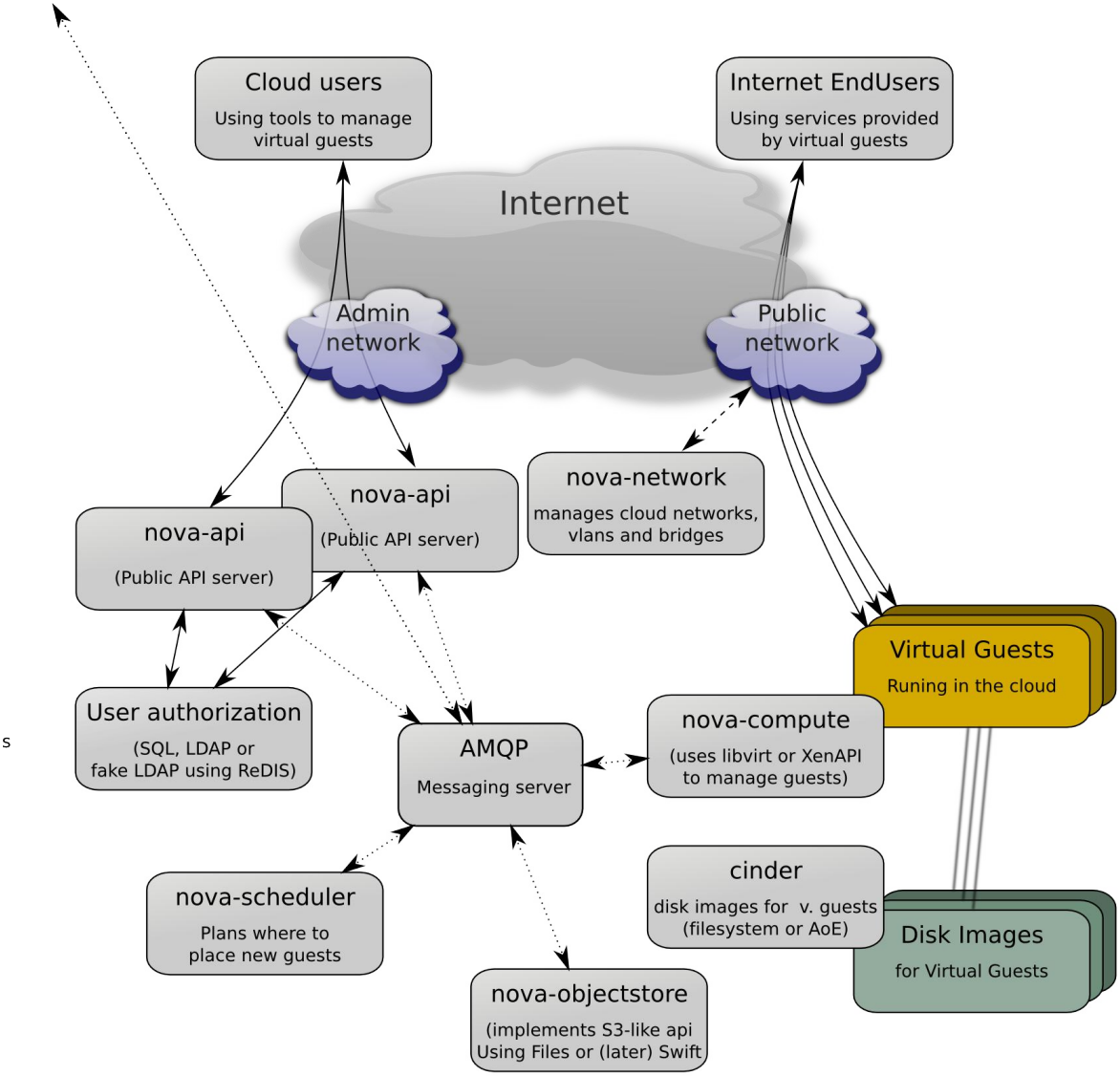
SmartCenter 的软件许可请参考《SmartCenter1.0 许可说明》

2.3.2. 客户操作系统

本书对客户操作系统没有特殊要求，客户操作系统需要能运行 IE、Chrome 或者 Firefox 较新版本的浏览器.支持 HTML5 会给你带来更美好的体检。

3. 技术简介

3.1. 服务架构



s

3.2. 主要特性

3.2.1. 高性能的虚拟机

为 SmartCenter 提供强劲驱动力的 SmartVM 虚拟机具备优异的性

能，在一般的软件计算中，SmartVM 最高能达到相当于物理机 90%或者更高的性能，而对于某些特定软件工作负载(如 LAMP)，经过优化后的 SmartVM 性能甚至超越了物理性能，达到物理机性能的 140%。

最新版本的 SmartVM 支持最大 160 个 VCPU 单位，支持最新的工业标准 CPU（包括 Intel Core I3,I5,I7 和 AMD 15h，Opteron G4），每个虚拟机的内存可扩展到 2TB.

3.2.2. 良好的用户界面

SmartCenter 提供良好的用户界面，无论你的身份是管理员还是用户，通过基于 HTML 标准的浏览客户端，都可以灵活的操控 SmartCenter。

在管理员控制面板，您可以：

1. 管理用户
2. 管理虚拟机
3. 管理镜像
4. 管理网络
5. 查看平台运行报告

在用户管理面板，您可以：

1. 新建虚拟机
2. 为虚拟机创建快照
3. 管理自己的虚拟机放火墙规则
4. 创建自己的虚拟机内部网络
5. 查看虚拟机的运行状态报告

3.2.3. 连续的高可用性

活动快照技术可以让你快速保存虚拟机的工作状态，并在需要时快速恢复

基于统一存储的虚拟机热迁移技术让您可以灵活地话虚拟机从一台物理机迁移到另一台物理机。

基于存储的热插拔技术可以让管理员在不停机的情况随时增加或者替换磁盘。

3.2.4. 广泛的操作系统支持

采用 SmartVM 的 SmartCenter 支持在各种操作系统的虚拟化及镜像安装，包括 Windows Server 2003、Windows Server 2008、Windows Server 2008 R2 及 Redhat、CentOS、Ubuntu、Open SUSE、Debain 等各种主流 Linux,包括 32 位与 64 位.

3.2.5. 优异的安全性

3.2.6. 统一的ＡＰＩ接口

SmartCenter 提供统一的、全面的ＡＰＩ接口供客户程序调用。

4. 关键技术

4.1. ＣＰＵ虚拟化技术

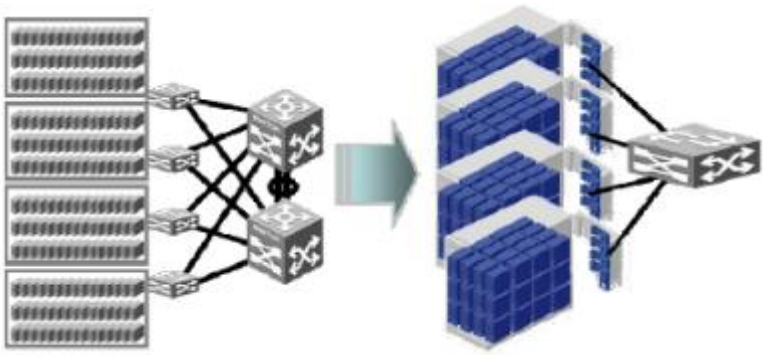
CPU 的虚拟化技术可以使单 CPU 模拟多 CPU 并行，允许一个平台同时运行多个操作系统，并且应用程序都可以在相互独立的空间内运行而互不影响，从而显著提高计算机的工作效率。

虚拟化技术与多任务以及超线程技术是完全不同的。多任务是指在一个操作系统中多个程序同时并行运行，而在虚拟化技术中，则可以同时运行多个操作系统，而且每一个操作系统中都有多个程序运行，每一个操作系统都运行在一个虚拟的 CPU 或者是虚拟主机上；而超线程技术只是单 CPU 模拟双 CPU 来平衡程序运行性能，这两个模拟出来的 CPU 是不能分离的，只能协同工作。

支持虚拟技术的 CPU 带有特别优化过的指令集来控制虚拟过程，通过这些指令集，VMM 会很容易提高性能，相比软件的虚拟实现方式会很大程度上提高性能。虚拟化技术可提供基于芯片的功能，借助

兼容 VMM 软件能够改进纯软件解决方案。由于虚拟化硬件可提供全新的架构，支持操作系统直接在上面运行，从而无需进行二进制转换，减少了相关的性能开销，极大简化了 VMM 设计，进而使 VMM 能够按通用标准进行编写，性能更加强大。

服务器虚拟化的直接效果是导致数据中心具有更高的应用密度，在相同物理空间内逻辑服务器(虚拟机)数量比物理服务器大大增加。由此，服务器的总体业务处理量上升，使得服务器对外吞吐流量增大。



虚拟化带来高密度的逻辑服务器群集

4.2. 软件定义网络技术(SDN)

SDN 是由美国斯坦福大学 cleanslate 研究组提出的一种新型网络创新架构，实现了网络流量的灵活控制，为核心网络及应用的创新提供了.良好的平台。

在 2006 年，斯坦福的学生 MartinCasado 领导了一个关于网络安全与管理的项目 Ethane，该项目试图通过一个集中式的控制器，让网络管理员可以方便地定义基于网络流的安全控制策略，并将这些安全策略应用到各种网络设备中，从而实现对整个网络通讯的安全控制。受

此项目(及 Ethane 的前续项目 Sane)启发，Martin 和他的导师 NickMcKeown 教授(时任 CleanSlate 项目的 FacultyDirector)发现，如果将 Ethane 的设计更一般化，将传统网络设备的数据转发(dataplane)和路由控制(controlplane)两个功能模块相分离，通过集中式的控制器(Controller)以标准化的接口对各种网络设备进行管理和配置，那么这将为网络资源的设计、管理和使用提供更多的可能性，从而更容易推动网络的革新与发展。于是，他们便提出了 OpenFlow 的概念，并且 NickMcKeown 等人于 2008 年在 ACMSIGCOMM 发表了题为 OpenFlow:EnablingInnovationinCampusNetworks[4]的论文，首次详细地介绍了 OpenFlow 的概念。该篇论文除了阐述 OpenFlow 的工作原理外，还列举了 OpenFlow 几大应用场景，包括： 1)校园网络中对实验性通讯协议的支持(如其标题所示);2)网络管理和访问控制;3)网络隔离和 VLAN;4)基于 WiFi 的移动网络; 5)非 IP 网络;6)基于网络包的处理。当然，目前关于 OpenFlow 的研究已经远远超出了这些领域。基于 OpenFlow 为网络带来的可编程的特性，Nick 和他的团队(包括加州大学伯克利分校的 ScottShenker 教授)进一步提出了 SDN(SoftwareDefinedNetwork,目前国内多直译为“软件定义网络”)的概念--其实，SDN 的概念据说最早是由 KateGreene 于 2009 年在 TechnologyReview 网站上评选年度十大前沿技术时提出[5]。如果将网络中所有的网络设备视为被管理的资源，那么参考操作系统的原理，可以抽象出一个网络操作系统(NetworkOS)的概念—这个网络操作系统一方面抽象了底层网络设备的具体细节，同时还为上层应用提

供了统一的管理视图和编程接口。这样，基于网络操作系统这个平台，用户可以开发各种应用程序，通过软件来定义逻辑上的网络拓扑，以满足对网络资源的不同需求，而无需关心底层网络的物理拓扑结构。

SDN(软件定义网络)应该是自扁平化数据中心网络架构提出之后，又一令整个网络产业极度关注的技术概念。其核心技术之一 OpenFlow 通过将网络设备控制面与数据面分离开来，从而实现了网络流量的灵活控制，为核心网络及应用的创新提供了良好的平台。SDN 和 OpenFlow 的横空出世，绝不是做网络变革的调味剂那么简单，当网络仍然在性能、架构扩展性，稳定性等方面纠结的时候，一场网络虚拟化的真正变革却悄然来临。

软件定义网络能够从路由器和交换机中的控制平面分离出数据平面，这个控制平面原本是专有的，只有开发它们的供应商知道，而在 SDN 中，控制平面将是开放的，并且受到集中控制，同时将命令和逻辑发送回硬件(路由器或交换机)的数据平面。 这提供了整个网络的视图，并且提供了集中更改的能力，而不需要在每个路由器或交换机上分别进行以设备为中心的配置更改。通过开放协议(例如 OpenFlow 标准)管理控制平面的能力，允许对网络或设备做出精确的更改，这将帮助增加网络的速度和安全性。

软件定义网络有以下优点：

- 拥有了自由移动的 SDN 网络后，工程师将能够通过快速且高水平地查看网络的所有区域以及修改网络来改变规则。
- 这种自由和控制还能为你的系统带来更好的安全性。通过快速

限制以及从中央视角查看网络内部的能力，管理人员可以有效地作出更改。例如，如果你的网络中爆发了恶意软件，通过 SDN 和 OpenFlow，你将能够迅速地从集中控制平面阻止这种流量来限制这种爆发，而不需要访问多个路由器或交换机。

- 快速对网络作出调整的能力使管理人员能够以更安全的方式来执行流量整形和数据包 QoS。这种能力现在已经存在，但速度和效率不好，当管理人员在试图保护网络安全时，这将限制他们的能力。

OpenFlow 交换机将原来完全由交换机/路由器控制的报文转发过程转化为由 OpenFlow 交换机（OpenFlow Switch）和控制服务器（Controller）来共同完成，从而实现了快速数据包转发（数据面）和高水平路由决策（控制面）分离。控制器可以通过事先规定好的接口操作来控制 OpenFlow 交换机中的流表，从而达到控制数据转发的目的。

OpenFlow 使传统的二层和三层交换机具备了细粒度流转发能力，即传统的基于 MAC 的网包转发，基于 IP 的路由转发，被拓展到了基于多域网包包头描述的流转发。同时，传统的控制层面从转发设备中剥离出来，所有转发行为的决策从交换机自身“迁移”到了某个集中控制器上。

OpenFlow 将给他们带来一个开放的硬件和软件路由，交换，让他

们拥有更全面的控制能力。

4.3. 虚拟机动态迁移技术

动态迁移则是云计算及虚拟化技用户尤为关注的技术。上个世纪80 年代,进程的迁移成为研究操作系统关注的焦点。然而进程的迁移却显示了很多的缺点例如:对原进程的依赖性。因此人们提出了在虚拟机上 OS 实例的迁移。基于虚拟机的 OS 迁移思想是在原域运行的过程中,将 OS 迁移到其它域上运行,并且对原域的资源访问能够延续,简单的说,OS 的迁移过程就是 OS 状态的保存,迁移,及在其它域上运行的过程。

迁移操作系统实例对数据中心和集群系统是一个很有力的工具,加强了系统负载均衡管理,降低了维护的费用,加强了服务器的整合性,提高了物理资源的管理性、系统的容错性。由于当前工作站性能价格比不断提高, 以及网络技术的飞速发展,从客观上为提高迁移效率提供了较好的基础, 同时为迁移提供了更多的机会。对于迁移的研究将进入一个更为迫切的阶段。

评价迁移性能的主要标准是迁移总的消耗时间和迁移的宕机时间,理论上可以从两个方面提高迁移性能,一种是减小迁移单位的大小,另一种是提高迁移的传输速度。针对目前迁移内存映像,常使用如下三种策略:

1) pure stop-and-copy[2]

该策略需要对原主机 halt,直接拷贝所有的页到目标机,然后重起

新的虚拟机,这种方法有很多的缺点,迁移的整个时间和宕机时间同分配给虚拟机的物理页的大小有关。这种方法适合小内存的迁移。目前国内外的产品中采用该技术的主要是 SWsoft 的 Virtuozzo。由于需要对原主机一个 stop 的过程,如果要迁移的 OS 映像比较大,直接影响到迁移的宕机时间,并且这种迁移技术是拷贝所有的页,当然迁移的单位相对很大。

2) pure demand-migration[1]

该策略在开始传输的时候,首先利用 stop-and-copy 的方法传输一些关键的内核数据结构,然后目标虚拟机重起,通过网络传输其余的页。这种方法有很多的优势,首先避免了 stop-and-copy 的直接停止原主机所造成的服务的间断。但是增加了迁移的总的时间,主要的缺点是迁移中很大程度上增加了对原虚拟机的依赖性。

3) pre-copy(预拷贝)

V[9]系统中基于进程迁移提出的预拷贝算法,目前采用的虚拟机软件 xen、SWsoft 的 virtuozzo 产品,还有 Virtual Iron 产品。这种方法在原域冻结之前,预先拷贝 OS 的关键状态,这时原域处于运行状态,之后冻结原域,重复拷贝这段执行时间内修改的脏页,这种方法的主要思想是减少 OS 冻结时间,从而避免因冻结时间长而导致的开销和错误,该策略虽然减少了 OS 的冻结时间,但是却增加了拷贝的次数,延长了总的迁移时间。如果第一次拷贝的是那些经常被访问的页,在下一次的拷贝中增加了拷贝脏页的次数,增加了拷贝的时间。

迁移系统映像的三种策略各有优缺点，其中 pure stop-and-copy 适

合于小内存系统迁 移 ,pre-demand-migration 的缺点是增加了对原域的依赖性,pre-copy 技术避免了这种依赖,但是增加了迁移的总的消耗时间,根据不同的系统,可以按照需求三种策略融合在一起。

4.3.1. 基础网络高可用设计

数据中心作为承载企业业务的重要 IT 基础设施，承担着稳定运行和业务创新的重任。伴随着数据的集中，企业数据中心的建设及运维给信息部门带来了巨大的压力，“数据集中就意味着风险集中、响应集中、复杂度集中……”，数据中心出现故障的情况几乎不可避免。因此，数据中心解决方案需要着重关注如何尽量减小数据中心出现故障后对企业关键业务造成的影响。为了实现这一目标，首先应该要了解企业数据中心出现故障的类型以及该类型故障产生的影响。影响数据中心的故障主要分为如下几类：

硬件故障

软件故障

链路故障

电源/环境故障

资源利用问题

网络设计问题

高可用数据中心网络设计思路：

数据中心出现的故障类型较多，风险也无法避免。那怎样才能做到当故障发生时对企业关键业务造成的影响最小呢？其实，我们可以看出虽然数据中心的故障类型众多，但故障发生产生的后果却大同小异。即，导致数据中心中的设备、链路或 server 发生故障，无法对外提供正常服务。对于这些故障的缓解最简单的方式就是冗余设计，可以通过对设备、链路、server 提供备份，从而将故障对用户业务的影响降低到最小。

但是否是一味的增加冗余设计就可以达到我们缓解故障影响的目的呢？有人可能会将网络可用性与冗余性等同起来。但事实上，冗余性只是整个可用性架构中的一个方面。一味的强调冗余性反而可能会降低可用性减小冗余所带来的优点，因为冗余性在带来好处的同时也会引入一些缺点：

网络复杂度增加

网络支撑负担加重

配置和管理难度增加

因此，数据中心的高可用设计是一个综合的概念。我们在选用高可靠设备组件、提高网络的冗余性的同时，还需要加强网络构架及协议部署的优化，从而实现真正的高可用。设计一个高可用的数据中心网络，可参考类似 OSI 七层模型，在各个层面保证高可用，最终实现数据中心基础网络系统的高可用，如下图所示：

网络架构高可用设计

从近年来企业信息化发展趋势来看，未来企业业务的发展对数据中心的依赖将越来越强，数据中心将会成为企业信息系统的“原点”，企业所有新业务的开展都将在数据中心内进行部署，同时随着新的数据处理技术（如数据仓库、商业智能）的应用，数据中心将会成为企业分析、评估和决策过程中的重要工具和数据支撑。这些都会对数据中心网络架构的设计产生影响。如下图所示：

传统的以局域网为基础、业务驱动不断扩展的“数据中心”网络架构思路将会越来越不适应企业业务的变更，在不断扩容、改造的过程中极易引入新的人为故障和设备故障，影响数据中心的可用性。以下是当前大多数企业用户的数据中心局域网络拓扑：

在上述拓扑中，主要有以下几个方面的问题：

局域网核心与数据中心核心交换设备共用，此核心承接了广域、局域和服务器之间的数据交互，不同类型的流在这里纵横交汇，随着网络不断扩容，此设备上的配置将会越来越复杂，风险集中；

局域网和服务器区的安全策略集中在核心部署，为实现安全分域管理，所有服务器的网关均设置在核心的防火墙上，核心防火墙极易受到攻击而瘫痪。一旦瘫痪将会导致所有服务器均无法对外提供服务，所有业务中断，风险大而且集中；

服务器基本处于无序的堆砌状态，网络层仅仅提供了提入互通的通道，没有进行系统的分区管理。在数据集中后服务器的数量会大量增加，安全及管理问题将会突显，无法满足业务的持续扩展。

从严格意义上讲，上述的结构并不是真正意义上的“数据中心”架

构，而且在传统的局域网基础之上，不断的扩展和修补形成的结果。这种架构可能在服务器规模不大的情况下可以满足大部分的业务系统部署要求。但随着业务系统的不断增多，服务器规模增加大一定数量（约200台）的情况下，网络配置将越来越复杂，最终导致网络运维管理员不敢去碰去动核心设备（交换机、防火墙）上的配置，信息部门的运维压力非常大。

为解决这些问题，企业在进行数据中心架构重新规划设计时，需要严格按照模块化、层次化原则进行，避免在后续规模越来越大的情况再进行大规模的整改，费时、费力且费钱。

模块化

模块化设计是指在对一定范围内的不同功能或相同功能不同性能、不同规格的应用进行功能分析的基础上，划分并设计出一系列功能模块，模块之间松耦合，力求在满足要求的基础上使网络稳定可靠、易于扩展、结构简单、易于维护。

不同企业的应用系统可能有一定的差异，在网络层面，根据应用系统的重要性、流量特征和用户特征的不同，可大致分为以下几个区域，如下图所示：

上述分区中有几点比较关键：

企业园区网核心与数据中心核心分离，各司其职。园区网核心主要承接纵向流量和用户的接入控制（DHCP、认证等）；数据中心核心主要承接服务器间的流量(横向流量居多)，数据中心核心交换机上部署尽可能少的策略和配置，保证其互连互通的高可靠、高性能，同

时在扩展新的模块时力求达到核心设备配置的零更改，各模块之间互通的松耦合，避免某功能模块故障而影响其它功能模块，实现风险分散、灵活扩展；

分布式安全部署。与传统的防火墙集中在核心旁挂的方式不一样，在模块化数据中心网络架构中，安全设备部署应下移到各功能模块的出口（汇聚层）位置，如上图中的红色网格线所示。而不是旁挂部署在核心交换区，这样做的目的也是分散风险，实现各模块间的松耦合。数据中心核心交换区就像是连接各城市的高速公路，建设时应充分保证其高可靠和高性能，而不部署红绿灯调度；

Intranet 服务器区是企业应用系统的关键分区，此分区可根据应用业务的关键性、实时性等特征的不同，可考虑再进行子分区的划分，一般而言可分为“关键业务区”、“通用业务区”、“财务应用区”几类，子分区可以是物理的，也可以是逻辑的。如果是逻辑的，可为每个子分区分配一个虚拟防火墙来部署安全策略。

层次化

数据中心层次化设计包括网络架构分层和应用系统分层两个方面，在当前网络及安全设备虚拟化能够不断完善的情况下，应用系统分层可完全通过设备配置来实现逻辑分层，不影响网络的物理拓扑。

对于网络架构层次化设计，三层架构还是二层架构可能是不少企业进行数据中心网络建设时面临的选择。传统网络核心、汇聚、接入各层的职责定义如下：

核心层：主要负责的是数据的交换与路由，不负责处理；

汇聚层：主要负责的是数据的处理，选择和过滤等操作；

接入层：主要负责的是数据的接受与发送，负责端到端的链路建立释放

从可靠性的角度来看，三层架构和二层架构均可以实现数据中心网络的高可用。近年来随着云计算概念逐步炒热，二层扁平化网络架构更适合云计算网络模型，可以满足大规模服务器虚拟化集群、虚拟机灵活迁移的部署。对于二层和三层架构的选择，可参考以下表格的对比：

	三层架构	二层架构
可靠性	增加了一层网络设备，相对增加了故障点	网络故障点相对较少
安全性	风关在汇聚层，安全策略部署在汇聚层，容易部署	安全策略部署在接入层，相对比较分散，部署工作量大
服务器接入数量	服务器接入数量多	服务器接入数量较少
扩展性	同一功能分区内服务器数量扩展多，可灵活实现物理分区内的子量扩展受限 逻辑分区	同一功能分区内服务器数

运维管理	设备和管理点较多	设备少，管理点较少
成本	汇聚和接入设备可灵活选择配合，达到最佳的成本控制	接入设备要求较高，选型受限
适合场景	服务器数量多，安全策略控制严格的场合	服务器集群、虚拟机迁移应用较多，服务器搬迁移动频繁场合

二层还是三层架构没有决对的优劣之分，企业用户可根据自身的业务特点进行选择，也可以先二层，后续针对某些特定的功能分区采用三层组网。

模块化、层次化的架构设计将数据中心网络风险进行了分散，将出现问题后的影响降低到最小，同时模块之间的松耦合可增强数据中心的扩展，简化网络运维，降低在扩展的过程中管理员的人为故障，保证数据中心的可用性。

设备层高可用设计

设备可靠是系统可靠的最基本保证，对于数据中心核心交换区设备的可靠稳定更是重要，尽管可以通过架构、策略、配置等的调整和优化多种手段降低核心设备出问题的可能、降低出问题后的影响范围，但要解决最根本的设备本身软硬件故障，必须选用数据中心级的设备组网部署。

关于数据中心级设备的定义，业界目前还没有类拟的标准，但从

目前主流网络设备供应商提供的数据中心解决方案产品（如 H3C S12500、Cisco N7000、JuniperEX8200等）可以看出，数据中心级交换机产品应具备以下特征：

控制平面与转发平面物理分离

传统的园区网交换机（如 H3C S75E、Cisco C65 等）一般采用“Crossbar+共享缓存”的交换架构，引擎板即承担控制平面的工作，同时也承担数据转发平面的工作，跨槽位的流量转发报文需要经背板到引擎板的 Crossbar 芯片进行转发。这种架构限制了设备的可靠性和性能：

- 可靠性限制：引擎需要承接数据转发平面的工作，因此在引擎出现主备倒换时必然会出现丢包。此外引擎 1 + 1 冗余，也使得 Crossbar 交换网只能是 1 + 1 的冗余，冗余能力无法做的更高；
- 性能限制：受制于业界当前 Crossbar 芯片的工艺以及引擎 PCB 板卡布线等制造工艺，将 Crossbar 交换网与 CPU 主控单元集中在一块引擎板上的结构，一般单块引擎的交换容量不可能做的太高（一般约 1TB 左右）。

数据中心级交换机产品将控制平面与转发平面物理分离，采用 CLOS 多级交换架构可以大大提高设备的可靠性及性能，这类设备一般有独立的引擎板和交换网板，在交换架构上，数据中心级产品采用了 CLOS 多级交换架构，与传统的 Crossbar+共享缓存交换架构相比，有如下优势：

	SmartCenter+共享缓存	CLOS 多级交换
结构	1，单平面交换； 2，交换矩阵和控制统一，即引擎承担了交换和控制 双重功能；	1，多块交换网板共同完成流量交换 2，控制和交换硬件分离
转发能力	受限于单个交换芯片的交换能力，目前最大到 1TB 就很难提升。	多块交换网板同时分担业务流量，相当于 N 倍于单级交换的能力，可实现 5 ~ 10TB 交换容量
可靠性	引擎倒换会丢包	控制平面与转发平面硬件物理分离，引擎切换时不影响转发，可实现零丢包
冗余能力	引擎 1 + 1 冗余，双引擎负载分担式无冗余	引擎 1 + 1 冗余，交换网板 N+1 冗余

- 关键部件更强的冗余能力

除了上述的引擎和交换网的冗余外，此类设备的电源一般均可以配置多块，实现 N+M 的冗余，保证电源的可靠性更高；另外风扇的冗余也由原来的风扇级冗余，提高到了风扇框冗余，每个独立的风扇框内多个风扇冗余；

- 虚拟化能力

数据中心的复杂度越来越高，需要管理的设备也越来越多，设备

的虚拟化可将同一层面（核心、汇聚、接入）的多台设备虚拟化为一台，进行设备的横向整合，简化设备的配置和管理。

- 突发大流量的缓冲能力

随着业务整合、资源共享、数据仓库、数据挖掘及智能分析等业务的部署，将会使数据中心内部和业务服务器之间的横向流量越来越多。这种流量模型的变化会导致多服务器群向一个服务器群的流量、多个应用服务器向同一个数据库服务器的流量越来越频繁，这种多对一的流量模型是一种典型的拥塞模型，如果网络设备的缓存能力不够，将会导致丢包重传，导致业务系统的响应时间变长或中断。

数据中心级设备对端口的缓存容量进行扩容，并采用了新一代的分布式缓存机制，将原有的出方向缓存移至入方向，在同样的端口缓存容量条件下，这种分布式的缓存机制可以更好的缓存多对一的拥塞模型，能够更好的吸收数据中心的突发大流量。如下图所示：

- 绿色节能

数据中心是企业能耗的主要部门，同时高的能耗将会带来高的发热量，而这也是影响数据中心设备稳定运行的重要因素。选用低能耗设备降低发热量是提高可靠性的一个方面，另一方面设备本身的散热风道设计的合理与否？能否更好的配合机房的空调循环？也影响着数据中心的可靠性。

为更好的配合机房冷热风道的布局，机柜中发热量较大的设备最后是前后散热的风道设计。但普通的横插槽设备一般是左右散热的方

式，因此应优先考虑采用竖插槽的设备，实现前后散热。如下图中的理想散热风道设计：

链路层(L2)高可用设计

以太网是广播性质的网络，一旦链路成环路很容易导致广播风暴，耗尽网络链路及设备资源。然而在实际的数据中心网络部署中，在实现设备和链路冗余提高可靠性的同时，也带来了环路和复杂度的增加。

对于传统的数据中心服务器区接入～汇聚交换网络，针对无环设计和有环设计有多种选择方案。如下图所示：

拓扑	优点	缺点
1 倒 U 型	不启用 STP，好管理 VLAN 可以跨汇聚层交换机，服务器部署灵活	必须通过链路聚合保证高可用性 汇聚交换机故障时，服务器无法感知，无法实现高可用接入
2 正 U 型	不启用 STP，好管理 双 active 链路，接入交换机密度高	不能使 VLAN 跨汇聚层，服务器部署不灵活 接入交换机间链路故障，VRRP 心跳报文无法传递，整机做 VRRP 主备切换，故障收敛时间长。
3 三角形	链路冗余，路径冗余，故障收敛时间最短 VLAN 可以跨汇聚层交换机，服务器部署灵活	存在环路，需要启动 STP 协议

4 矩形	双 active 链路，接入交换机密度高 VLAN 可以跨汇聚层交换机	有一半的接入层流量要通过汇聚交换机之间的链路。当接入交换机上行链路故障时，所有流量将从一侧的交换机上行。收敛比变小，网络易拥塞，降低网络高可用性。 存在环路，需要启动 STP 协议

由上表可以看出，三角形组网提供了更高的接入可用性以及更灵活的服务器扩展能力，所以常见推荐的组网采用第 3 种拓扑方式。需要指出，接入交换机直接双上行与汇聚层设备相连，冗余连接并不是越多越好，而最小的三角形环能够提供最快的收敛速度和最高的可用性。例如下图中右侧图组网拓扑在接入层交换机和汇聚层交换机之间采用全交叉冗余，是一种过度冗余组网，反而增加交换机的生成树计算的复杂性以及故障排错的复杂性，所以不建议按这种方式部署。虽然三角形组网已经成为数据中心接入设计的最佳实践，但从网络的拓扑设计、环路规避、冗余备份等角度考虑，设计过程是极其复杂的。如 VLAN 的规划、生成树实例的拓扑阻塞、网关冗余选择，包括相应技术的参数选择、配置，故障切换的预期判断等，需要一套十分详细的流程，而在后期网络运行维护过程中面临的压力和复杂度是显而易见的。

引入虚拟化设计方式之后，在不改变传统设计的网络物理拓扑、保证现有布线方式的前提下，以 IRF2 的技术实现网络各层的横向整合，即将交换网络每一层的两台、多台物理设备使用 IRF2 技术形成

一个统一的交换架构，减少了逻辑的设备数量，同时实现跨设备的链路捆绑，消除环路的同时保证链路的高可用。如下图所示

对于服务器而言，目前的服务器绝大多数都标配了双网卡甚至更多的网卡，但在实际的部署时，大多数企业都是采用主备模式，双网卡的出口带宽没有得到充分利用，同时网卡主备切换需要较长的时间（秒级）。

在接入交换机部署了 IRF2 虚拟化之后，两台接入交换机与服务器双网卡实现跨设备的链路捆绑（采用 LACP 标准协议，服务器网卡驱动均支持），此时双网卡处于“双 Active 模式”，服务器出口带宽充分利用，而且此时网卡、链路、接入交换机出现故障时，切换时间有了数量级的提升（毫秒级），服务器接入的可靠性很好的得到保障。实际的部署如下图所示：

对于接入层设备来说，以 Top of Rack 配线接入为例：一般使用两台接入交换机对同类业务系统服务器进行接入，以满足服务器双网卡的上行要求。使用 IRF2 对网络汇聚层或服务器网关层的虚拟化整合是必要的，因为这是消除生成树和 VRRP 的关键网络层。对接入层网络来说，有下图所示的两种选择：

方式 A：保持原有网络拓扑和设备独立性不变，通过 IRF2 将汇聚网关层虚拟化，Top of Rack 接入交换机双归属上联的两条链路直接进行捆绑，消除了环路，服务器网卡归属到独立的两台交换机，双网卡采用传统的主备方式。

方式 B：在 Top of Rack 两台交换机之间增加 IRF2 互联线缆，使

得接入层也实现虚拟化整合，服务器双网卡连接的两台交换机虚拟化成一台，这两台交换机的所有上联线缆可实现跨设备的捆绑，服务器双网卡启用 LACP 捆绑，实现服务器双网卡(或多网卡)接入的高可用。IRF2 虚拟化整合之后，数据中心网络从服务器网卡接入至汇聚、核心交换机，二层链路可实现端到端捆绑。

IRF2 部署之后，相比 STP + VRRP 的协议收敛，IRF 设备及链路的故障切换时有了数量级的提升，下图为 IRF 实测数据：

从上表中的数据可以看出，IRF 部署后无论是设备级故障倒换还是链路级故障倒换，时延都是毫秒级，因此整个网络将是一个快收敛的网络，可用性得到很大的提升。

协议层(L3)高可用设计

数据中心协议层高可用设计可以从以下三个方面考虑：

- 路由协议部署
- 快速检测与切换
- 不间断转发

路由协议部署：

数据中心汇聚层到核心层间可采用 OSPF 等动态路由协议进行路由层面高可用保障。常见连接方式有两种，如下图所示。拓扑 1 采用了三角形连接方式，从汇聚层到核心层具有全冗余链路和转发路径；拓扑 2 采用了四边形连接方式，从汇聚层到核心层没有冗余链路，当

主链路发生故障时，需要通过路由协议计算获得从汇聚到核心的冗余路径。所以，三角形拓扑的故障收敛时间较短，但三角形拓扑要占用更多的设备端口。

在采用模块化、层次化设计之后，数据中心内部各分区与核心交换区的路由将会大大简化，因此针对拓扑 1 的组网方式，可进行 IRF2 横向整合，对汇聚层、核心层的双机设备进行虚拟化，实现跨设备链路捆绑实现汇聚层上行到核心层的多链路负载分担与备份，在此基础之上，核心层与汇聚层仅需要一个 VLAN 三层接口互联，直接在此 VLAN 三层接口上部署静态路由，简化数据中心内部的协议部署。

数据中心内部各服务器分区汇聚层与数据中心核心交换区之间采用 IRF 配合静态路由的方案部署，可简化后续运维的复杂度。但对于数据中心外联模块，由于外部路由相对较复杂，可部署 OSPF 动态路由，提高路由选择的灵活性。数据中心总体路由结构如下图所示：

快速检测与切换：

为了减小设备故障对数据中心业务的影响、提高网络的可用性，设备需要能够尽快检测到与相邻设备间的通信故障，以便能够及时采取措施，从而保证业务继续进行。

由于数据中心内部一般采用以太网(或 MSTP 城域)链路来进行互联，无法通过 SDH（Synchronous Digital Hierarchy，同步数字体系）告警检测链路故障，通常情况下只能依靠路由协议中的 Hello 报文机制。这种机制检测到故障所需时间为秒级。对于数据中心内部吉比特

速率级高速数据传输，超过 1 秒的检测时间将导致大量数据丢失。BFD（Bidirectional Forwarding Detection，双向转发检测）就是为了解决上述检测机制的不足而产生的，它是一套全网统一的检测机制，用于快速检测、监控网络中链路或者 IP 路由的转发连通状况，保证邻居之间能够快速检测到通信故障，50ms 内建立起备用通道恢复通信。BFD 检测可部署在广域/域城出口模块，数据中心核心层与外联模块（广域区、城域区）之前运行 OSPF 动态路由协议，并在核心层交换机上配置 BFD 与 OSPF 路由联动，广域、城域路由设备或链路出现故障时，核心交换机 CoreSW 快速感知，并通告 OSPF 进行快速收敛，缩短数据中心外联数据故障恢复时间。

OSPF 使用 BFD 来进行快速故障检测时，OSPF 可以通过 Hello 报文动态发现邻居，OSPF 将邻居地址通知 BFD 就开始建立会话。BFD 会话建立前处于 down 状态，此时 BFD 控制报文以不小于 1 秒的时间间隔周期发送以减少控制报文流量，直到会话建立以后才会以协商的时间间隔发送以实现快速检测。当网络出现故障时：

1. BFD 检测到链路/网络故障；
2. 拆除 BFD 邻居会话；
3. BFD 通知本地 OSPF 协议进程 BFD 邻居不可达；
4. OSPF 协议中止上层协议邻居关系；
5. 核心交换机 CoreSW 选择备用路径。

BFD 除了上述可以应用在数据中心外联模块（广域/城域）外，还可以部署在 IRF2 虚拟组内，快速检测出 IRF 分裂，提高 IRF 虚拟化部署的可用性。

当 IRF 正常运行时，只有 Master 上配置的 MAD IP 地址生效，Slave 设备上配置的 MAD IP 地址不生效，BFD 会话处于 down 状态；

当 IRF 分裂后会形成多个 IRF，不同 IRF 中 Master 上配置的 MAD IP 地址均会生效，BFD 会话被激活，此时会检测到 IRF 组分裂，IRF 会将优先级低的设备隔离，避免出现多 Active 冲突。

使用 BFD 进行 IRF MAD（Multi-Active Detection，多 Active 检测）检测时，需要有一条 BFD MAD 检测专用链路（千兆或万兆以太链路），这些链路连接的接口必须属于同一 VLAN。

不间断转发：

在部署了动态路由协议的数据中心网络中，当网络设备进行主备切换时，在路由协议层面会与邻居之间发生震荡。这种邻居关系的震荡将最终导致路由震荡的出现，使得主备切换路由器在一段时间内出现路由黑洞或者导致邻居将数据业务进行旁路，进而会导致业务出现暂时中断。

为了实现不间断转发 NSF 技术，除了设备本身需要支持数据转发与控制分离，支持双主控设计外。根据需要，可能需要部分保存协议的状态（控制平面），并借助邻居设备的帮助，实现发生主备切换时控制平面的会话连接不重置，转发不中断的目的。这些实现控制层面不重置的技术统称为路由协议的 Graceful Restart（平滑重启）扩展，

简称 GR，它表示当路由协议重启时保证转发业务不中断。

GR 机制的核心在于：当某设备的路由协议重启时，能够通知周边设备在一定时间内将到该设备的邻居关系和路由保持稳定。在路由协议重启完毕后，周边设备协助其进行路由信息同步，在尽量短的时间内使该设备的各种路由信息恢复到重启前的状态。在整个协议重启过程中，网络路由和转发保持高度稳定，报文转发路径也没有任何改变，整个系统可以不间断地转发 IP 报文。

在数据中心 OSPF 动态路由部署的区域（广域、外联、园区、互联网等），一般按照如下的组网结构部署 GR：

使用 GR 保证网络中的核心层节点和广域出口节点在出现协议重启时的转发业务不中断，避免出现不必要的路由振荡。

核心层节点和广域出口节点作为 GR Restarter（同时缺省也作为 GR Helper），分支节点作为 GR Helper。这样当广域出口节点发生主备切换或重启 OSPF 进程时，核心节点可以作为 GR Helper 协助其进行 LSDB 重同步，并且保持转发不中断；当核心层节点发生主备切换或重启 OSPF 进程时，广域出口节点和分支节点都可以作为 GR Helper 协助其进行 LSDB 重同步，并且保持转发不中断。

应用层(L4～L7)高可用设计

在数据中心网络层面实现 L4～L7 层的高可用，可采用负载均衡的方案。L4～L7 层负载均衡一方面可以提高服务器的响应能力和链

路的带宽利用率，另一方面可以保证单台服务器或单条链路出现故障后，业务数据无缝分摊到其它服务器和链路，从而实现数据中心的高可用。对于 L4～L7 层负载均衡，分为以下两个方面：

a) L4～L7 链路负载均衡（LLB）

b) L4～L7 服务器负载均衡（SLB）

链路负载均衡：

链路负载均衡常部署在数据中心的广域接入区和互联网接入区，通过静态表项匹配及动态链路检测，对多条链路状态进行实时的探测和监控,确保流量以最合理及快速的方式分发到不同链路上，,实现业务的高效传输。

对于数据中心广域接入区，由于广域网出口流量仍然是企业内网数据流，在 L4 层一般可通过 IP 报文的五元组特征区分出不同的业务流，因此可直接在路由器上通过分层 CAR、跨端口的流量转发实现负载分担、关键业务带宽保证、广域链路捆绑。无需专门的 LB 设备。如下图所示：

流量控制要求

基本业务分流：通常情况下，生产业务走主链路，办公和视频业务走备用链路。

超负荷流量调度：无论主备链路，超负荷流量走对方链路；备用链路视频业务不要进行超负荷流量分担；纵向出口进行多业务 QoS 调度。

设计实现

基本业务分流：通过 OSPF COST 设计，生产业务默认走主链路

转发，对办公和视频业务采用策略路由走备链路。

超负荷流量调度：以备链路为例，需要在数据中心广域网的入口进行流量监管 CAR，超过 10M 的流量结合策略路由调度到左侧路由器。为保证视频流量不会被调度到左侧路由器，必须采用分层 CAR 实现。

对于 Internet 出口链路负载均衡，由于内网用户访问的数据流不固定，特征复杂，很难在 L4 层区分出不同的业务流，因此需要部署专门的负载均衡设备实现多运营商出口的链路负载均衡。并启用 Inbound 和 Outbound 两个方向的负载均衡，一方面满足企业内网用户或服务器访问外部 Internet 站点的流量分担；另一方面满足外网用户通过 Internet 访问企业公共服务（如网站、FTP 等）的流量分担。

Outbound 链路负载均衡中，用户将访问外网的报文发送到 LB 负载均衡设备后，负载均衡设备根据就近性算法和调度策略，将内网访问外网的业务流量分别分发给相应的链路。

Inbound 链路负载均衡中，负载均衡设备作为权威名称服务器记录域名与内网服务器 IP 地址的映射关系。一个域名可以映射为多个 IP 地址，其中每个 IP 地址对应一条物理链路。外网用户通过域名方式访问内网服务器时，本地 DNS 服务器将域名解析请求转发给权威 DNS 服务器——LLB 负载均衡设备，负载均衡设备依次根据持续性功能、ACL 策略、就近性算法选择最佳的物理链路，并将通过该链路与外网连接的接口 IP 地址作为 DNS 域名解析结果反馈给外网用户，外网用户通过该链路访问内网服务器。

服务器负载均衡：

目前大多数应用系统都采用了 BS 架构，企业数据中心的 WEB 服务器需要承接来自内网和外网众多用户的连接请求，因此单台服务器的性能和可靠性可能都无法满足，为实现更多的用户接入数和服务器冗余，可在 WEB 服务器部署负载均衡。服务器的负载均衡部署可采用以下两种方式实现：

服务器集群软件

服务器负载均衡(SLB)设备

采用服务器集群软件的方式与网络的相关性不大，一般要求服务器群在同一 VLAN 内即可，本文将重点针对“服务器负载均衡(SLB)设备”方式的设计和部署进行介绍。

依据转发方式的不同，服务器负载均衡的部署分为 NAT 方式和 DR 方式。两种方式的处理思路相同：LB 设备提供 VSIP(虚拟服务 IP)，用户访问 VSIP 请求服务后，LB 设备根据调度算法分发请求到各个实服务。而在具体的处理方式上有所区别：

NAT 方式：LB 设备分发服务请求时，进行目的 IP 地址转换（目的 IP 地址为实服务的 IP），通过路由将报文转发给各个实服务。服务器响应的报文也要经过 LB 设备进行 NAT 转换，这种方式 LB 设备承担的性能压力较大。

DR 方式：LB 设备分发服务请求时，不改变目的 IP 地址，而将

报文的目的 MAC 替换为实服务的 MAC 后直接把报文转发给实服务。服务器响应的报文不需要经过 LB 设备，直接转发到用户，这种方式 LB 设备承担的性能压力相对较小。

DR 方部署时需要对每个服务器配置 VSIP，并要求其 VSIP 不能响应 ARP 请求。而一般的企业网络运维和服务器运维是不同部门的不同工程师负责，这种 DR 方式的配置涉及到两个部门之间的配合，比较复杂，因此一般在 LB 设备性能足够的情况下不推荐使用。采用 NAT 方式部署组网灵活，对服务器没有额外要求，不需要修改服务器配置，适用于企业数据中心各种组网。

高可用设计与部署是企业数据中心建设的永恒话题，“勿在浮沙筑高台”，网络做为数据中心 IT 基础承载平台，是 IT 系统高可用的基本保证。数据中心网络要实现高可用，技术并不能解决所有问题，还需要完善的运维流程、规章制度、管理体制等多方面的配合。结合企业业务的发展趋势，不断的总结与积累，是一个长期的、循序渐进的过程。

4.4. 监控、管理、远程控制

站点监控

云监控可以对您的站点进行性能监控，其中，可用率和响应时间是两个重要指标。轻松创建不同类型的站点监控项目云监控支持多种

站点监控类型，它们对应着不同的网络访问传输协议，您可以利用它们来快速创建监控项目，从而监控您的站点。目前支持的站点监控类型包括：

云监控

提供不同的站点监控频率供您选择，从 1 分钟到 60 分钟。显然，更短的监控频率可以获得更加准确的站点可用率统计和平均响应时间统计，以及更加及时的告警通知。

当然，你并不需要担心过高的监控频率会导致您的站点负载，因为即便是 1 分钟的频率，监控请求导致您的站点资源消耗也是微不足道的。

服务器性能监控

云监控通过标准的网络管理协议 SNMP 来帮助您远程监控服务器性能，这一切只需要您在服务器上配置 SNMP 监控代理，按照我们的指引，SNMP 是安全的。安全的 SNMP 代理云监控对于 SNMP 的身份验证支持 v2c 和 v3，我们提供了多项安全配置建议，通过 v3 的加密身份验证，以及防火墙的保护，您完全可以放心的使用 SNMP。丰富的服务器监控项目类型目前云监控支持 Linux/Unix 服务器以及 Windows 服务器的性能监控，您可以创建各种类型的监控项目，包括：

・CPU 使用率

・CPU 负载

・内存使用率

・磁盘空间使用率

・磁盘 I/O

・网络流量

・系统进程数

对于 Linux 服务器，您可以看到详细的 CPU 使用率变化曲线图，包括用户态、内核态、IOWait 等，它们的使用率比例可以反映出您的服务器正在处理哪些性质的计算。

当用户态 CPU 使用率较高时，意味着服务器上应用程序需要大量的 CPU 开销，比如数据库服务器进行大量的查询和排序等计算。而当内核态 CPU 使用率较高时，则说明服务器花费大量的时间进行进程调度或者系统调用。

如果 IOWait 使用率较高，则意味着大部分 CPU 时间在等待磁盘 I/O 操作，这时候您的确应该检查一下磁盘 I/O 是否过高。

如果您正在使用 Windows 服务器，您一定很关心它的内存使用率变化，很多时候 Windows 服务器的内存使用率会越来越高，这同时也导致虚拟内存的异常消耗，使得服务器系统响应速度变慢。

通常情况下，我们将 Windows 虚拟内存大小设置为物理内存的两倍，不过对于具体应用，您需要更加经过详细的分析，监控它们的变化是必不可少的，通过了解它们的变化，您便可以不断调整内存使用策略。

网站安全扫描服务（Websec）

Websec 覆盖了 OWASP 所定义的常见 Web 应用程序漏洞，包括 SQL 注入漏洞，跨站脚本漏洞，文件上传漏洞，目录遍历，敏感文件

等，并且支持多种语言如 PHP,JSP,ASP,.net 编写的网页。扫描速度快，误报率低。

故障分析

云监控通过分布在各地的监控节点，运用各种故障分析手段，在故障发生时抓取各种信息，例如网络信息、域名解析信息等。帮助网站主判断故障原因，快速定位问题。

用户访问速度监控

你可以通过跟踪页面来快速采集全国 500 多个主要城市的真实用户访问速度，直观的地图展现让你一目了然。

在站点页面上加入用于自动跟踪用户访问速度的 Javascript 脚本您只需要按照云监控的指引，在您的站点服务器上放置一定尺寸的文件，然后将云监控的跟踪脚本嵌入到您的网页 HTML 中即可。其余的事情将由云监控自动完成，您要做的就是查看跟踪报告。

查看不同省份的用户访问速度云监控会根据跟踪到的用户 IP 来将它们统计到不同的省份中，并且按照响应时间来排序。

通常情况下，更多的跟踪次数意味着更多的抽样，所以访问速度更加的准确。所以，您需要让这些页面跟踪更多的用户。

在地图上直观的查看省份访问速度分布我们已经用不同颜色来标记不同的访问速度级别，绿色代表非常快，而红色代表非常慢，中间的过渡颜色则表示速度介于两者之间的不同程度。

如果您的站点没有采取南北分布策略或 CDN 加速,那么从地图上会非常容易看到南方和北方省份的速度差异。

警告通知

将告警消息通知到您

我们努力将各种告警消息及时告诉您，让您尽早的采取措施。相信您可以灵活的应用我们提供的各种通知方式，不论在您的工作时间还是休息时间，您都可以选择适合您的通知方式。

不放过任何新告警消息云监控管理中心具备告警消息站内实时提醒，故障消息、提醒消息和系统消息，并且配合闪烁的浏览器标题栏，让您快速知晓告警信息。

如果您不电脑旁，您也可以选择下边丰富的通知方式。

站内实时告警消息，支持 RSS 只要您的监控项目触发告警条件，比如网页无法打开，或者服务器 PING 数据包全部丢弃，云监控便会将这些情况记录到您的站内告警消息中，并通过新消息浮动层来通知您。

在告警消息列表中，可以直观的看到故障发生以及故障恢复的时间，以及必要的历史快照信息，比如当时的 HTTP 响应头信息。

支持多种常用的通知方式有了告警消息还是不够的，大多数情况下您可能并不知道，我们需要将告警消息及时通知给您，这就需要用到云监控提供的通知方式。我们支持以下几种通知方式：

EmailMSNGtalk 手机短信 RSS

为站点监控项目快速进行告警设置对于任何的站点监控项目，包括 HTTP、FTP、PING、DNS 等，您都可以直接查看它们的告警消息，并为它们设置常规告警和自定义告警。

常规告警是当站点不可用或者恢复可用时发送的通知，举个例子，对于 HTTP 网页监控来说，当网页不存在，或者服务器无响应时，便会触发告警。

如果站点持续故障，云监控并不会一直发送告警通知，而是会在故障恢复的时候发送通知。

傻瓜式的告警设置

对于常规告警设置，非常简单，您只需要勾选不同的通知方式在站点不可用或者恢复时是否开启。我们建议您灵活的组合使用这些告警通知方式，它们会互相弥补，既满足及时性，又可以让您了解详细的快照信息。

Email 告警通知来自云监控的告警通知 Email，包含了故障信息以及快照链接，您可以点击它来登录云监控查看详细的历史快照。

告警通知每日统计

通过告警通知每日统计，您可以清楚的了解各种通知方式的使用情况，特别对于付费的短信配额，通过每日统计，可以帮助您更好的分配短信告警配额。

4.5. 存储技术

SmartCenter 的存储解决方案主要分为对象存储和块存储

4.5.1. 对象存储

对象存储系统（Object-Based Storage System）是综合了 NAS 和

SAN 的优点，同时具有 SAN 的高速直接访问和ＮＡＳ的数据共享等优势，提供了高可靠性、跨平台性以及安全的数据共享的存储体系结构。

传统块存储与对象存储结构对比示意图：

- **对象存储系统由对象(Object) 、OSD（ Object-based Storage Device)、文件系统和网络连接组成:**

1. 对象(Object)

对象存储的基本单元。每个 Object 是数据和数据属性集的综合体。数据属性可以根据应用的需求进行设置，包括数据分布、服务质量等。在传统的存储中，块设备要记录每个存储数据块在设备上的位置。Object 维护自己的属性，从而简化了存储系统的管理任务，增加了灵活性。Object 的大小可以不同，可以包含整个数据结构，如文件、数据库表项等。

2、OSD（ Object-based Storage Device)

每个 OSD 都是一个智能设备，具有自己的存储介质、处理器、内存以及网络系统等，负责管理本地的 Object，是对象存储系统的核心。

OSD 同块设备的不同不在于存储介质，而在于两者提供的访问接口。

OSD 的主要功能

数据存储和安全访问

OSD 使用 Object 对所保存的数据进行管理。它将数据存放到磁盘的磁道和扇区，将若干磁道和扇区组合起来构成 Object，并且通过此 Object 向外界提供对数据的访问。每个 Object 同传统的文件相似，使用同文件类似的访问接口，包括 Open、Read、Write 等。但是两者并不相同，每个 Object 可能包括若干个文件，也可能是某个文件的一部分，且是独立于操作系统的。除了具体的用户数据外，OSD 还记录了每个 Object 的属性信息，主要是物理视图信息。将这些信息放到 OSD 上，大大减轻了元数据服务器的负担，增强了整个存储系统的并行访问性能和可扩展性。

3、文件系统

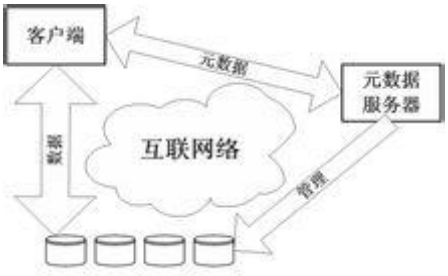
文件系统对用户的文件操作进行解释，并在元数据服务器和 OSD 间通信，完成所请求的操作。

现有的应用对数据的访问大部分都是通过 POSIX 文件方式进行的，对象存储系统提供给用户的也是标准的 POSIX 文件访问接口。

接口具有和通用文件系统相同的访问方式，同时为了提高性能，也具有对数据的 Cache 功能和文件的条带功能。

同时，文件系统必须维护不同客户端上 Cache 的一致性，保证文件系统的数据一致

文件系统读访问实例：



① 客户端应用发出读请求;

② 文件系统向元数据服务器发送请求，获取要读取的数据所在的 OSD;

③ 然后直接向每个 OSD 发送数据读取请求；

④ OSD 得到请求以后，判断要读取的 Object，并根据此 Object 要求的认证方式，对客户端进行认证，如果此客户端得到授权，则将 Object 的数据返回给客户端；

⑤ 文件系统收到 OSD 返回的数据以后，读操作完成。

4.元数据服务器 (Metadata Server)

为客户端提供元数据，主要是文件的逻辑视图，包括文件与目录的组织关系、每个文件所对应的 OSD 等。

在传统的文件系统中，元数据由本机或者文件服务器负责维护，每次对数据块的操作都要获取元数据。

在对象存储系统中，由于每次操作只有一次对元数据的访问，具体的数据传输都由 OSD 和客户端通过直接连接进行，大大减少了元数据的操作，降低了元数据服务器的负担，从而为系统的扩展提供了可能性。

特点

客户端采用 Cache 来缓存数据

当多个客户端同时访问某些数据时，MDS 提供分布的锁机制来确保 Cache 的一致性。

为客户端提供认证

为了增强系统的安全性，MDS 为客户端提供认证方式。OSD 将依据 MDS 的认证来决定是否为客户端提供服务。

5. 网络连接

为客户端提供认证

为了增强系统的安全性，MDS 为客户端提供认证方式。OSD 将依据 MDS 的认证来决定是否为客户端提供服务。

网络连接是对象存储系统的重要组成部分。它将客户端、MDS 和 OSD 连接起来，构成了一个完整的系统。

4.5.2. 块存储(SAN)

这里的块存储主要是指存储区域网络(Storage Area Network)

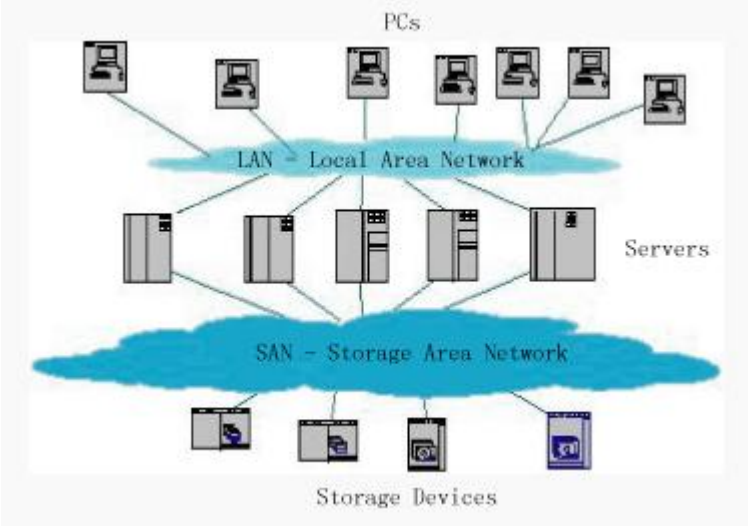
存储区域网络(SAN)是一种高速网络或子网络，提供在计算机与存储系统之间的数据传输。存储设备是指一台或多台用以存储计算机数据的磁盘设备，通常指磁盘阵列。一个 SAN 网络由负责网络连接的通信结构、负责组织连接的管理层、存储部件以及计算机系统构成，从而保证数据传输的安全性和力度。

典型的 SAN 是一个企业整个计算机网络资源的一部分。通常

SAN 与其它计算资源紧密集群来实现远程备份和档案存储过程。SAN 支持磁盘镜像技术(disk mirroring)、备份与恢复(backup and restore)、档案数据的存档和检索、存储设备间的数据迁移以及网络中不同服务器间的数据共享等功能。此外 SAN 还可以用于合并子网和网络附接存储(NAS:network-attached storage)系统。

当前常见的可使用 SAN 技术，诸如 IBM 的光纤 SCON，它是 FICON 的增强结构，或者说是一种更新的光纤信道技术。另外存储区域网络中也运用到高速以太网协议。SCSI 和 iSCSI 是目前使用较为广泛的两种存储区域网络协议。

SAN 的典型结构：



局域网、城域网和广域网都有相同的一个目的——让计算机相互通信。而存储区域网络(SAN)则不是以此为目的。它的目的是让计算机和存储设备进行通信。

对于一般的 PC 来说，存储设备通常就是在 PC 内部的磁盘驱动器。但是，当你建立一个大型的服务器群. 或是许多计算机要访问相同数据的时候，最好将磁盘驱动器或相关硬件安置在计算机的外部。

为了能够访问和记录那些磁盘驱动器上的数据，需要在计算机和磁盘之间使用网络。这种网络就称为存储区域网络。

以下列表概扫了存储区域网络的特性

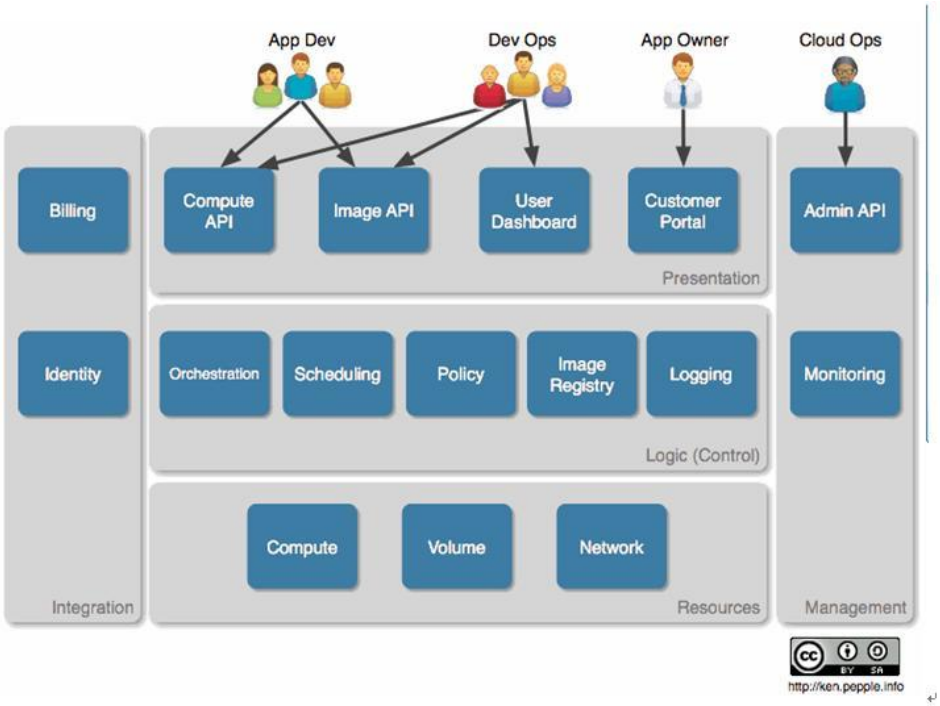
- 性能：存储区域网络支持两台或多台服务器对磁带或磁带队列的高速并行访问，这增强了系统性能；
- 有效性：存储区域网络通常在区外场所备份数据、常常超过 10 公里(6.2 英里) ，这大大增加了系统的有效性；
- 可扩展性——存储区域网络能够使用多种技术；这就使得系统间的数据备份、操作、文件转移和数据复制很容易实现重定向。

4.6. 分层设计

SmartCenter 帮助用户建立自己的 IAAS,提供基础设施服务给客户,为实现这一点，需要几个高级特性:

a. 允许应用拥有者注册云服务，查看运用和计费情况;
b. 允许应用开发者创建和存储他们应用的自定义镜像;
c. 允许用户启动、监控和终止实例;
d. 允许云的管理员配置和操作基础架构

以上四点都是 IaaS 的核心功能。这四个功能在 SmartCenter 中的示意图如下：



在此模型中，我们假设了需要与云交互的四个用户集：

应用开发者、应用管理员、应用用户、云管理员。

并为每类用户划分了他们所需要的功能。该架构采用的是非常普通的分层方法(presentation,logicandresources)，它带有两个正交区域展示层，组件与用户交互，接受和呈现信息。Webportals 为非开发者提供图形界面，为开发者提供 API 端点。如果是更复杂的结构，负载均衡，控制代理，安全和名称服务也都会在这层。

逻辑层为云提供逻辑(intelligence)和控制功能。这层包括部署(复杂任务的工作流)，调度(作业到资源的映射)，策略(配额等等)，镜像注册 imageregistry(实例镜像的元数据)，日志(事件和计量)。

假设绝大多数服务提供者已经有客户身份和计费系统。任何云架构都需要整合这些系统。

在任何复杂的环境下，我们都将需要一个 management 层来操作

这个环境。它应该包括一个 API 访问云管理特性以及一些监控形式(forms)。很可能，监控功能将以整合的形式加入一个已存在的工具中。当前的架构中已经为我们虚拟的服务提供商加入了 monitoring 和 adminAPI，在更完全的架构中，你将见到一系列的支持功能，比如 provisioning 和 configurationmanagement。

最后，资源层。既然这是一个 compute 云，我们就需要实际的 compute、network 和 storage 资源，以供应给我们的客户。该层提供这些服务，无论他们是服务器，网络交换机，NAS(networkattachedstorage)还是其他的一些资源。

5. 典型应用

5.1. ＩＤＣ公有云出租

公有云通常指第三方提供商为用户提供的能够使用的云，公有云一般可通过 Internet 使用，可能是免费或成本低廉的。这种云有许多实例，可在当今整个开放的公有网络中提供服务。

随着近年来云计算技术的逐步发展与普及，越来越多的用户开始选择可以提供按需、弹性模式交付计算，存储，IP 以及网络资源，同时可以支持按照实际使用情况付费的云服务提供商，这其中的佼佼者就是亚马逊公有云服务 AWS。

多年来亚马逊公有云服务是云计算领域一直是行业标杆。亚马逊

AWS 不断诱惑着企业 IT，其一系列的云服务为客户适应自身的 IT 环境提供了大量选择。而亚马逊凭借其公有云服务也在不断稀释着其他 IT 巨人在云端的份额。比如：微软(Azure)、谷歌(GCE)、Oracle 等。

公有云被认为是云计算的主要形态。目前在国内发展如火如荼，根据市场参与者类型分类，可以分为以下几类：

- 一类为传统电信基础设施运营商，包括中国移动、中国联通和中国电信；
- 一类为政府主导下的地方云计算平台，如各地如火如荼的各种“XX 云”项目
- 一类为互联网巨头打造的公有云平台，如盛大云；
- 一类为部分原 IDC 运营商，如世纪互联；
- 一类为具有国外技术背景或引进国外云计算技术的国内企业，如风起亚洲云。

由于目前国内并未开放外国公司在中国直接进行云计算业务，因此像亚马逊、IBM、Joyent、Rackspaces 等国外已有多年云计算业务经验的厂商在进入中国市场途中仍障碍重重。2012 年 11 月 1 日，微软终于实现旗下公有云计算平台 Windows Azure 在中国的落地，这将掀开外资企业进军中国云计算市场的序幕。

SmartCenter 以及后来即将发布的 SmartCloud 将会为国内的云计算市场部注入新的活力。

SmartCenter 针对国内ＩＤＣ企业的运营特点制定了"最小投资、平衡受益与安全"的云平台建设方案，首先在硬件层面采用服务器多

硬盘冗余、多网卡汇聚保证可靠性与性能;配置多交换机的汇聚冗余保证网络 i/o 的可用性;部署品 SmartCenter 的分布式存储模式，保证数据安全与高带宽,从底层保证公有云的可用性与性能。

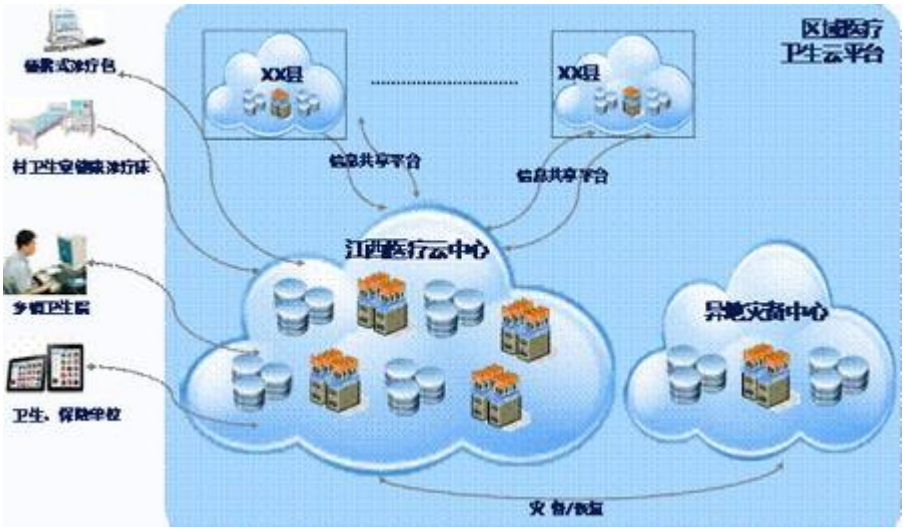
SmartCenter 能够以低廉的价格，提供有吸引力的服务给最终用户，创造新的业务价值，公有云作为一个支撑平台，还能够整合上游的服务（如增值业务，广告）提供者和下游最终用户，打造新的价值链和生态系统。

5.2. 医疗行业

医疗云是在医疗护理领域采用现代计算技术，使用“云计算”的理念来构建医疗保健服务的系统。这种医疗保健服务系统能有效地提高医疗保健的质量、控制成本和能够便捷访问的医疗保健服务。

这几年随着国家对卫生医疗领域的重视，对卫生医疗的信息化的资金投入在不断加大，随之惠及各个层面的信息系统应运而生。医疗信息化变革的时代恰逢云计算“风起云涌”，基于 SmartCenter 可以的“医疗云”解决方案，为中国基层医疗机构提供了一个借助信息化提升医疗行业管理能力的有效途径。

江西医疗云项目正是 SmartCenter 在“医疗云”建设之路上的真实案例，其作为江西省搭建在无锡城市云计算中心的医疗信息系统可使医疗信息扁平化，各个层面的人实现信息共享，从而有助于降低医疗成本，提高医疗整体服务水平。



从上图我们可以看到，江西医疗云中心的数据融化了下属市县、乡镇卫生院、村卫生室以及卫生所及保险单位的信息，以及提供了异地灾难备份中心，可有效保证数据的完整性。 通过将各市县及乡镇、村卫生室的数据连入江西医疗云中心，医疗机构的管理人员可通过数据中心了解联网内的所有入档的病人信息，病人的档案可随时调取，并能够监控各大小医院的治疗情况及用药情况。

目前江西医疗云项目覆盖了江西省宜春、赣州、南昌和九江四个市及县的一级医院，共计 4000 多个终端，将其部署在无锡城市云计算中心，有以下几大优势：

1. 数据安全。利用中心的网络安全措施，断绝了数据被盗走的风险;利用存储安全措施，使得医疗信息数据定期的本地及异地备份，提高了数据的冗余度，使得数据的安全性大幅提升。
2. 信息共享。将多个县市的信息整合到一个环境中，有利于各个部门的信息共享，提升服务质量。
3. 动态扩展。利用无锡云计算中心的云环境，可使医疗云系统的

访问性能、存储性能、灾备性能等进行无缝扩展升级。

4. 布局全国。借助曙光在全国各地的云计算中心，可使各地的一朵朵小云最终连片，形成覆盖全国的医疗云，医疗信息在整个云内共享，惠及更广大的群众。

“十二五”规划将医疗信息化提升到了新的战略高度，在国家大力推动和新医改背景之下，医疗行业信息化的盘子不可谓不大。据IDC 报告，2010 年我国医疗行业 IT 花费 114.1 亿元人民币，较 2009 年增长 26.7%;预测到 2015 年这一市场规模将达到 290.2 亿元。

面对如此大的市场诱惑，各大 IT 厂商也纷纷摩拳擦掌。相信未来，有 SmartCenter 的助力，中国的云计算不再让老百姓“雾里看花”，而更多的是切身感受云计算所带来的便捷生活。

5.3. 数字校园与高校实验室

创建节约型校园，是建设资源节约型、环境友好型的社会迫切需要，是共同应对我国资源相对紧缺，生态环境脆弱的唯一出路，是高校科学发展的必然选择。无锡市江南大学走在了创建数字化节约型高校的前列，优化资源配置体系，提高资源使用效益是创建数字化节约型校园的必要手段之一。

但是传统的数据中心纵向结构的建设模式，硬件和操作系统完全绑定，使得服务器之间无法复用计算资源，只能通过为不同业务单元

分别堆加服务器来满足业务要求，随着规模发展，硬件利用率低下、管理复杂、响应速度滞后，运行成本居高不下等问题正逐渐显现。

云计算技术的出现，在快速响应和节省成本之间找到平衡点。服务器虚拟化使得操作系统不再直接安装在硬件上，形成了逻辑层和物理层分离的横向结构，不仅可以方便地复用硬件资源，管理效率也大大提高。同时云计算结合服务器虚拟化、应用虚拟化和流技术，提出了新一代动态数据中心的建设模式，能够根据不同业务模块的资源消耗，自动地分配硬件资源，从而最大限度满足数据中心的高效率、高性价比和自动化管理等要求。

新系统改变了原有 IT 项目建设模式，其目标和意义在于：

1. 减少重复投资，优化资源分配：

   - 需要灵活的系统环境，根据不同的应用，数字化校园需要有跨平台环境，这就需要不同系统环境支持，例如 Windows 或 Linux 等多样化环境。为了提高服务质量和加快数字化节约型校园的发展，高校系统建设要达到国内高校先进水平，为将来百年发展奠定基础。

   - 按需动态分配的系统资源，实现资源聚集和共享的数字化利用，避免重复建设。高校应用有着受众范围广、使用时间规律化的特点，为了防止服务的中断，传统往往都是按照峰值进行配置以防止突发情况可能对系统带来的影响。导致平均利用率较低（10%-30%），而集中配置和分配系统资源将可以提高系统的利用率，通过动态调整系统资源把空闲资源或新采购的

资源迅速补充到不同的应用中。硬件资源不需要一次采购，而是可以按照使用量进行动态调整。

- 集中监控和运维管理，通过对数据中心的统一监控和运维管理来提高整个系统的可靠性和服务质量，降低了对维护技术人员的要求。

- 需要方便易行的安全管理手段，减少安全管理复杂度，例如，系统安全软件需在大量的服务器中安装，不仅费时费力，而且版本和补丁管理也繁琐低效，数据中心可以通过更加便捷高效的手段进行安全管理。

2. 加强统一管理，提高系统标准化

- 对应用环境进行统一的管理，可以减少管理和维护的投入，通过标准化提高管理能力，从而保障充足的、可动态分配的系统资源。统一的数据中心可以避免系统的多样性和管理维护的高成本，从而可以基于统一映像来提高系统部署的能力。

3. 降低能源开销，万向绿色 IT:

- 通过共享计算资源来降低能源开销，统一的数据中心平台基于需求来动态分配计算资源，同时监控能源开销，这帮助高校管理者了解数据中心的能源使用，从而向绿色 IT 迈进。

基于“云计算”的数据中心服务平台是满足高校实现节约发展、科学发展、可持续发展最佳途径。

6. 结束语

在当今的全球竞争市场中,企业必须创新并最大程度利用资源才能取得成功。这就要求为员工、业务合作伙伴和用户提供能推动创新的平台和协作工具。云计算基础是下一代平台,可以为各类规模的企业带来巨大价值。它们可以帮助企业更有效地利用其 IT 硬件和软件投资,并提供加快创新采用速度的手段。云计算通过提高资源利用来增强盈利能力。由于仅在需要时才提供资源,使得成本得以降低。云计算帮助团队和企业组织精简冗长的采购流程。

云计算使创新者无需费力寻找资源以开发、测试并向用户社区展示其创新。创新者可以全神贯注于创新,而不必理会为了支持创新而寻找和管理资源这类琐事。云计算和 SmartCenter 的结合提供了全面的协作环境,使企业转变为创新的源泉。

智网科技是国内云计算和创新技术的引领者和实践者。通过 SmartCenter ,员工或者用户能够迅速获得计算资源。智网科技 SmartCenter 提供托管的生态系统或本地安装的解决方案。

7. 附录 A 术语

1. SDN
2. IAAS
3. SmartAPI

4. SmartVM
5. LAMP
6. IntelVT/AMD-V

   Intel VT 和 AMD’s AMD-V 是一套和支持该技术的虚拟机监视器相结合的硬件增强特性(指令集扩展)

7. 租户

8. 附录 B 参考资料