虚拟化环境下 网络 服务器 平台的协作经验

腾讯 网络平台部

Tencent 腾讯 Tencent Technology Company Limited

前言-希望您这一个小时的投资，值！

- 业务兄弟：
  - 做云计算，需要考虑到基础设施这条产业链，可能会遇到诸多挑战；
  - 做云计算，基础设施团队可以很给力；

- 服务器、网络、系统、平台开发、运营等兄弟：
  - 在互联网海量用户服务、云计算、虚拟化路途中的技术思考和探讨；
  - 将腾讯的摸索经验、问题和思考，真切地分享给大家，共同学习探讨；

- 今天的小厨：
  - 腾讯网络平台部　马志强/MartyMa

研讨议题

- 浮云来得如此真实，基础设施何来从容应对
- 互联网研发能力强，所有问题软件能否独挑
- 各方互补合体给力，方显云计算核心竞争力

迅猛发展的业务，海量服务用户

- 腾讯五大领先的互联网平台

网络媒体

- 流量最高的中国门户网站

无线业务

- 国内领先的无线门户网站

即时通信

- 国内最大的在线社区
- 活跃账号数6.476亿
- 最高同时在线数1.275亿

网络社区

- 国内最大的互动社区网站
- 活跃账户数4.92亿

网络游戏

- 国内第一互动娱乐游戏平台
- QQ游戏平台最高同时在线账户数680万

注：所有数字的统计口径为2010年第4季

构造面向海量用户服务的基础设施

- 腾讯基础设施规模
  - IDC资源范围：全国数十个大中城市部署业务和IDC资源；
  - 截止至2010年底，服务器量突破十万台；
  - 全国数万台网络设备；
  - 城域骨干网络容量达数百G；
  - 全国骨干网络达2.5G/10G级别；

- 预计2011年度建设的服务器量将达以前所拥有的服务器总量
  - 高峰时期同时开工建设的IDC数量达5~6个；
  - 高峰月度服务器建设量达10000台；

网络架构 | **IDC资源** | **服务器和操作系统** | **平台服务**

全面开放，即将迎来业务的又一波放量，不同性质的爆发

- 腾讯开放云平台 - 已经形成八大开放平台，包括腾讯朋友、QQ空间、腾讯微博、财付通、电子商务、搜索、彩贝及QQ；

- 注册开放商/个人超过7000家，每天接近150万条内容通过腾讯开放的分享组件被分享到QQ空间；
- 腾讯开放QQ登录，10天内就有上千家网站申请，现在已有超过9000家网站使用QQ互联账号系统登录；
- 接入腾讯各大开放平台的第三方应用已超过1000款，合作伙伴最高单月分成收入已经超过1000万元，数款应用进入千万DAU俱乐部、超过10款超过百万DAU；

浮云来得如此真实，基础设施何来从容应对

- 全方位的开放，要求强大云计算平台支持

- 腾讯开放云平台的优势：

数亿的活跃用户
利用强大的传播平台
您的产品能最快速度送达用户

注重双赢
腾讯注重双赢
收入按比例分成给第三方开发商

强大云服务支持
腾讯提供强大云服务支持
节省您的运营成本

浮云来得如此真实，基础设施何来从容应对

- 提供“强大的云计算开放平台”、实现“强大云服务支持”需要基础设施全角度考虑问题

从基础设施的角度看待云计算开放平台

●为合作伙伴提供应用级别互联互通

●为合作伙伴/应用开发商提供IT基础设施及相关的云计算服务

浮云来得如此真实，基础设施何来从容应对

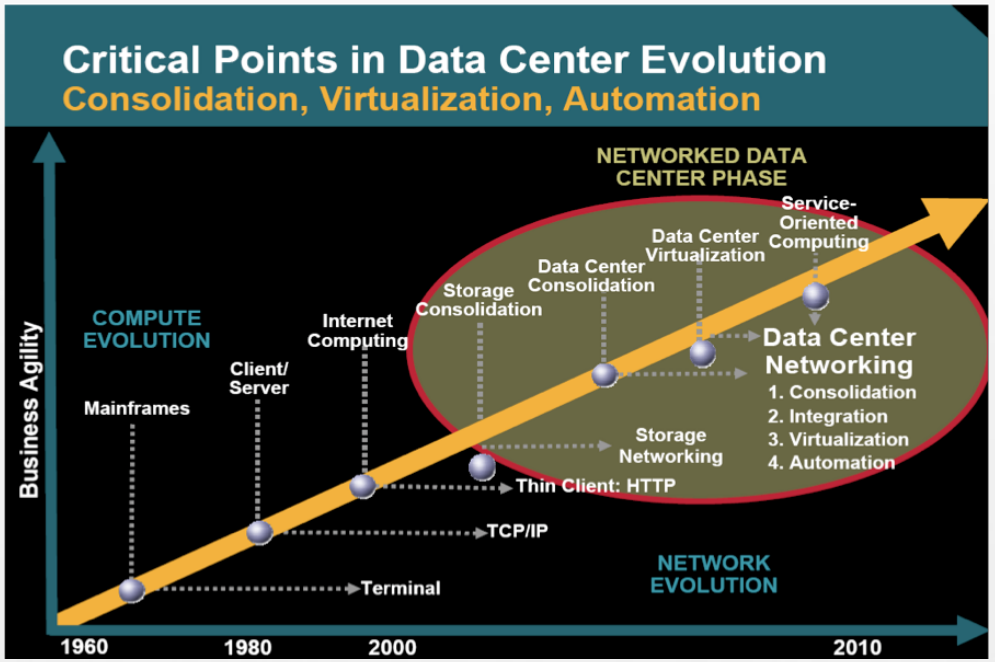
云计算平台支持需求	IT及网络架构
合作伙伴快速接入需求 ▪ 以小时或分钟为颗粒度进行基础设施的供给或定制化	**资源的快速交付** • 标准化、可快速部署的服务器和网络资源
灵活多样的平台服务 ▪ 轻松实现的资源申请和回退 ▪ 多样的基础网络服务	**安全隔离** • 为合作伙伴提供可控的安全隔离
合作伙伴个性化需求 ▪ 服务质量监控和保障 ▪ 灵活多样的数据分析	**灵活扩展性** • 为海量合作伙伴接入做好准备，基础设施扩展灵活
	集中管理平台 • 通过统一的平台对服务器、网络、基础服务等进行集中调度 • 为合作伙伴提供简洁的申请、调整、回退资源的管理平台




腾讯朋友
pengyou.com

浮云来得如此真实，基础设施何来从容应对

- 基础设施的发展是否已经就绪云计算---标准化
  - 及时的IDC资源供给
  - 标准化服务器型号和操作系统
  - 标准化网络架构方案和部署流程
  - 标准化IDC验收交付流程和规范
  - 3~5个月->1个月->1小时？



- 基础设施的发展是否已经就绪云计算---虚拟化
  - 通过标准化的虚拟机和网络资源，加快资源供给速度，提升基础设施整体的标准化程度

- 基础设施的发展是否已经就绪云计算---自动化
  - 内部基础设施的统一管理和调度
  - 对合作伙伴服务的规范和自动化

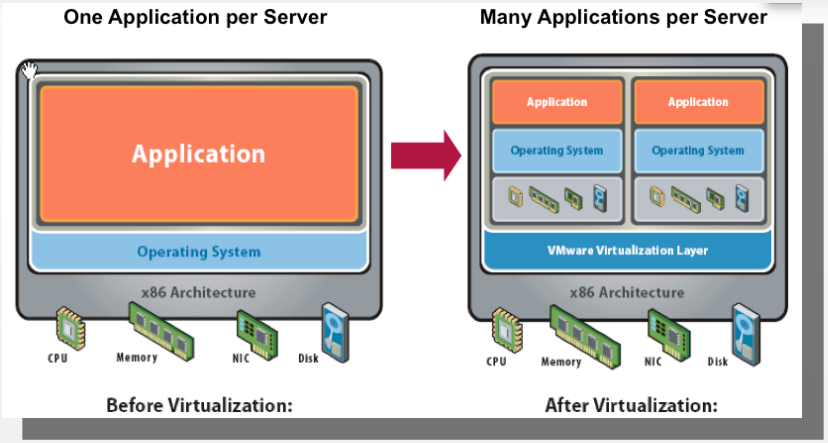
研讨议题

- 浮云来得如此真实，基础设施何来从容应对
- 互联网研发能力强，所有问题软件能否独挑
- 各方互补合体给力，方显云计算核心竞争力

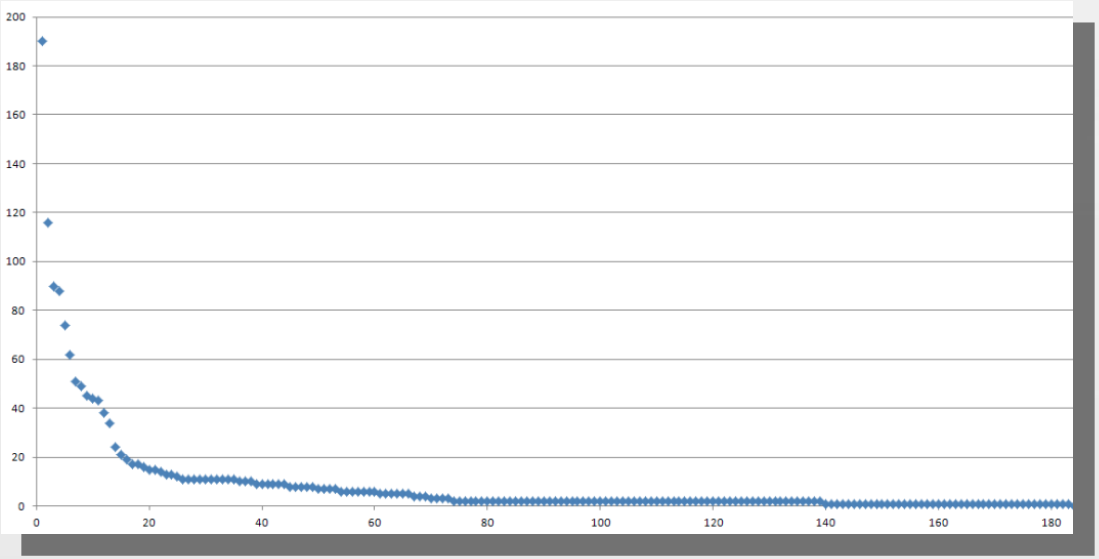
互联网研发能力强，所有问题软件能否独挑

第一回合-虚拟化来了，初始需求是什么

- 物理服务器虚拟化，我们服务的对象是虚拟机
  - Linux系统，基于XEN/KVM等开源虚拟化平台
  - 1pXv，可对外服务的服务器数量成倍增长
  - 物理服务器网络接口吞吐利用率大幅提升
  - 虚拟化效果的探索和验证阶段



- 合作伙伴/应用供应商/Tenant
  - 除“网络模块容量规格”、“VLAN”、” TOR”之外，另外一个规格层次，一个“合作伙伴”的设备可能跨TOR、跨VLAN；
  - 萌芽态的合作伙伴尺寸较小，有木有可能呈爆发性增长态势；

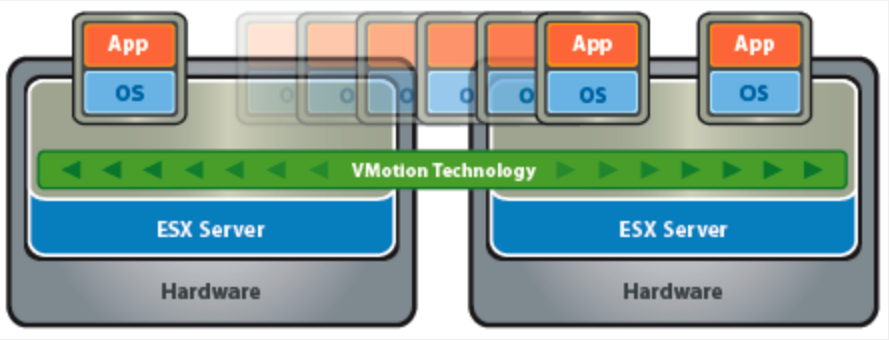


第一回合-挑战与尝试

- 如何解决服务器IP地址需求
  - ❖ 击穿原有接入层网段设计容量，并势必一定程度上造成IP地址浪费
  - – 扩大现有IP地址段
    - 业务能否接受服务器IP地址变更
  - – 使用Secondary IP地址
    - 运维复杂度稍有增加
  - – 直接增加更多VLAN
    - 运维复杂度稍有增加
    - 可能出现虚拟机跨VLAN互访需求
- 能否构建稳定的网络而同时适应合作伙伴的成长
  - ❖ 现阶段一个合作伙伴/应用供应商/Tenant的OS数量小，后续爆发性增长如何平滑扩展
  - – 构建容纳更多OS的大二层/VLAN网络

第二回合-要迁移要隔离

- 要提供虚拟机的无缝迁移
  - 保持虚拟机IP地址不变；
  - 无中断迁移，网络信息（VLAN、安全策略等）联动；
  - 合作伙伴/应用供应商/Tenant范围内的无缝迁移；



- 要提供虚拟机间的安全隔离
  - 同一物理服务器可能承载多个合作伙伴/应用供应商/Tenant的虚拟机；
  - 安全策略可扩展可变更；

第二回合-挑战与尝试（1）

- 无缝迁移，一个TOR内部好搞定
  - 无论路由还是交换的CLOS架构-交换域仍可覆盖至少一台接入层交换机；

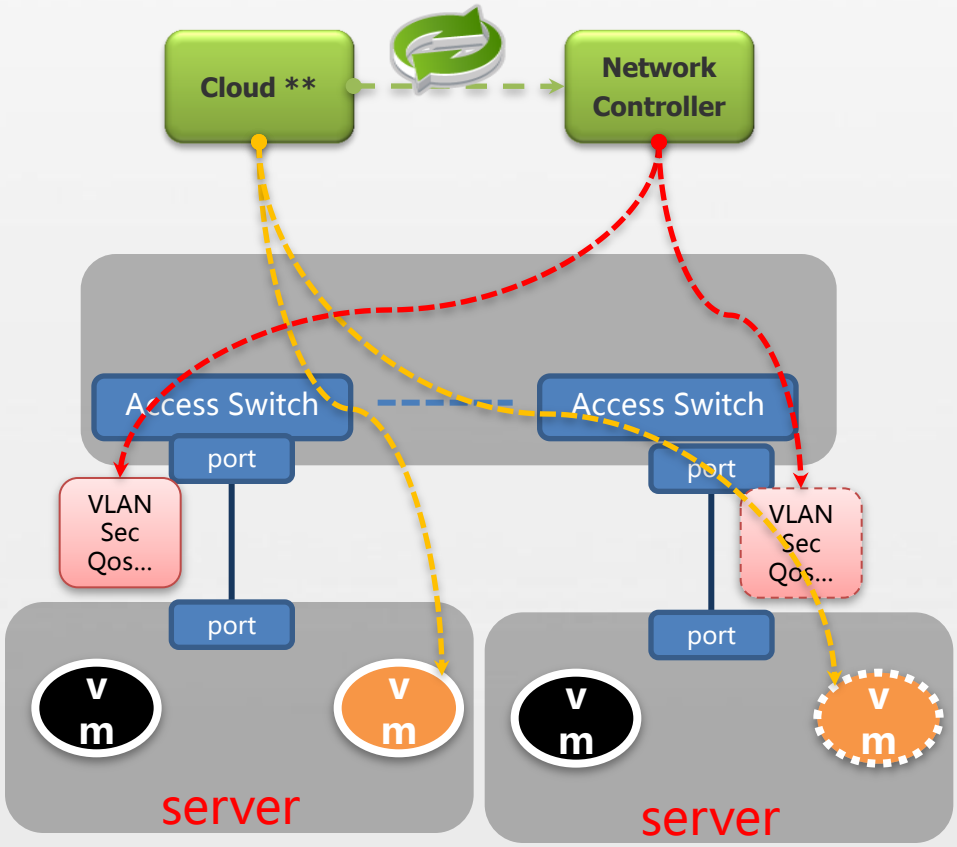
- 无缝迁移，一个数据中心网络模块内，路由的CLOS架构。。。
  - 追求稳定的路由CLOS架构，当虚拟化迁移真的来临时，有些捉襟见肘

- 无缝迁移，一个数据中心网络模块内，交换的CLOS架构。。。
  - ❖ 生成树叶节点增加、MAC地址容量需求增加、ARP容量需求增加、STP域被动扩大
  - ❖ VLAN敢跨多大范围、敢承载多少台虚拟机？二层无环网络是否真正可扩展的二层网络？
  - 确保接入、核心交换机的容量指标满足需求和增长
  - 妥当设计VLAN大小，满足现状同时适当允许增长
  - 采用二层无环网络设计避免依赖STP - 思科VPC/H3C IRF
  - 利用TRILL/FabricPath技术构建真正稳定的二层网络

第二回合-挑战与尝试（2）

- 无缝迁移，跨数据中心间。。。
  - 数据中心间直联链路；
  - Cisco OTV；
  - EoMPLS/VPLS等二层VPN技术；

- 真正无缝，要服务器、网络、平台合体给力才能实现
  - 虚拟机的motion动作由虚拟机集中管理平台实现；
  - 与网络集中管理平台联动，同步调度VLAN信息、安全策略、Qos策略等信息；



第二回合-挑战与尝试（3）

- 安全隔离，首先网络能否看见虚拟机？
  - 同一个物理端口上，多台虚拟机出现，如何分辨VLAN归属、安全策略等

- VEB
  - 成熟的产品Cisco Nexus 1000v
  - 开源社区的OpenVSwitch
  - ❖ 对Hypervisor的特殊要求
  - ❖ 软件实现对性能的影响

- 802.1Qbg和802.1Qbh
  - 协议成熟性
  - 产品成熟性
    - 思科7+5+2
  - ❖ 要求网卡和Hypervisor特殊支持
  - ❖ 普遍在万兆网卡上的应用，与现阶段千兆网卡服务器广泛部署，是否同步

第二回合-挑战与尝试（4）

- 安全隔离，软件层面能否多做一点点？
  - 虚拟机上的IPTABLES
  - ❖ 安全策略复杂化带来的性能挑战
  - ❖ 安全策略分散管理带来的运营压力

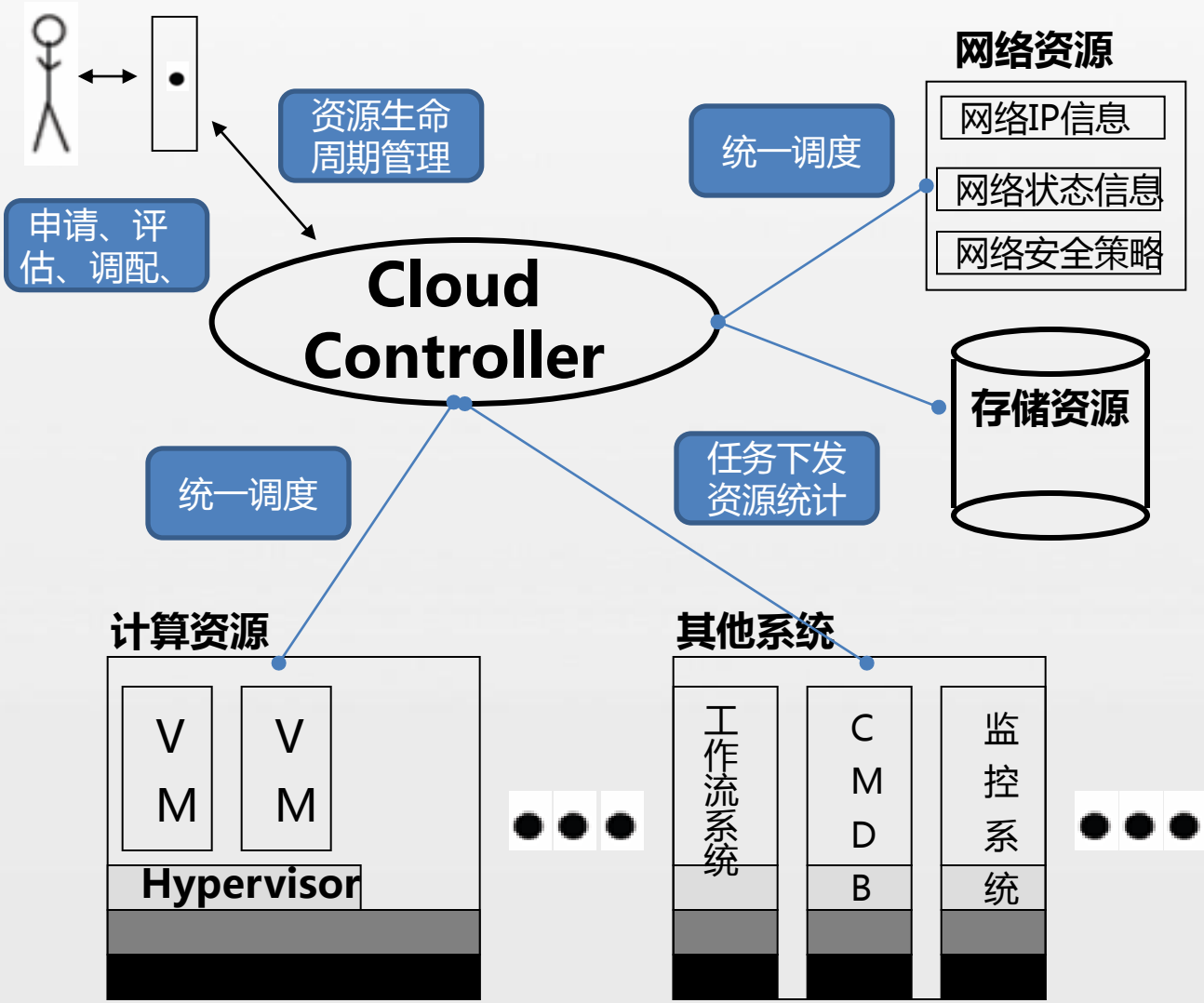
- 回归网络传统手段？
  - VLAN+ACL
  - VRF
  - 端口ACL
  - ❖ 仍面临安全策略较为分散，需网络调度系统进行集中统一管理
  - ❖ 要求平滑过渡到802.1Qbg和Qbh，避免工具系统研发重构

第三回合-奉献更多核心价值

- 合作伙伴网络带宽服务质量保障和流量统计
  - 局域网级别的QOS服务质量保障，定制化的网络流量统计，均要求更加特性更加充足的接入层交换机、更加细致颗粒度的带宽管理流程和规范、更加灵活强大的流量分析系统；
- 无阻塞通信或万兆服务器接入需求
  - 无阻塞通信需求将对CLOS架构网络的规格和网络设备密度提出新的挑战
  - 万兆服务器网卡及万兆网络端口的成本，与虚拟化的总体步伐能否匹配
- 业务部署的容灾和高可靠性
  - 跨数据中心的资源快速供给，仍然要求跨网络模块的无缝迁移
  - 城域网和广域网能否为合作伙伴提供安全独立的基础设施服务

第三回合-奉献更多核心价值

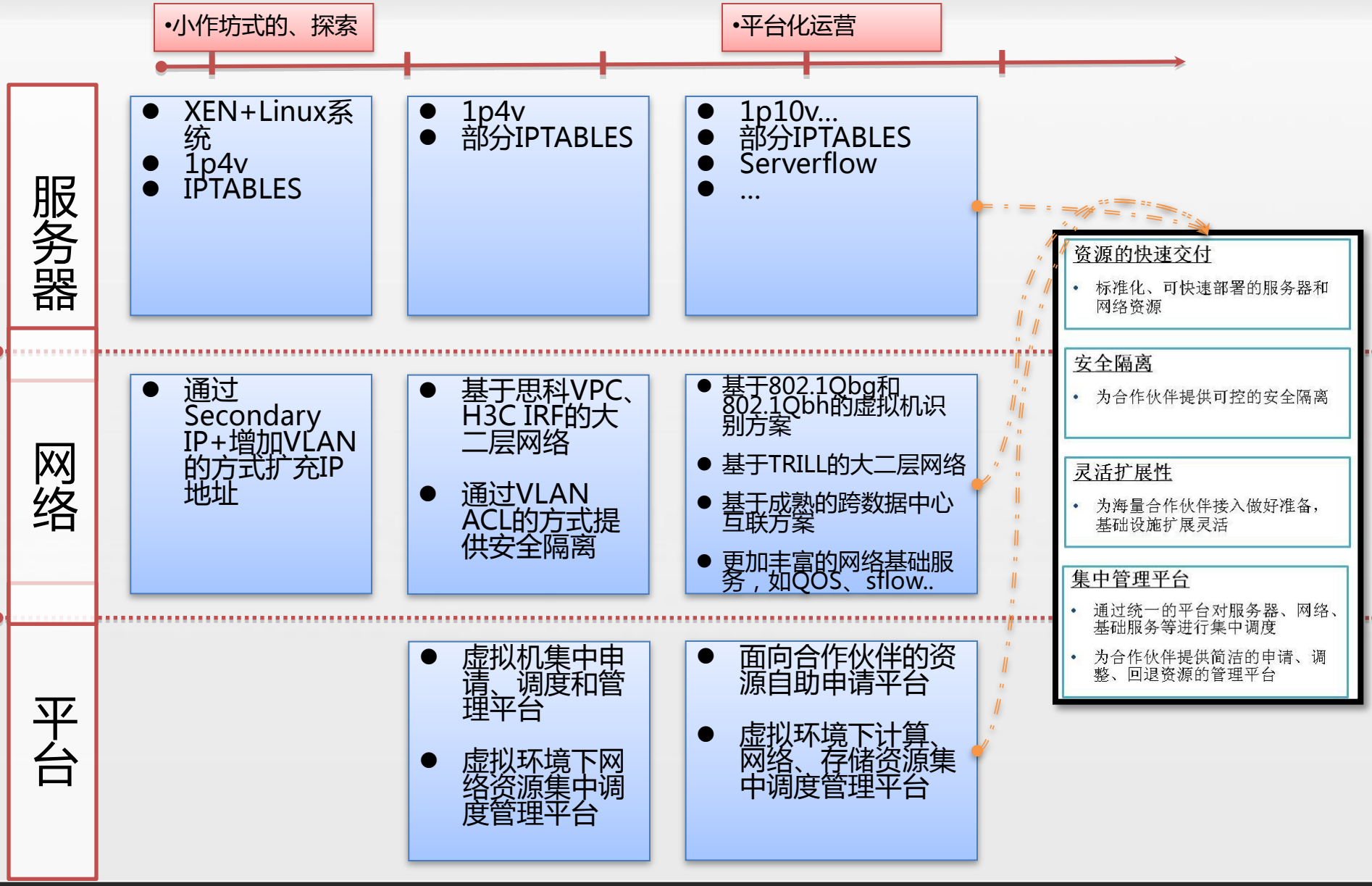
合作伙伴/业务BU:



研讨议题

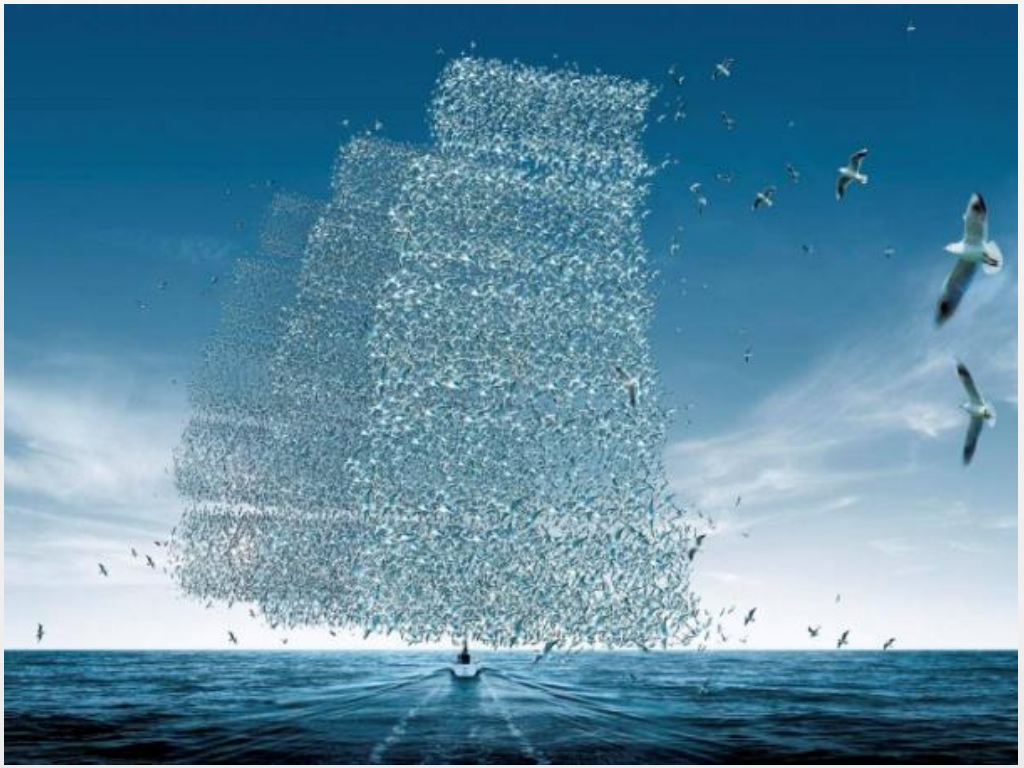
- 浮云来得如此真实，基础设施何来从容应对
- 互联网研发能力强，所有问题软件能否独挑
- 各方互补合体给力，方显云计算核心竞争力

各方互补合体给力，方显云计算核心竞争力



为了更好的云计算平台和云服务支持，我们在努力！

- 更加快捷的基础设施资源供给
  - 模块化数据中心
  - 自动化的网络部署
  - 虚拟化
- 更加强大的基础设施云计算平台
  - 网络和服务器联合
  - 计算资源、网络统一调度的集中管理平台
- 更加丰富的基础设施云服务支持
  - 突出核心竞争力的基础设施服务
  - 面向合作伙伴的自助管理平台



Key Takeaways

- **基础设施团队要与业务团队建立良好的沟通体系和信任关系**
  - 开放、云计算，业务团队要认识到基础设施的复杂性和重要性；
  - 基础设施团队要带着强烈的服务意识加强与业务的沟通；
  - “合作伙伴-业务团队-基础设施团队” - 沟通是解决问题、服务提升的最佳渠道；

- **强大的云计算开放平台和强大云服务支持需要基础设施团队转变思路**
  - 资源的快速供给需求、服务的多样性需求，要基础设施团队更加快捷地提供解决方案；
  - 基础设施团队要相互补偿，联合给力；

- **腾讯以开放的心态面对业界，推动产业链的健康成长**
  - 联合运营商、设备供应商，积极探索解决方案，共同面对云计算带来的挑战；
  - 积极与业界同仁学习探索、分享经验，并推动业内相关标准的建立的话，受益整个行业；

乐 于 倾 听　　乐 于 分 享

Tencent 腾讯 Tencent Technology Company Limited