

日期：2012年5月31日星期四

Huawei Enterprise **A Better Way**

数据中心网络总体设计方案

www.huawei.com/enterprise

HUAWEI TECHNOLOGIES CO., LTD.

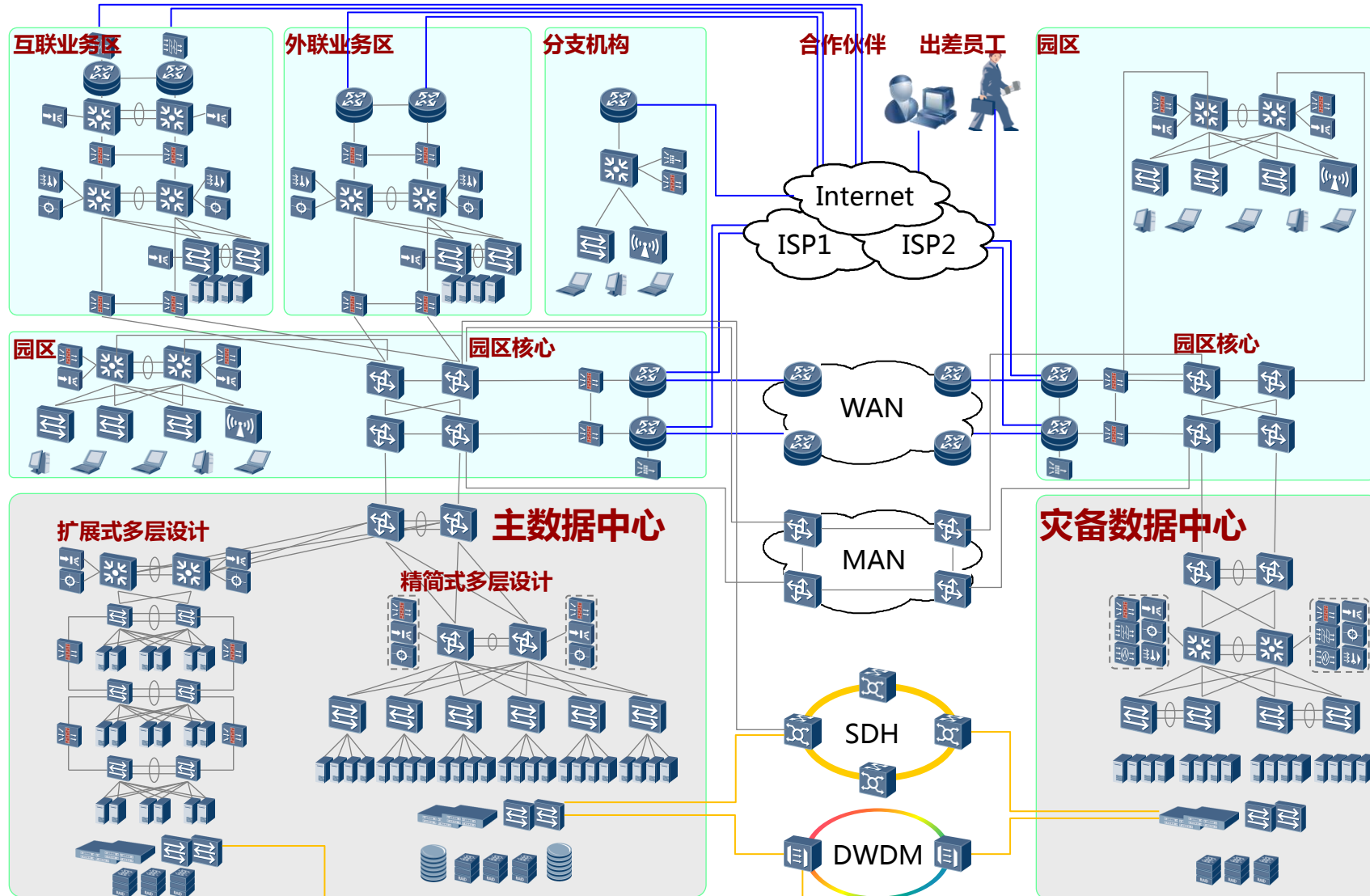


目录

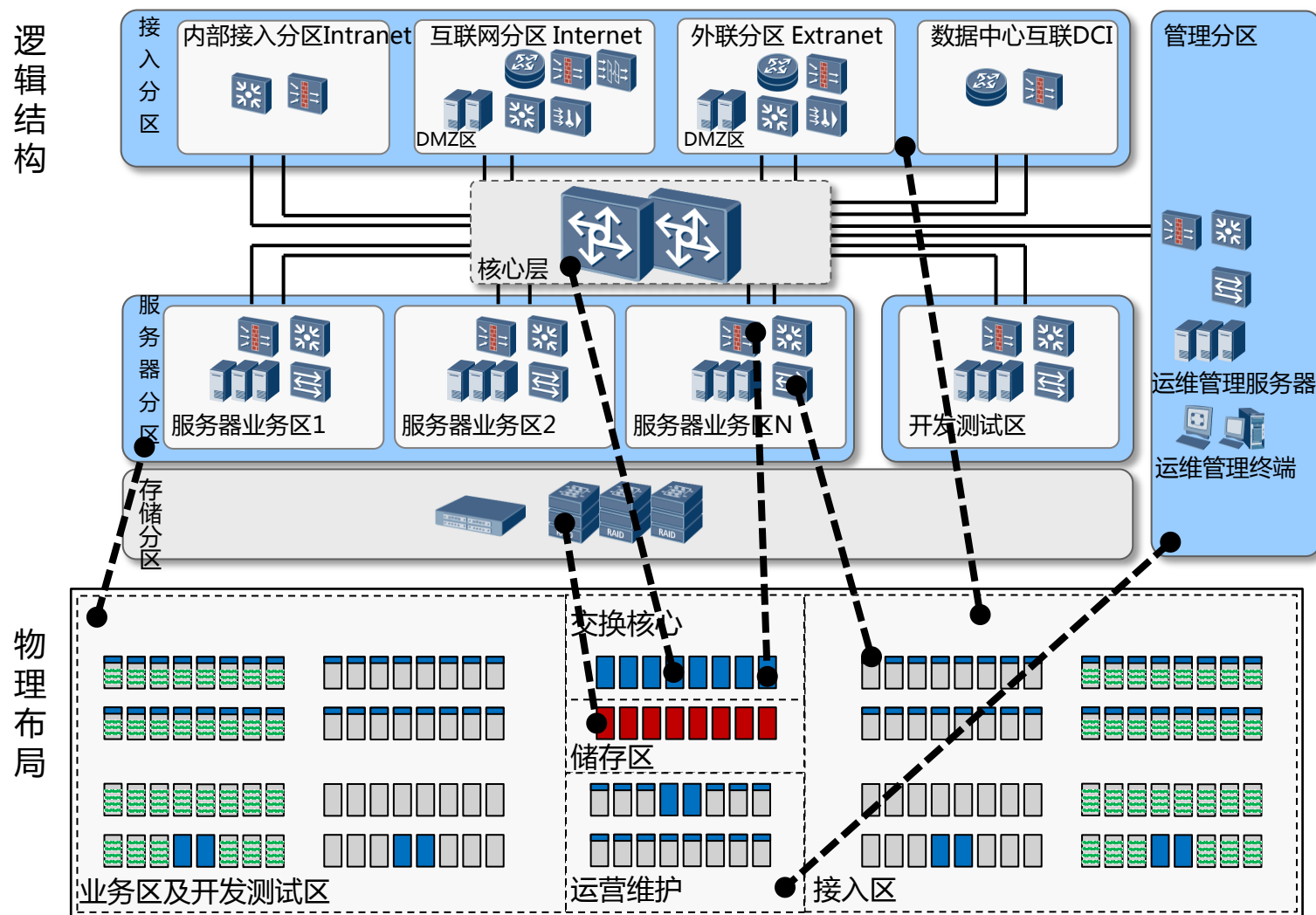
数据中心网络全貌

数据中心网络总体设计

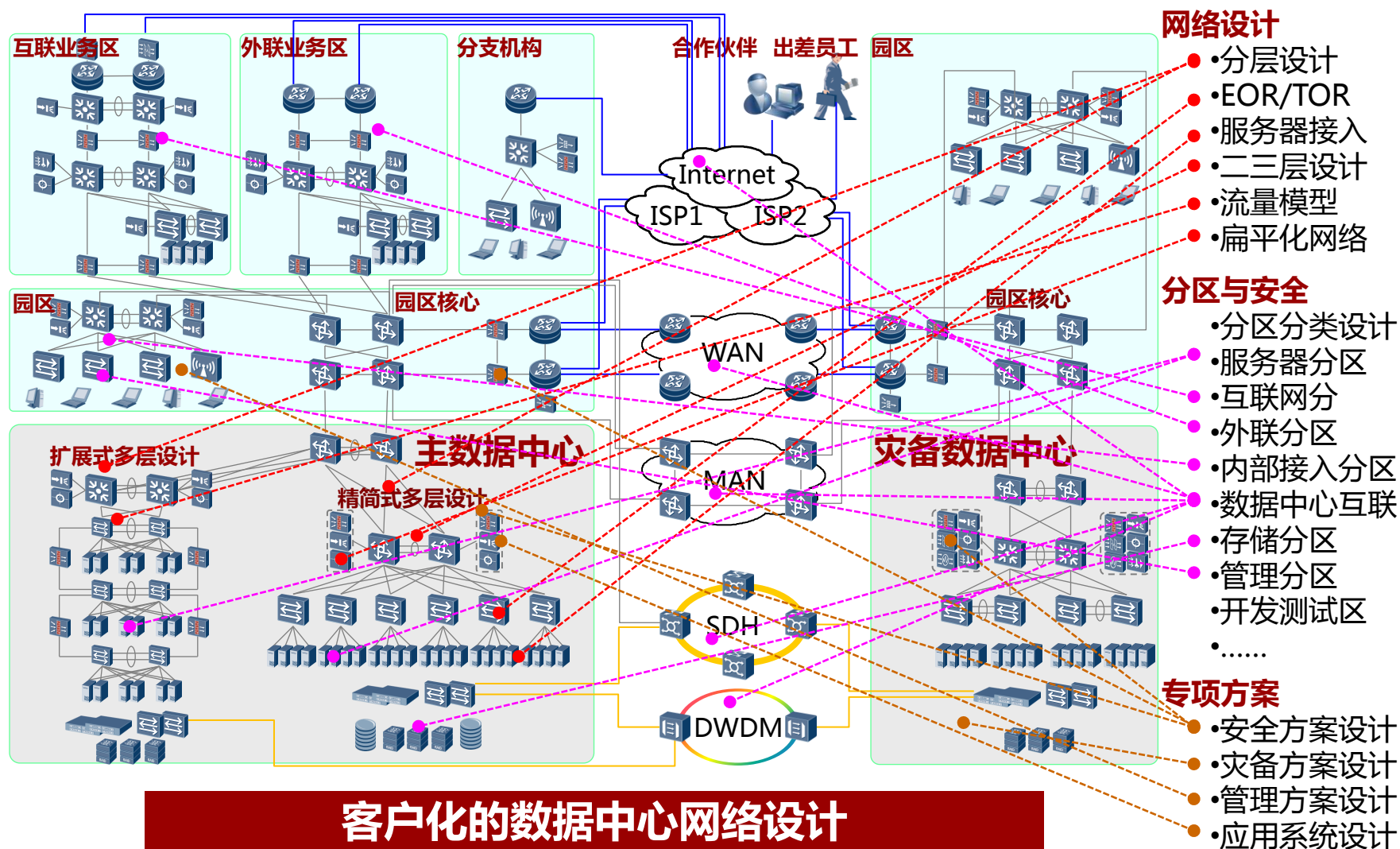
数据中心网络解决方案全景架构概图



数据中心逻辑结构和物理布局



数据中心网络解决方案设计全景图



目录

数据中心网络全貌

数据中心网络总体设计

分层设计

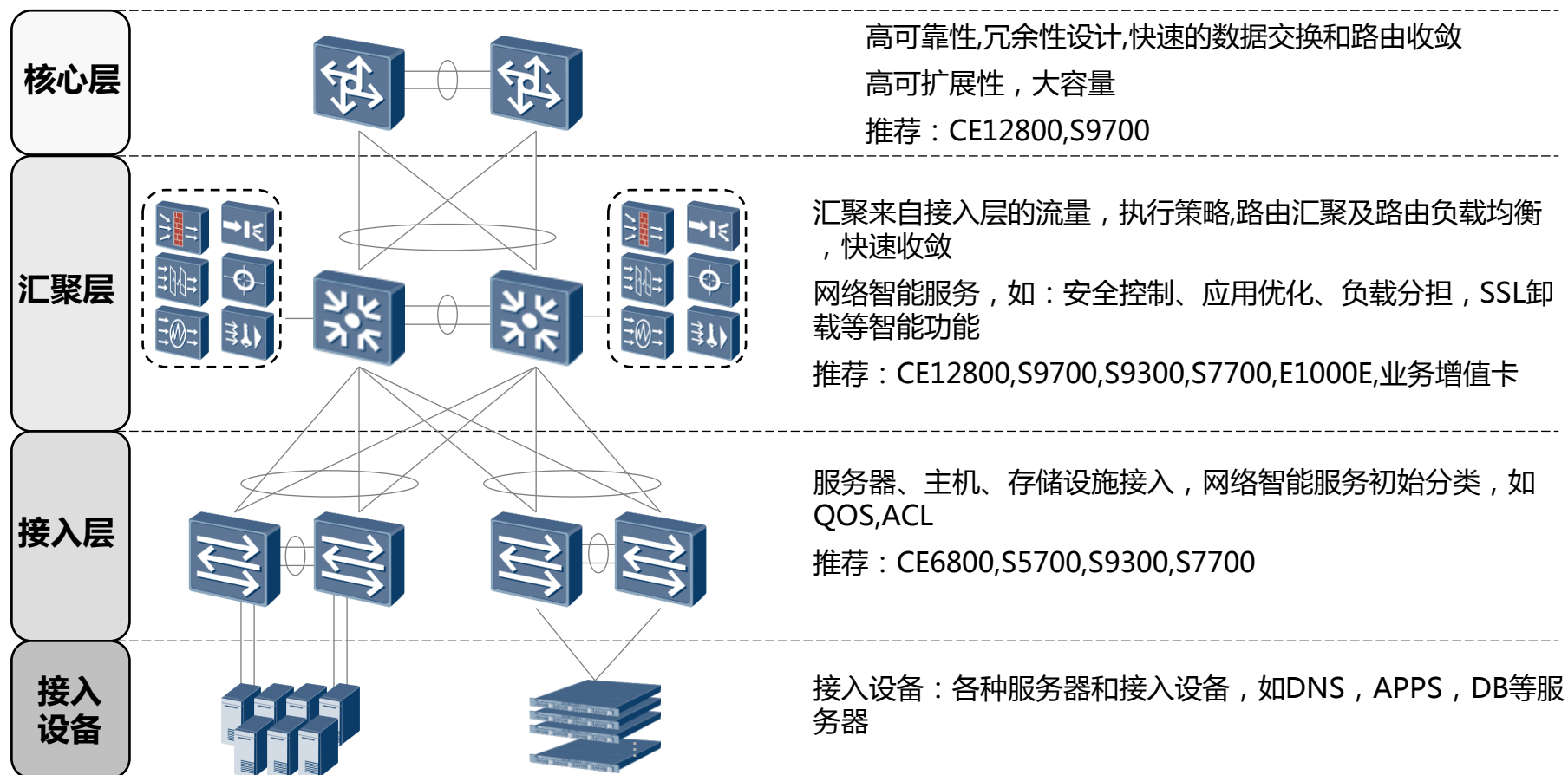
服务器接入设计

网关及路由设计

流量模型分析

扁平化网络设计

网络层次化设计



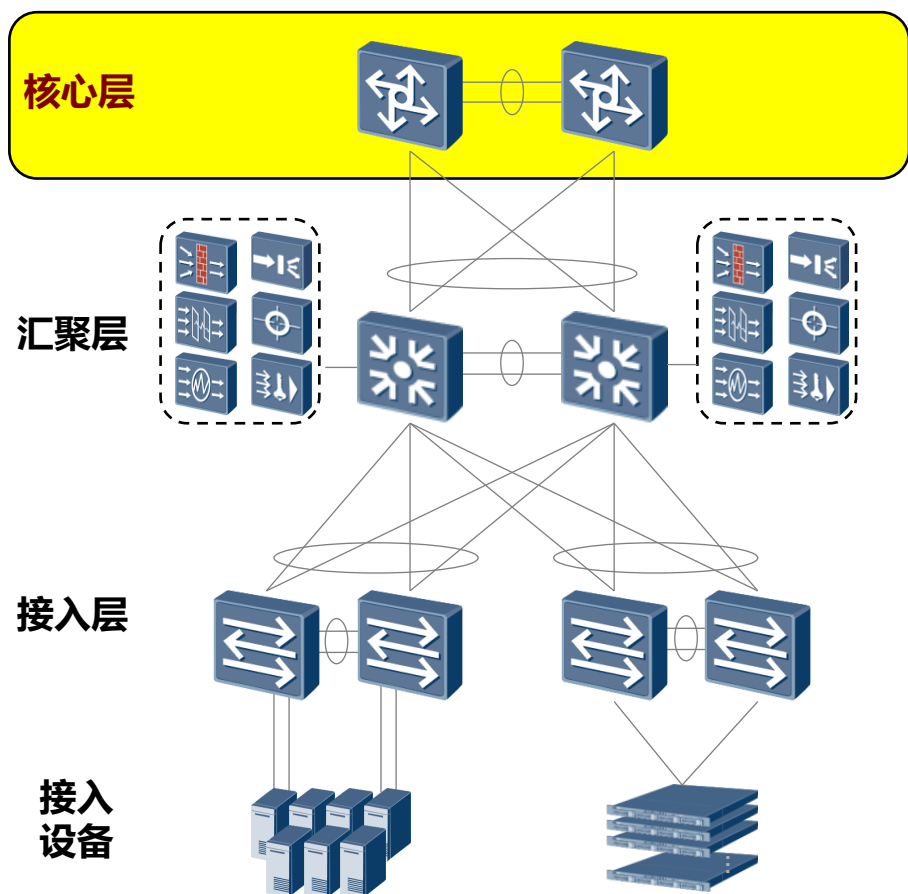
层次化设计的优势

可靠性高:汇聚区设备独立,冗余设计,避免单点故障

易于扩展:易于模块化设计,汇聚区间相对独立,易于业务扩展,故障隔离和故障定位

方便管理:便于部署路由,安全控制策略,

核心层设计



设计目标

核心层网络是数据中心的核心
连接数据中心内部各个功能分区，如互联接入区，服务区，内部接入区等
是数据中心内部总线。

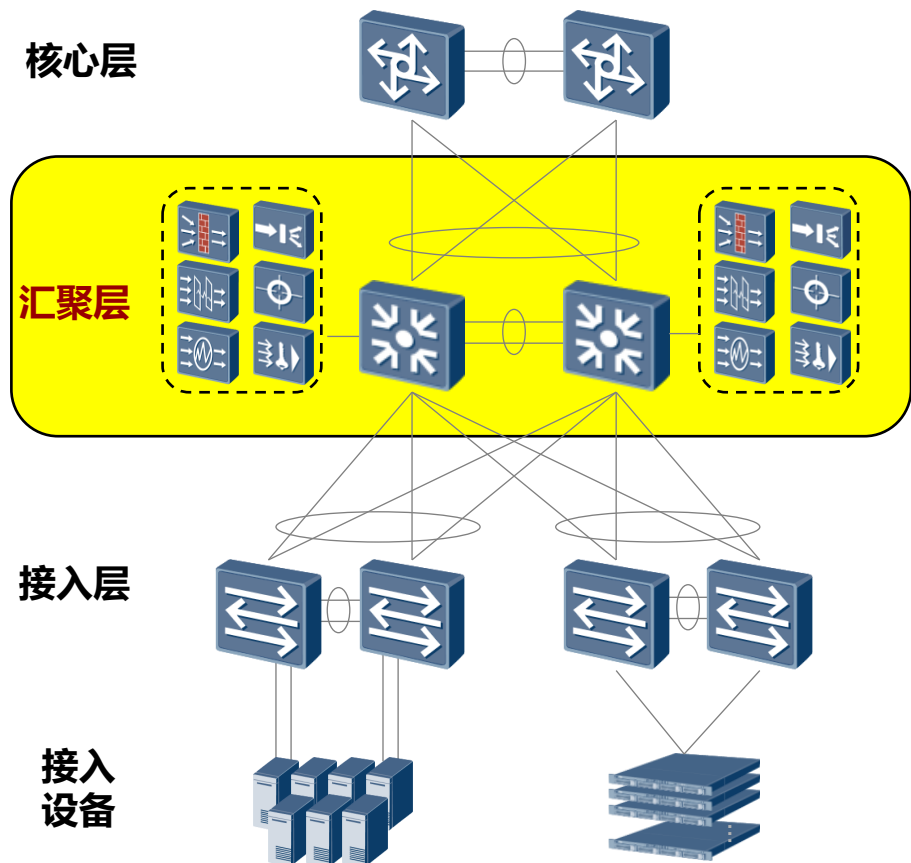
设计要求

高性能快速转发;高密度10GE/40GE连接
高可靠性/可用性,支持不间断转发
可扩展性高,满足数据中心业务扩展
收敛比尽可能小
较高的稳定性

部署方案

核心由两台高性能交换机组成
对于更高可靠性可使用4台交换机组成双平面

汇聚层设计



设计目标

汇聚层连接数据中心核心网络
 汇聚区内部的東西向流量转发
 服务器到数据中心外部的南北向流量转发。
 汇聚层一般作为业务网关，是业务和安全策略的控制点。

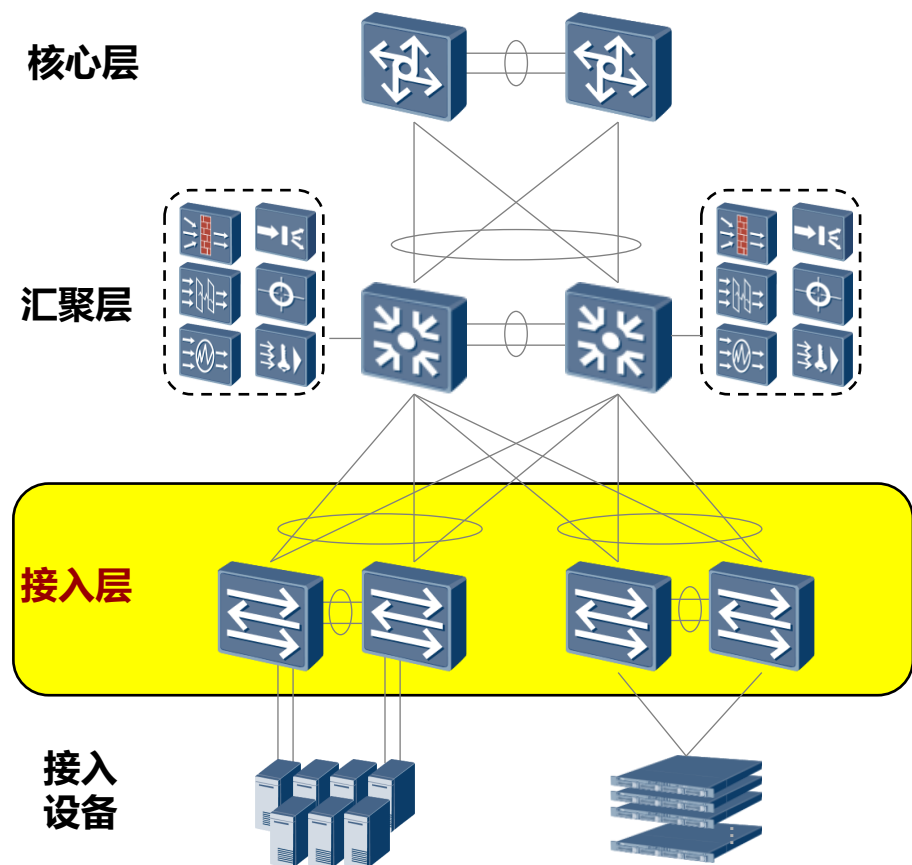
设计要求

强转发能力，高端口密度
 可靠性是最重要的特性，如三层支持NSF，二层支持Smartlink
 丰富的二、三层特性，承载多业务，VRRP、OSPF等

部署方案

两台或多台高性能交换机组成
 网络智能服务，如：应用优化、负载分担，SSL卸载等

接入层设计



设计目标

接入层连接服务器，主机，存储等设备的接入，上行到汇聚交换机

提供基本ACL，QOS分类等网络服务

设计要求

丰富的二层特性：VLAN、IGMP Snooping、环网避免

安全特性：端口安全、DHCP snooping、DAI、IP Source Guard，MFF安全

可靠性：系统级的引擎和电源冗余，LAG，iStack等

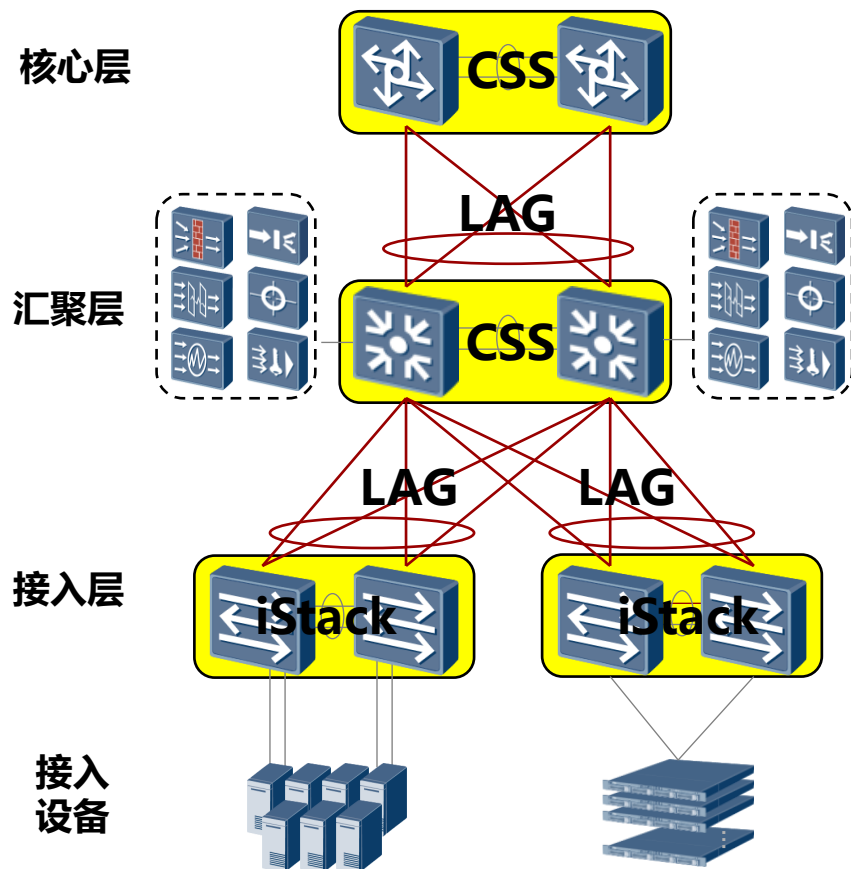
带宽管理：QoS等

高密度GE/10GE接口，接入交换机数量大，易管理

部署方案

一般双归到汇聚层，部分区域接入和汇聚合并
服务器接入部署可采用EOR/TOR等方式

高可用设计：LFR无环以太网



高可靠的物理和逻辑拓扑

解决方案

通过堆叠 (iStack) / 集群 (CSS) 技术保证节点的可靠性；一台设备故障后，另外一台设备自动接管所有的业务

CSS + LAG + iStack 端到端可靠性架构打造无间断数据中心，保证业务持续运营

多台接入层堆叠，2台汇聚层集群，2台核心交换机集群

方案特色

无环以太网简化网络架构，收敛时间远优于xSTP，收敛时间达到ms级

提升链路利用率，采用逐流方式负载均衡，支持不同负载分担方式，二层报文，IP报文。

转发支持本地优先转发策略，减少设备间转发

配置简单，不易引入配置故障，不需要配置多数可靠性的协议，如VRRP等，减化配置和维护工作量，减少出错的机率

链路的可靠性保证，通过Trunk技术，一条或多条链路故障后，流量自动切换到其他正常链路

CSS(Cluster Switch System) 交换机集群系统

目录

数据中心网络全貌

数据中心网络总体设计

分层设计

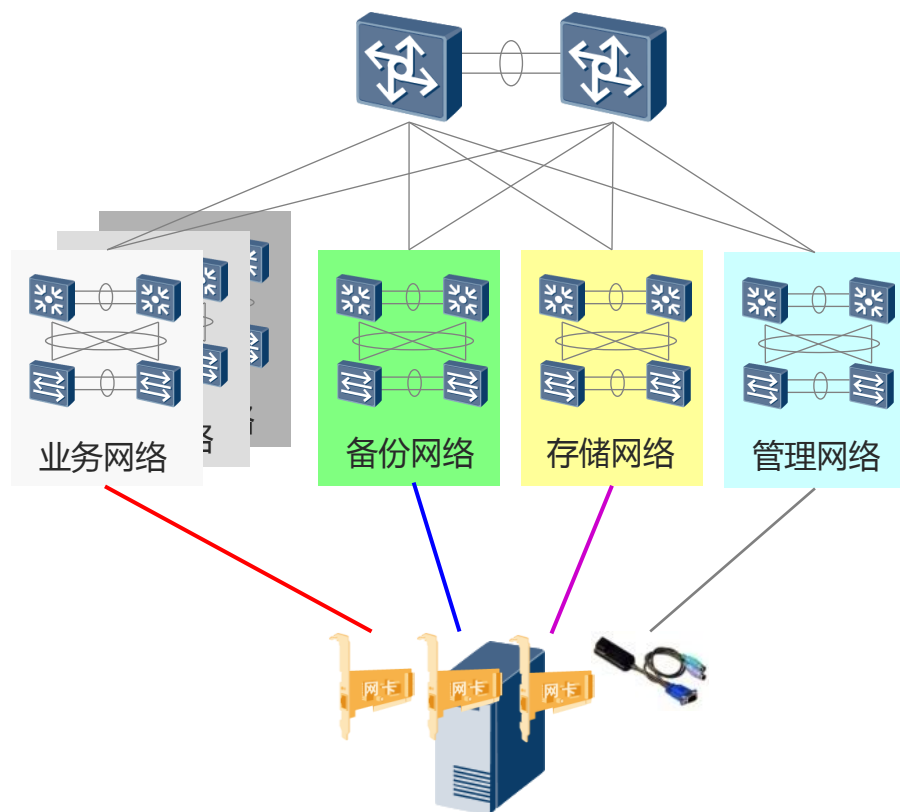
服务器接入设计

网关及路由设计

流量模型分析

扁平化网络设计

服务器多通道接入多个网络



服务器的多通道，隔离和提高IO能力

服务器服务器包括四类接口，分别接入到业务网络、网管/KVM网络、SAN网络、备份和IP存储网络

服务器采用多通道的方式，采用物理隔离的多个网络接口，充分发挥CPU的高性能

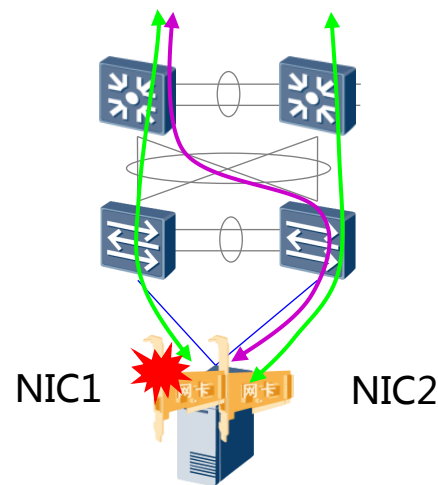
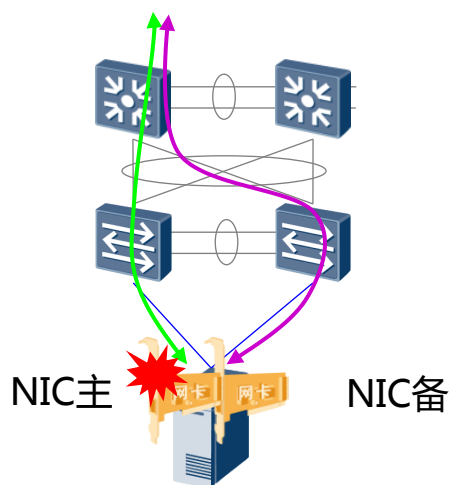
四网多通道分离优点

整网按区域分为多个业务网络

整网一个管理网络，一个备份网络，一个IP存储网络

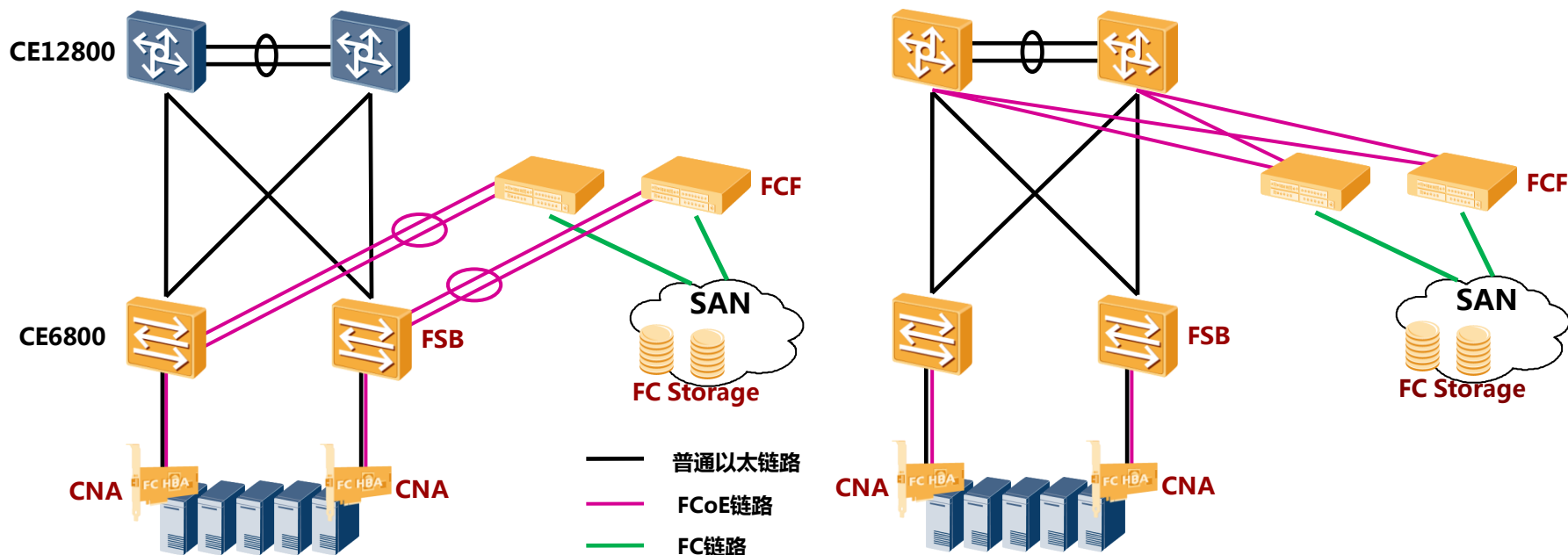
业务网络、管理网络、存储网和备份网络流量分离，服务器通过不同的网卡分别接入不同的网络，提升网络整体可靠性。

服务器冗余接入设计



冗余方式	主备方式	负载分担方式
正常情况	两个网卡的MAC相同 流量在主网卡转发	两个网卡的MAC相同 两个网卡都可以发送和接收流量
异常处理	故障时，备网卡发出免费ARP，网络设备必须正确处理这个免费ARP报文，才能将发给服务器的流量切换到新的转发路径上 主网卡故障后，转发路径需要从绿色曲线切换到紫色曲线	没有故障时转发路径时绿色曲线，两个网卡都有流量。左边网卡故障后，转发路径需要从绿色曲线切换到紫色曲线。

服务器接入融合设计



	接入存储融合	核心存储融合
应用场景	TOR支持FCoE/FC接口和DCB无丢包以太网 存储流量在TOR直接分流到SAN存储网络 范围较小，占用FCoE端口较多	TOR/核心都支持FCoE/FC接口和DCB无丢包以太网 存储流量在核心分流到SAN存储网络 范围较大，FCoE端口较少

目录

数据中心网络全貌

数据中心网络总体设计

分层设计

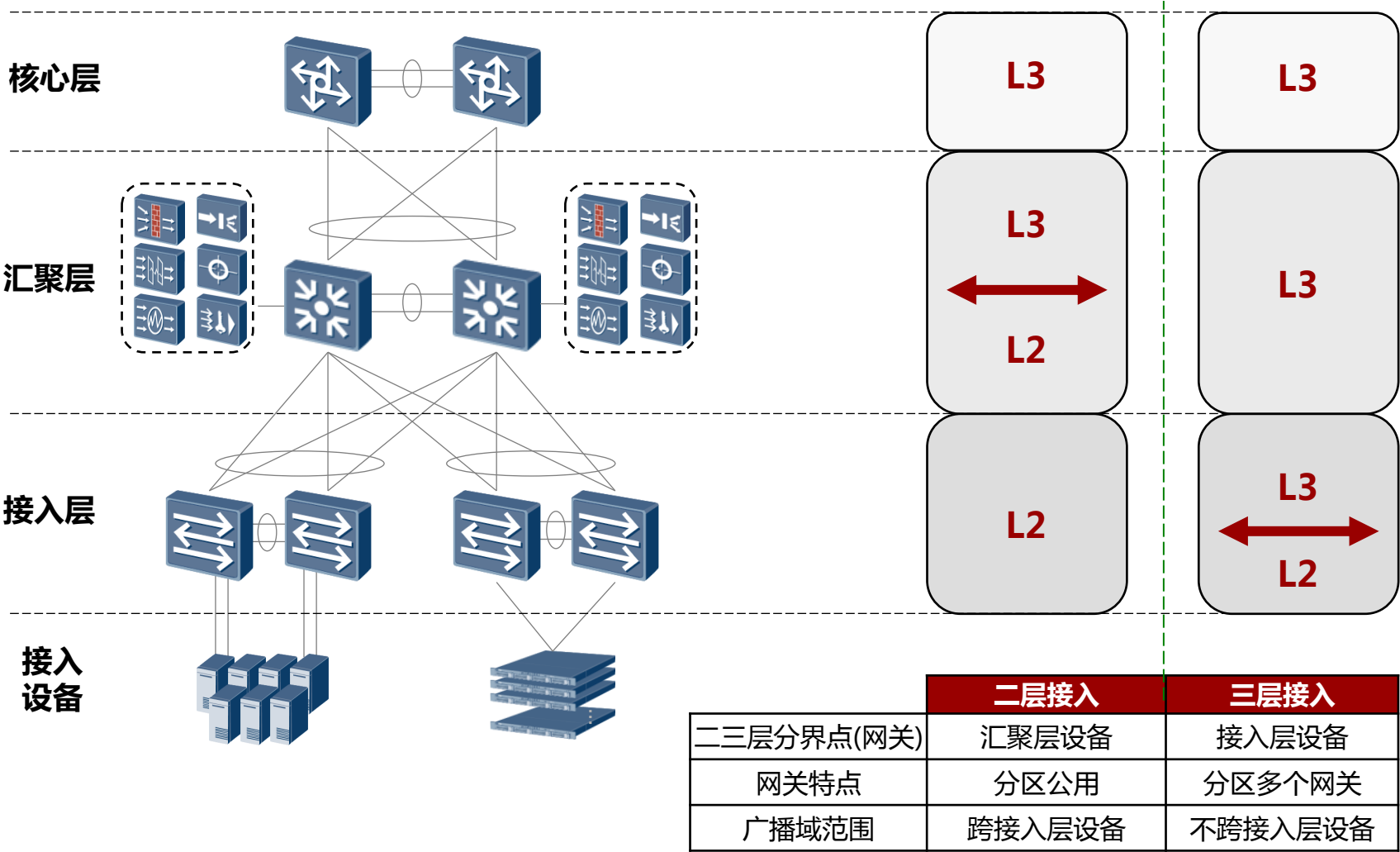
服务器接入设计

网关及路由设计

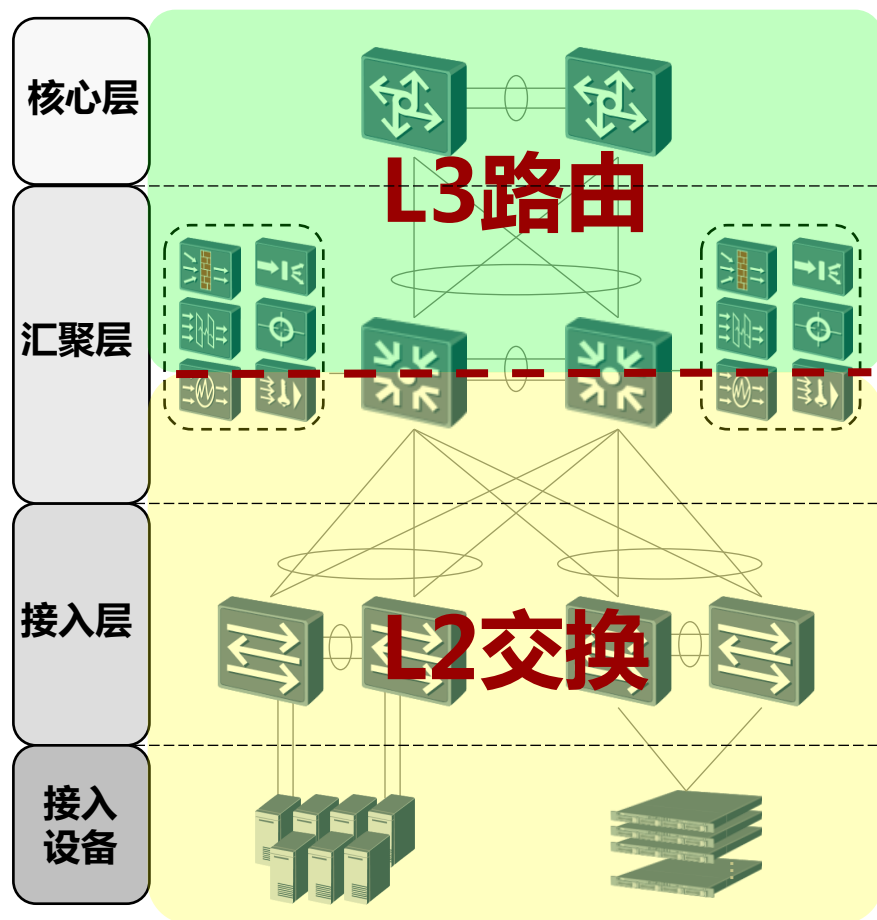
流量模型分析

扁平化网络设计

二三层分界点设计方案



二层接入



方案

汇聚设备作为二三层网络的分界点，作为服务器业务网关

汇聚、核心设备采用三层路由互联，通过IP FRR做快速路由收敛

接入设备仅仅为用户提供二层接入功能，根据服务器分组的具体情况划分VLAN

接入层到汇聚层通过设备堆叠、集群，采用链路汇聚保证设备间可靠性.通过MSTP防止错误配置或连线引起的网络环路

一般情况下推荐本方案

优点

低成本，接入侧交换机采用二层交换机，保护和节省用户投资.

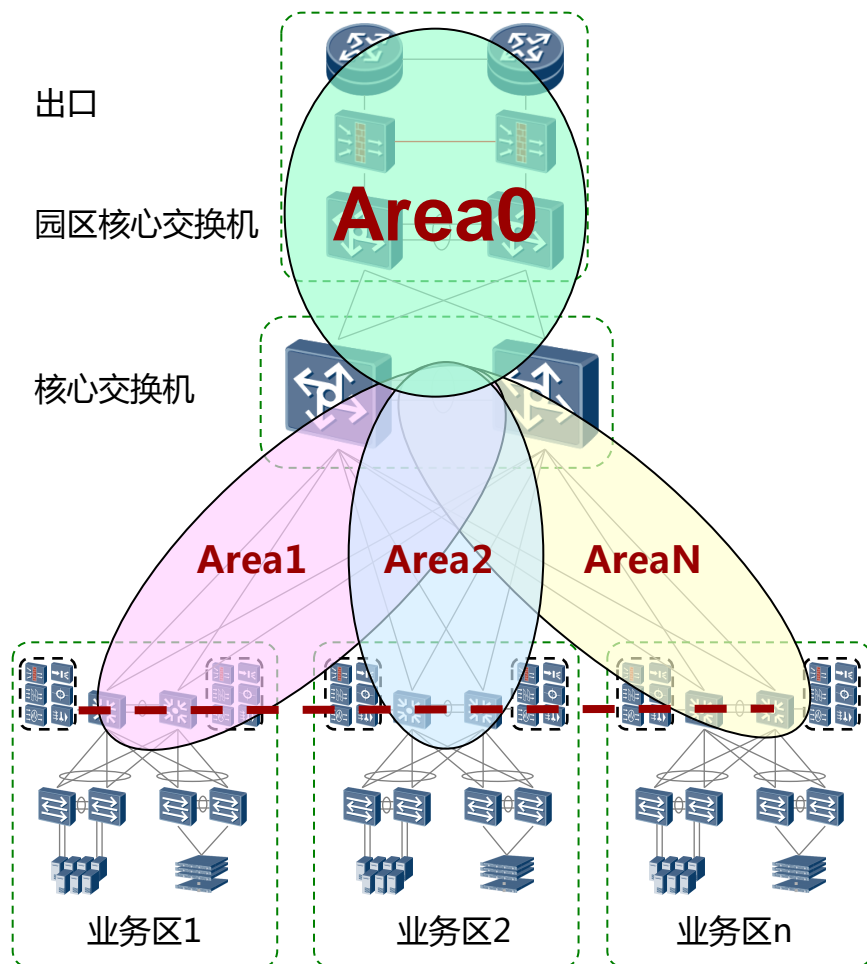
满足分区内业务的二层互通需求

缺点

二层网络内的故障定位难度较大

二层广播域范围较大，增加了流量管理的复杂度

二层接入方案IGP设计



路由设计

方便管理和维护，建议采用OSPF动态协议

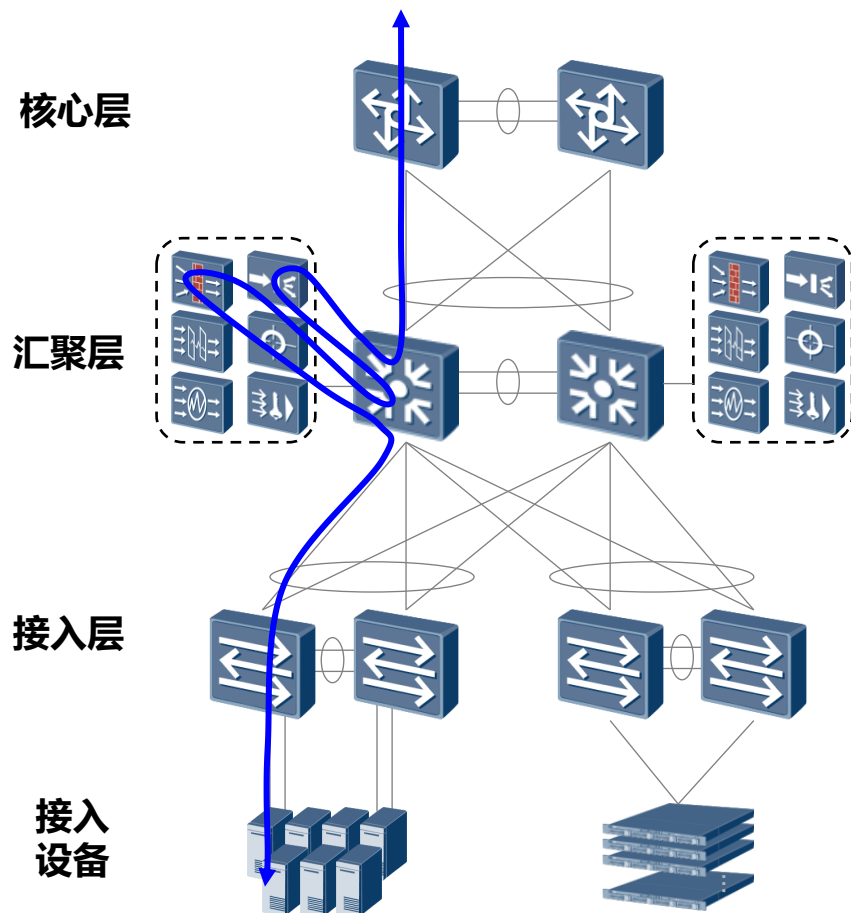
核心交换机(和园区核心或者出口路由器)和防火墙组成骨干区域 (Area 0)

每个分区的汇聚交换机和核心交换机，防火墙，组网部署为不同的OSPF区域 (Area ID 1,2,N) 。

Area 1,2..N 使用OSPF NSSA 区域，限制LSA在区域间的传播

若分区数量少，建议只配置Area 0

二层接入方案业务服务设计



FW和LB设计

FW和LB旁挂在汇聚交换机，推荐运行三层模式

对于可信的业务，如可信分区间流量，不需要通过防火墙，业务网关部署在汇聚交换机

需要负载均衡的业务

外部流量策略引流到防火墙，处理完成后回注到LB设备，LB选择业务服务器

LB采用非对称部署方式，业务网关部署在LB上，否则部署在防火墙。

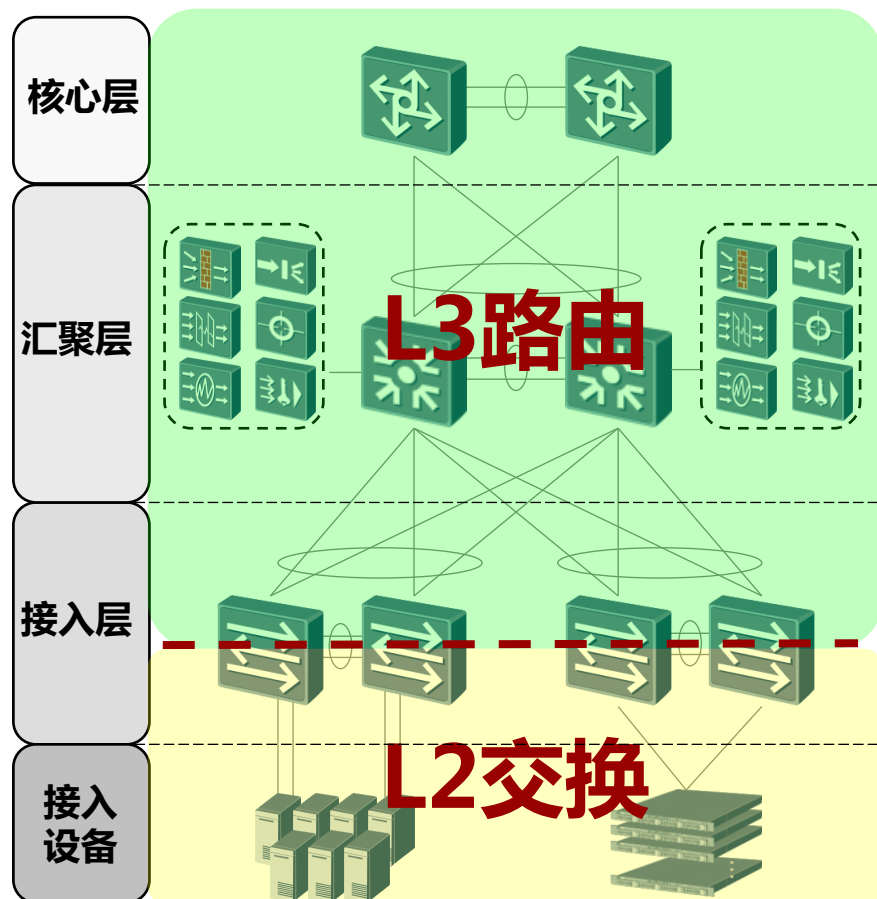
不需要负载均衡的业务

外部流量策略引流到防火墙，处理完成后回注到交换机。

业务网关部署在防火墙

其他业务与FW和LB组合方式类似

三层接入



方案

接入设备是二三层网络的分界点，接入设备作为服务器业务网关，提供二层终结，三层路由。

服务器采用VLAN方式接入网关。

全网采用IP FRR做快速路由收敛

优点

纯三层结构，网络结构简单清晰，不依赖CSS等技术简化网络

扩展性强，网络拓扑依赖度低，可以任意网络拓扑形式扩展

易维护，无二层环路网络风险，无需配置生成树协议

易配置，无需规划二层配置

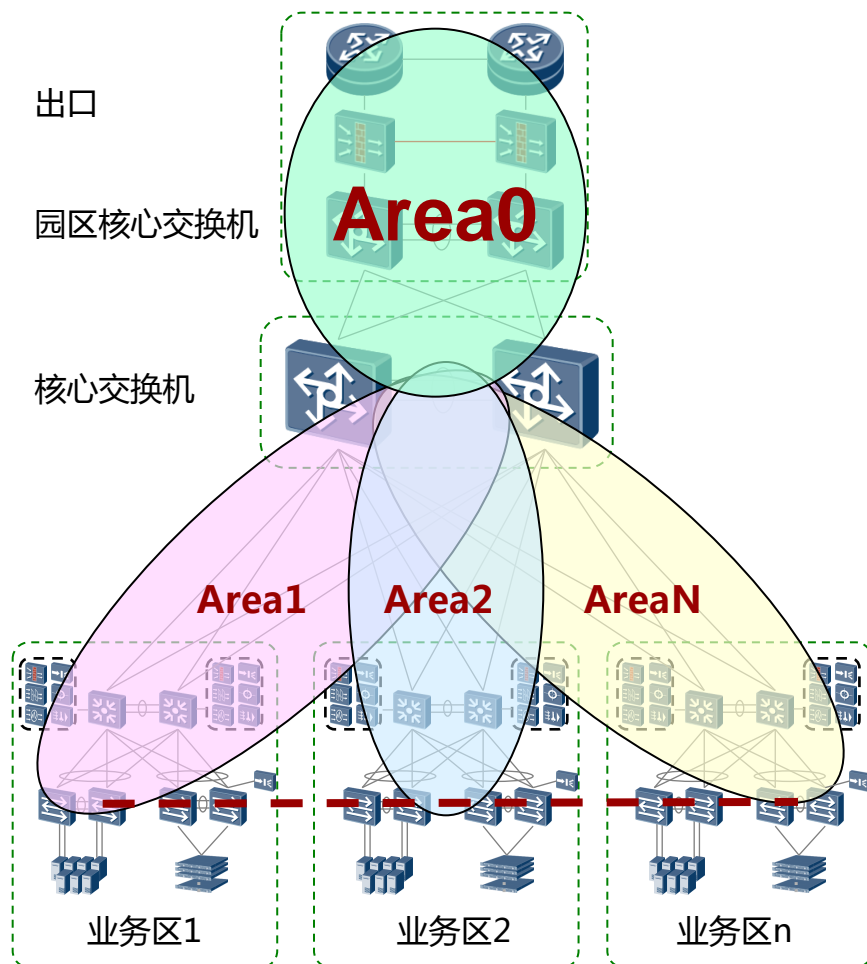
缺点

对接入交换机要求较高，导致成本提升

不同接入交换机下的服务器无法进行二层互通，某些特殊业务无法运行

收敛速度相对二层可能会略慢

三层接入路由设计和业务服务设计



路由设计

核心交换机和园区核心(或者出口路由器)和防火墙组成骨干区域 (Area 0)

每个分区的汇聚交换机和核心交换机, 防火墙部署为不同的OSPF区域 (Area ID 1,2,N)。

Area 1,2..N 使用OSPF STUB 区域, 限制LSA在区域间的传播

业务部署

LB设备部署在接入层

业务网关部署在接入交换机或者LB设备

防火墙部署在汇聚层, 使用路由模式

目录

数据中心网络全貌

数据中心网络总体设计

分层设计

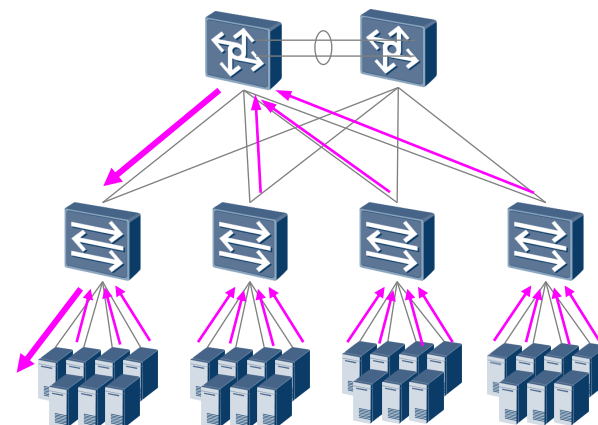
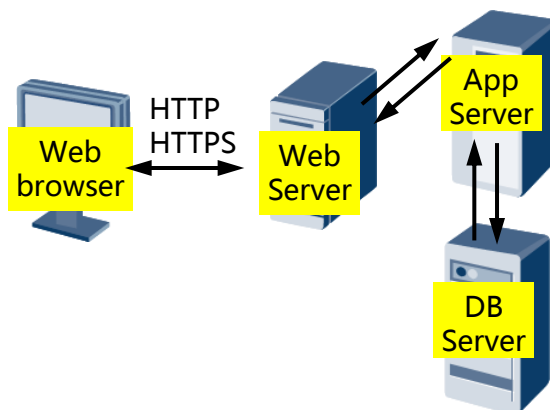
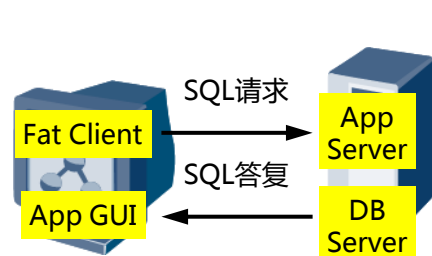
服务器接入设计

网关及路由设计

流量模型分析

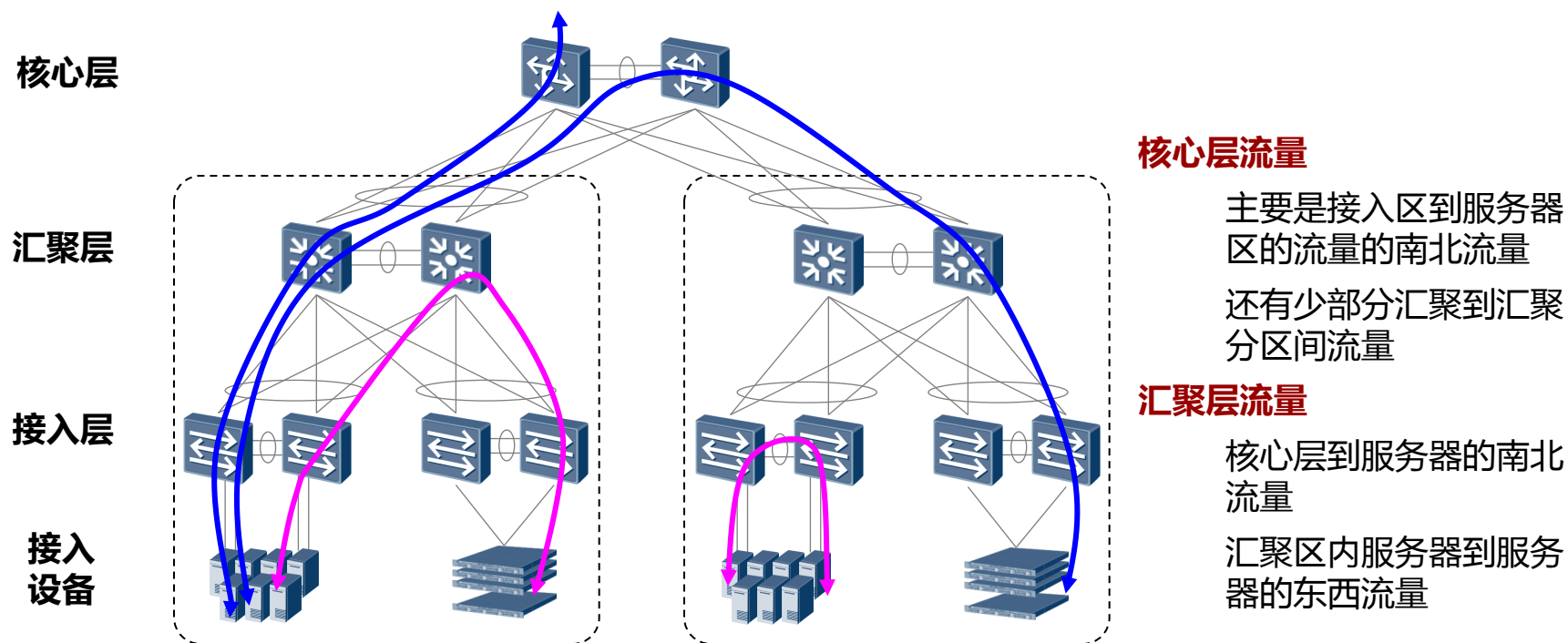
扁平化网络设计



典型业务模式



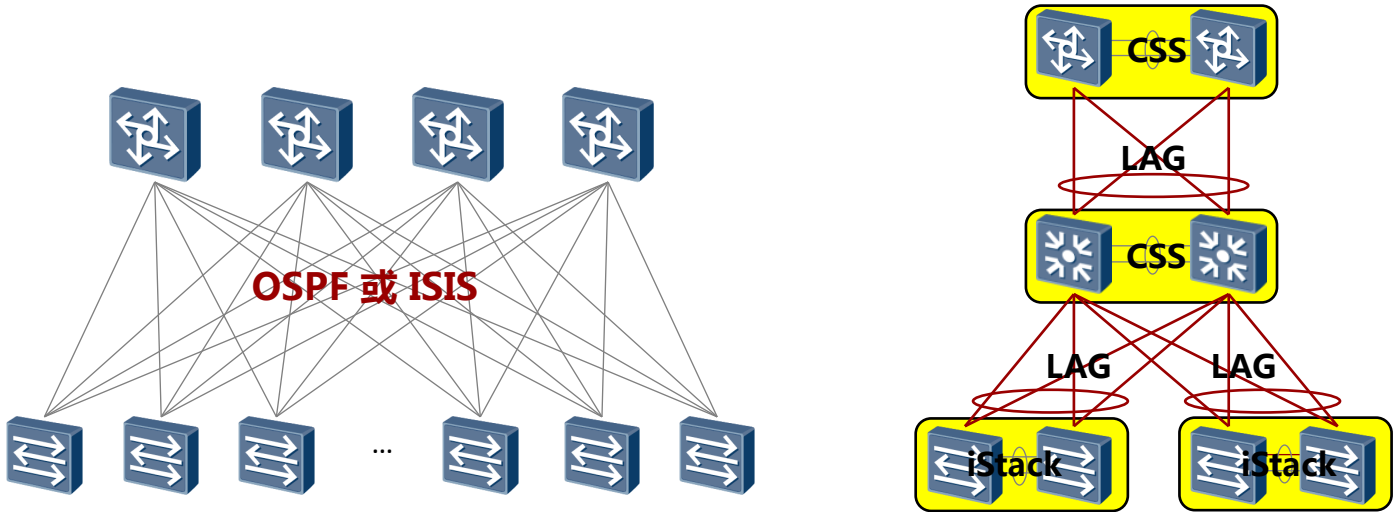
	客户/服务器模式：C/S模式	N层模式：Web业务	集群模式：协同计算业务
模型	One-Tier 或 Two-Tier	Three-Tier 或者 N-Tier	Fat-Tree
描述	前端是客户机，通常是PC或其他终端 后端是服务器，部署在数据中心，服务器通过本机存储或独立的存储设备存取数据。	业务处理分工为：界面（Web服务器负责完成）、业务处理（应用服务器负责完成）和数据库（数据库服务器和存储系统负责完成）	流量主要是服务器之间的流量，调度服务器将计算业务分发给大量计算服务器处理，计算服务器将处理结果返回给调度服务器
优点	客户机本地化处理信息	灵活性、安全性和可扩展性好 业务系统的发展变化可以在Web或应用服务器上实现，客户端改变少	服务器和网络利用率高，网络可扩展性强
缺点	可扩展性、灵活性和安全性较差	时延较长	对网络时延和带宽要求高，易形成瞬间峰值流量
典型业务	有文件存储、邮件系统、ERP	Web、互联网访问、电子商务等	3D渲染、药物研究、基因分析、Web搜索业务

数据内部流量分布



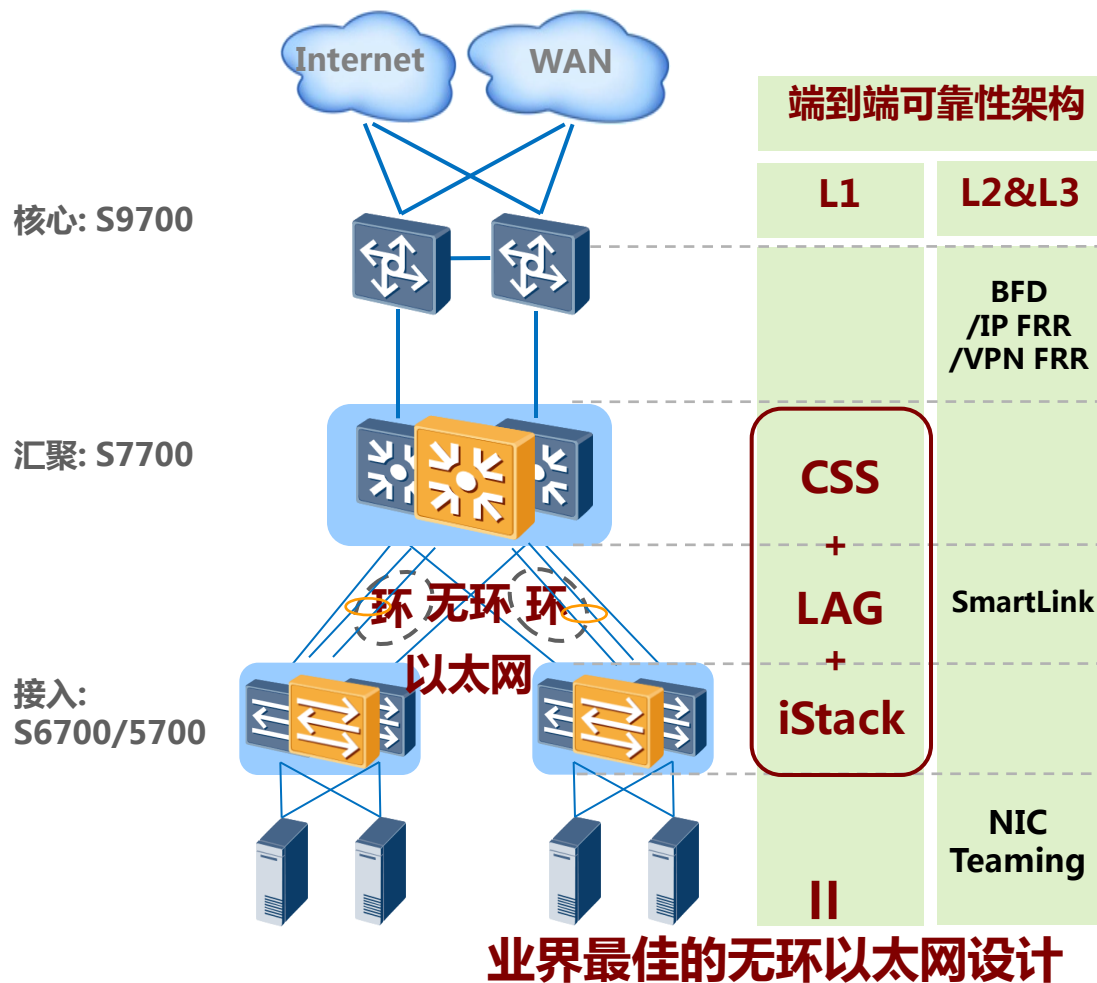
	南北流量	东西流量
流量特点	数据中心内部服务器与外部通信的流量，如客户端和服务器之间的数据请求和应答 也称为C/S流量	数据中心内部服务器之间流量， 又称为S/S流量
流量示意	蓝色曲线 	紫色曲线 

流量模型：负载分担设计



	L3负载分担	L2负载分担
分担特点	流量均衡无阻塞，收敛比1:1 ~ 2:1 通常用扁平化多路径网络 运行OSPF或ISIS，形成IP等价路由 流量按五元组进行Hash，实现逐流负载分担， 也可配置为逐包负载分担方式	多台设备采用堆叠(iStack)+集群(CSS)组网，实现设备的负载分担 链路捆绑（Eth-trunk）提升链路利用率 逐流方式负载分担 本地优先转发，尽量不通过堆叠线缆转发
优点	低时延，无阻塞	高可靠，高带宽
应用场景	hadoop等SaaS、PaaS云计算环境	无环以太网，无间断数据中心

无环以太网技术支撑典型流量模型



LFR无环以太网技术

- 端到端200ms可靠性架构打造无间断数据中心，保证业务持续运营
- LFR “无环以太网” 简化网络架构，收敛时间远优于传统STP，且100% 利用网络带宽（STP为50%）
- “CSS + LAG + iStack” 构建业界最佳的快速无环以太网
 - 业界最可靠的物理层“硬”集群技术
 - 业界最快的200ms级收敛能力
 - 业界最大的256G集群带宽

目录

数据中心网络全貌

数据中心网络总体设计

分层设计

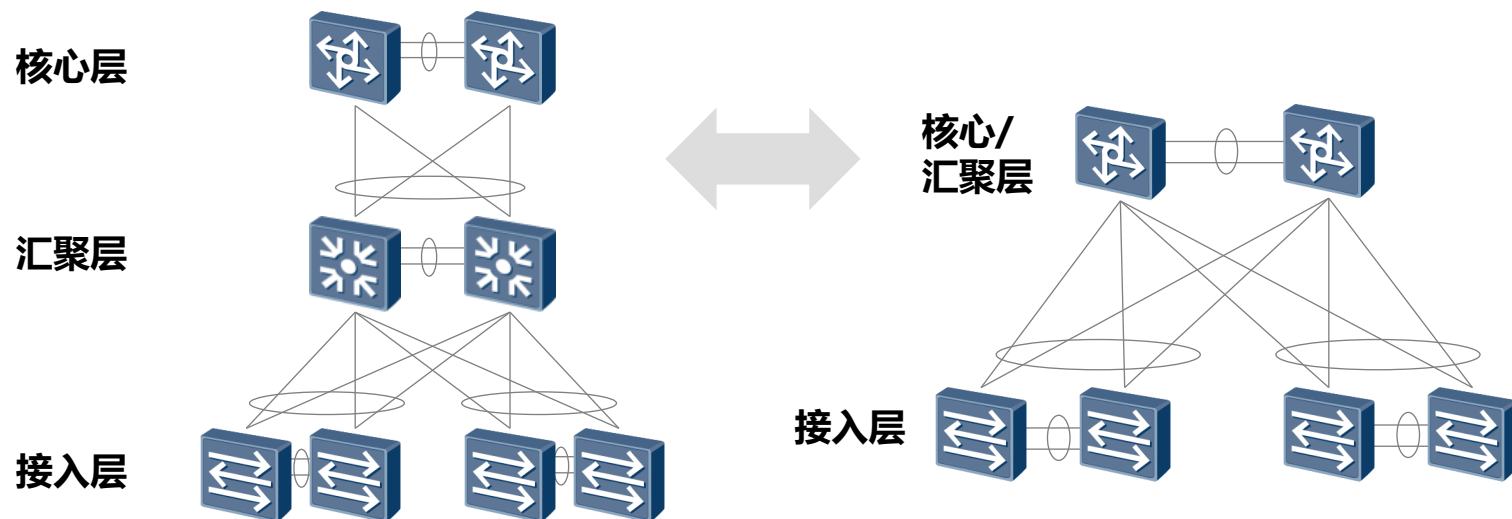
服务器接入设计

网关及路由设计

流量模型分析

扁平化网络设计

扁平化网络— 核心、汇聚合并



应用场景

中小规模数据中心，或者业务类型简单场景可以合并核心层和汇聚层，不同分区间用VLAN、L3VPN隔离

优点

设计、安装、维护简单
减少网络设备，降低成本

缺点

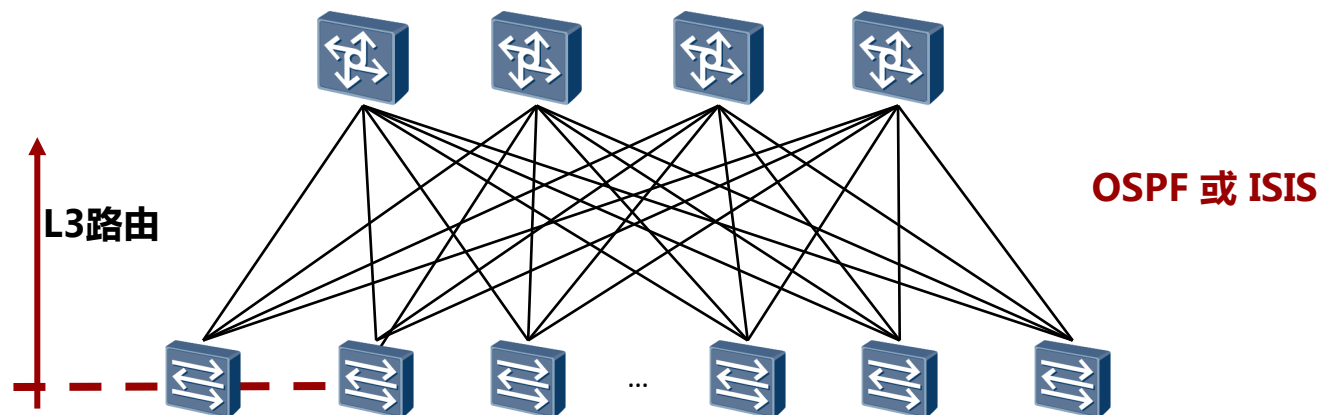
缺乏扩展性、布线管理较复杂

部署说明

在数据中心发展扩容时，可以根据需要将合并的核心/汇聚层再分解开，演变到核心、汇聚、接入三层结构。

为了简化网络，降低维护成本，也可以通过更换更高容量的交换机设备，将原来的核心、汇聚、接入三层结构演变为核心/汇聚合并的两层结构

胖树架构：L3互联



应用场景

在hadoop等SaaS、PaaS云计算环境,或者协同计算业务,流量收敛比小(1:1~1:2),要求有一个高效、均衡的,无阻塞网络。

解决方案

采用FAT Tree的三层组网方案可满足需要GE和10GE服务器接入方式

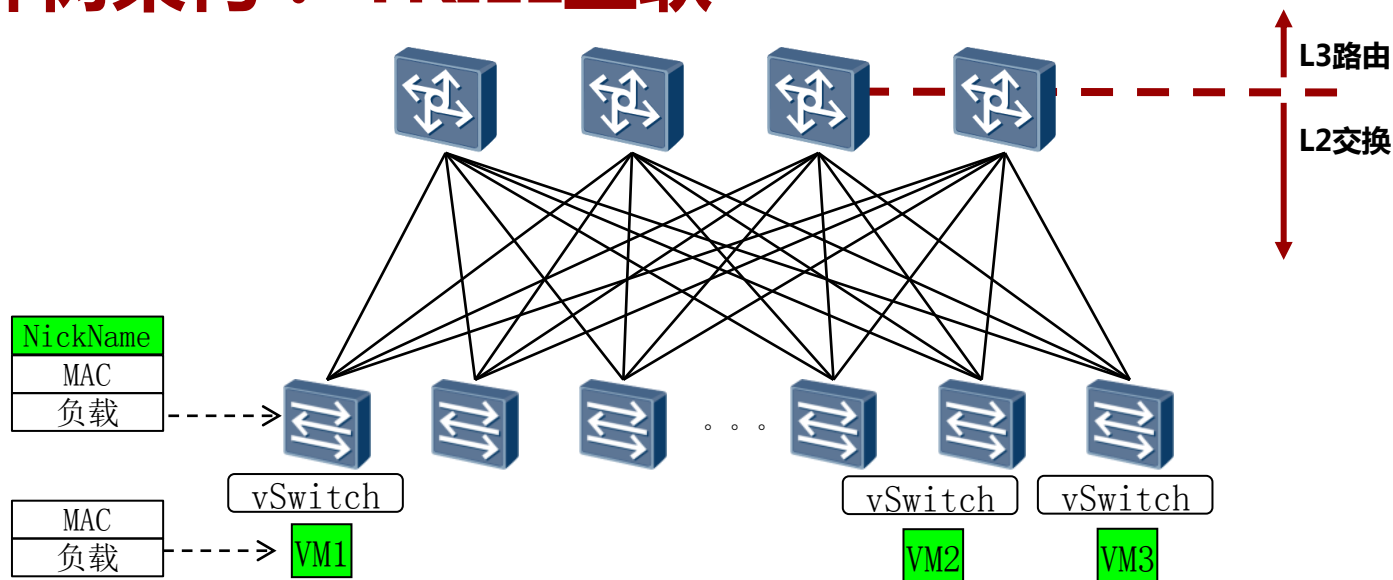
方案特点

任何两台服务器间的通信不超过3台设备。

实现IP路由的ECMP,采用五元组进行HASH,逐流负载分担,链路利用率高。

网络规模可弹性扩展,最多可以扩展到16平面,还可以增加网络层次扩大网络规模

胖树架构：TRILL互联



应用场景

大范围的资源共享，虚拟机大范围的迁移，需要构建更大范围的二层网络；
传统的xSTP靠阻断链路防环导致链路利用不充分，同时故障切换速度慢

解决方案

TRILL部署到网络TOR边缘，构建真正整网大二层；
TRILL和网关的叠加；
支持华为增强OAM功能，方便维护

方案特点

构建整网大二层网络，链路负载分担，提高利用率；
整网的资源共享和虚拟机迁移；
通过核心TRILL和网关的叠加，节省额外网关设备，降低成本

T级平台打造全线速40GE无阻塞网络



提供4608个40GE或18432个10GE端口接入

胖树架构

- 采用L2/L3多路径技术，构建胖树架构
- 打造无阻塞、低时延数据中心网络

40GE线速互联

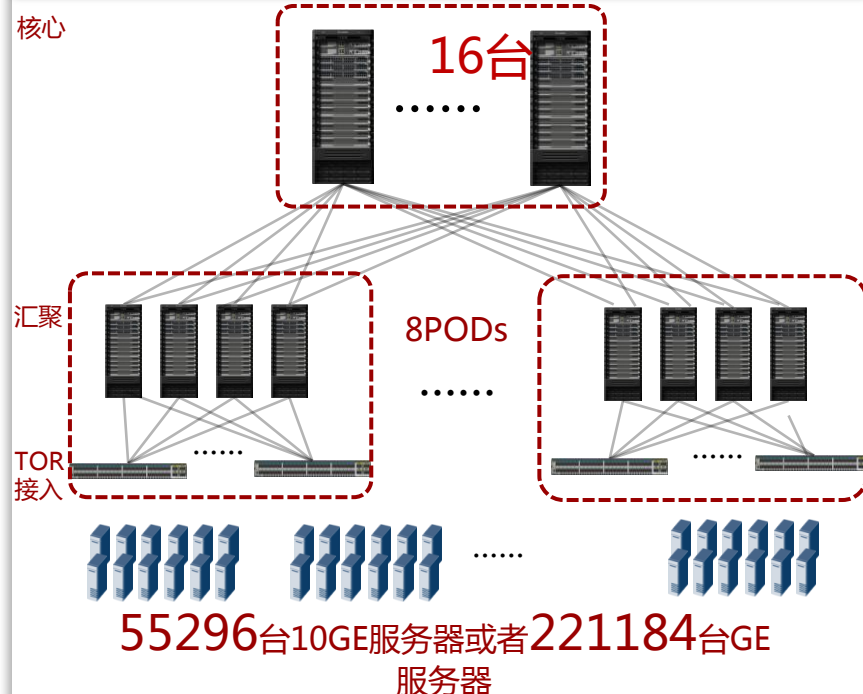
- CE12800单板支持24×40GE
- CE6800支持4×40GE上行

动态分布式缓存

- 动态分布式缓存，可有效吸收突发流量

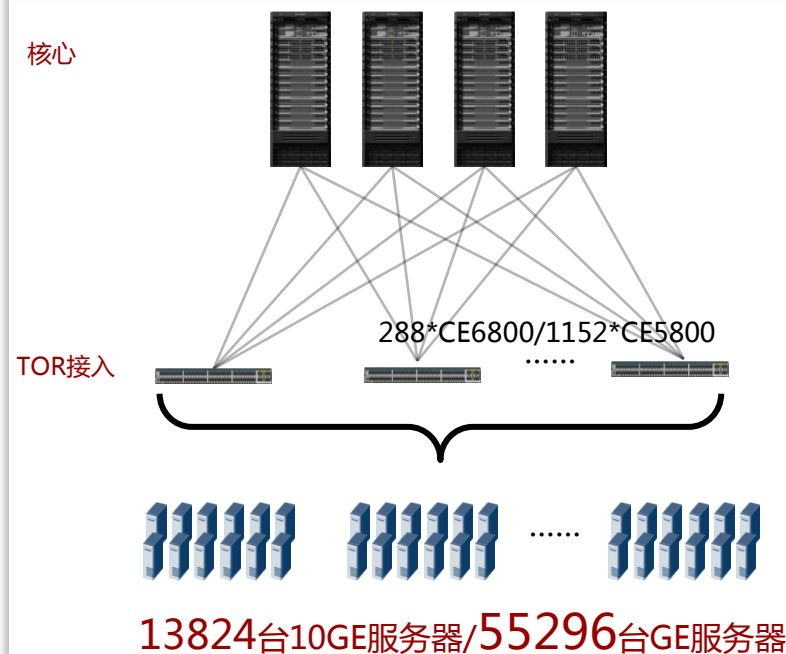
胖树架构提供海量服务器接入

三层架构



- 核心和汇聚之间无阻塞，TOR接入1:3收敛

二层架构



- 配合TRILL等多路径协议使无阻塞胖树架构延伸至TOR



Huawei Enterprise *A Better Way*