
Architect
SACC
2022 中国系统架构师大会
SYSTEM ARCHITECT CONFERENCE CHINA 2022
激发架构性能 点亮业务活力
云上会议 网络直播 | 2022年10月27-29日
ITPUB

边缘云网络架构演进

阿里云 高级技术专家
曹超

01 边缘云背景

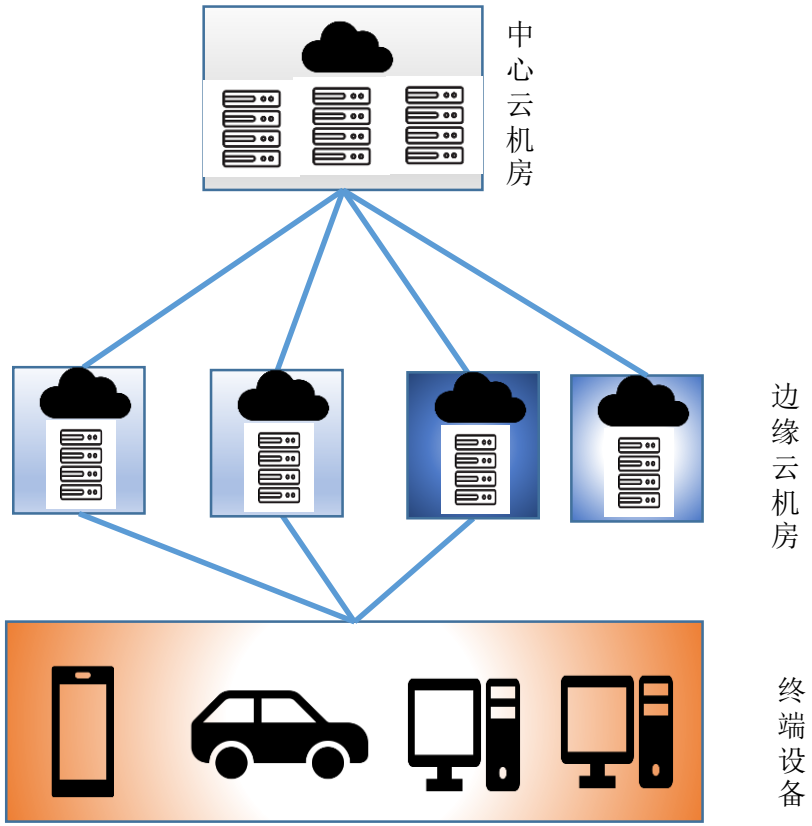
02 边缘云网络产品介绍

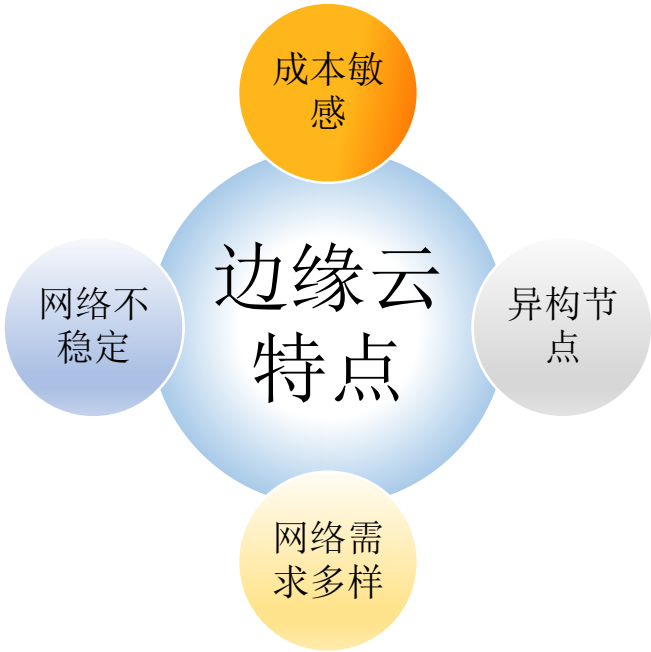
03 边缘云网络发展

04 未来方向

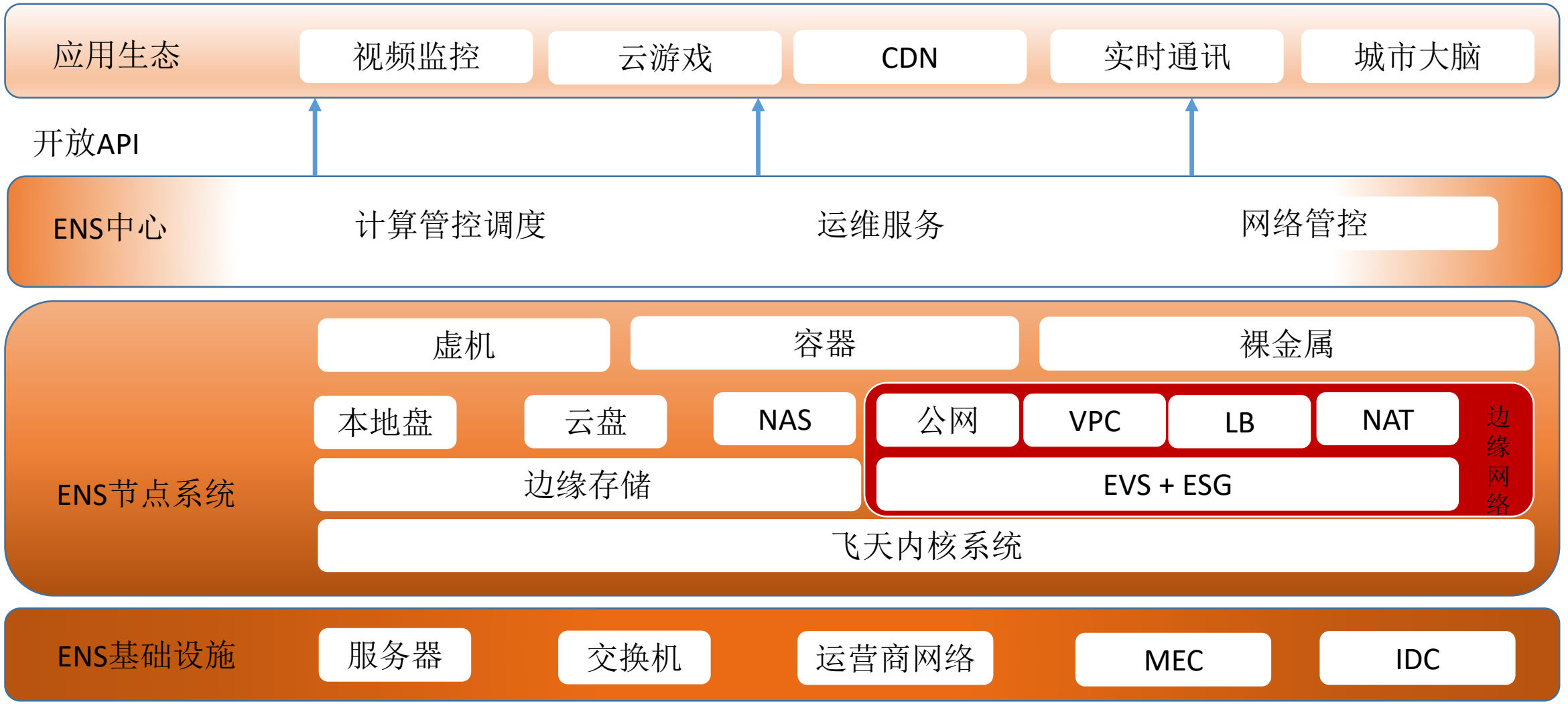
一 背景

- 边缘覆盖广，单节点规模小，对成本时延敏感
- 节点异构，自建机房+合作机房
- 网络不稳定，运营商时常割接
- 网络需求多样，对二三层网络都有要求





产品架构


应用生态
视频监控
云游戏
CDN
实时通讯
城市大脑
开放API
ENS中心
计算管控调度
运维服务
网络管控
ENS节点系统
虚机
容器
裸金属
本地盘
云盘
NAS
公网
VPC
LB
NAT
边缘网络
边缘存储
EVS + ESG
飞天内核系统
ENS基础设施
服务器
交换机
运营商网络
MEC
IDC

二 边缘云网络产品介绍

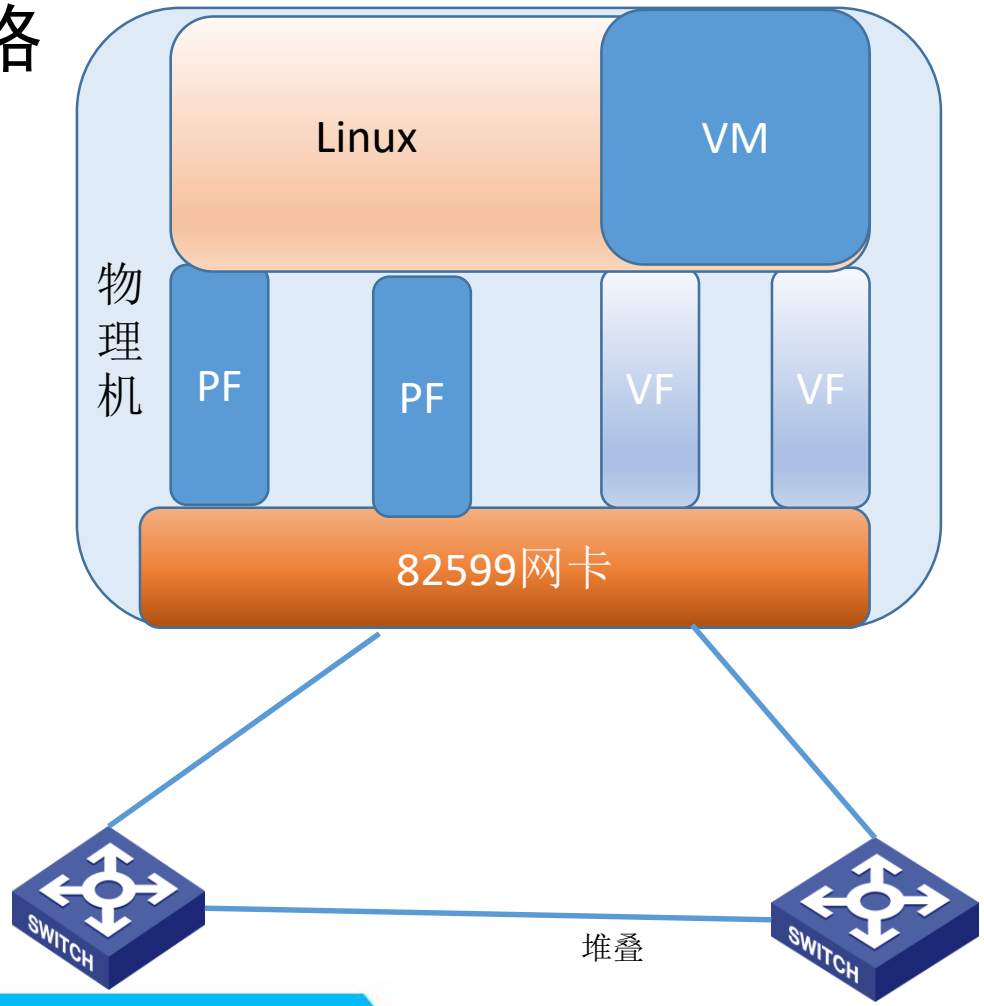
1. 公网下沉模式
    1. 实例携带公网IP，具有低延时高性能特性
    2. 按照节点可选二三层网络模式
    3. 具备安全组，限速基础功能
    4. 支持IPV6
2. 公网上移模式
    1. 实例只携带内网IP，默认不能访问公网，安全隔离
    2. 支持弹性IP，动态申请释放
    3. 支持四层负载均衡，负载均衡支持DNAT，FullNAT，DR等模式
    4. 支持七层负载均衡，HTTP/HTTPS,继承Tengine，稳定，特性丰富
    5. 支持DNAT/SNAT
    6. 支持跨节点VPC互通
    7. 支持公网四层加速

三 边缘云网络发展

1.0 直通经典网络

客户需求

- 网络能通
- 能支持LVS DR
- 延时低
- 能计费
- 三线运营商
- 快速上量



实际实现

1. 支持限速，但不均衡
2. 运行在二层网络，支持DR模式
3. 没有软件介入，延时低
4. 不用开发，上线快

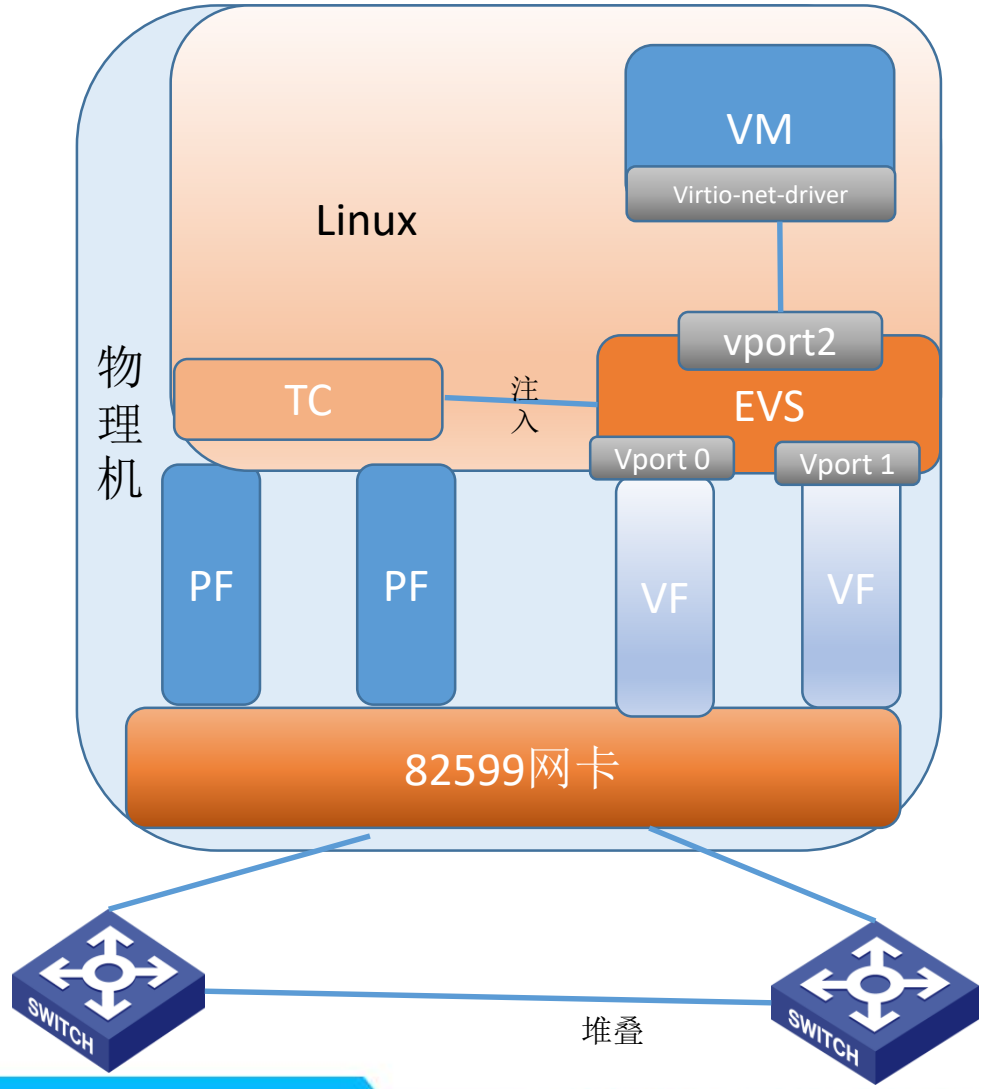
缺点

1. 不支持ACL安全组
2. 不支持热迁移
3. 虚机内部需要适配对应驱动，并且无法感知网络变化
4. 虚机内部网卡队列数受网卡限制
5. 给虚机网卡个数受限硬件网卡

2.0 EVS经典网络

客户需求

1. 安全组
2. 限速
3. 热迁移
4. 网卡多队列



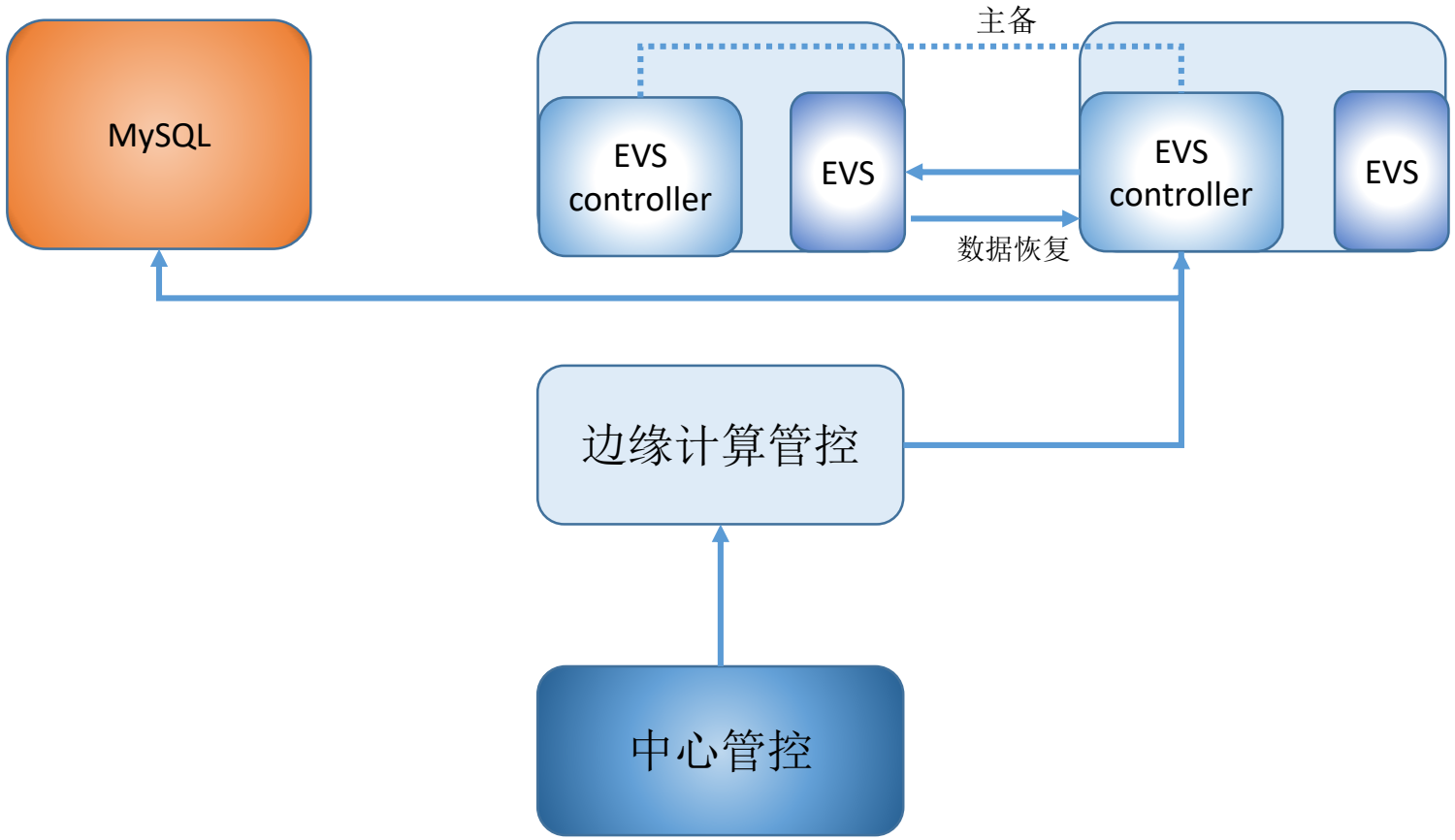
实际实现

1. 引入自研EVS虚拟交换机
2. 支持左侧所有的客户需求
3. 屏蔽硬件网卡与链路抖动
4. 支持热升级，快速迭代
5. 支持热迁移，主动运维

缺点

1. 引入软件层，增加网络延时，性能比直通有所下降
2. 队列数增多，vm cpu均衡度提升
3. EVS占用4 core，减少资源售卖

2.0 EVS经典网络-控制面



1. 边缘计算管控负责存储，网络，计算统一管理，EVS controller负责网络管理

2。 创建实例之前通过网络管控创建对应的EVS Vport，每个虚拟网卡对应一个Vport

3. 虚拟网卡的ACL，限速，Vlan等配置均对应EVS的Vport

4. EVS不持久化存储vport配置等，统一由EVS controller管理

2.0 EVS经典网络--一些坑

1. EVS采用DPDK开发，cpu绑定需要采用isolcpus与qemu隔离，并且注意超线程影响
2. 提前大页预留，避免内存碎片
3. Virtio的虚拟队列合理调度到EVS cpu
4. 82599的TSO有长度限制，分片注意丢包问题
5. 网卡的queue size在brust报文比较大的时候，需要进行调优
6. 虚机内存软中断需要打散，避免出现因为虚机接收不过来导致丢包

3.0 EVS混合网络

客户需求

1. 多网关引流 （公网加速，边缘节点互通）
2. 纯内网实例 （减少公网IP浪费）
3. 去堆叠三层网络支持 （overlay网络）
4. VPC网络 （VLAN隔离资源受限）
5. 四七层LB， HAVIP
6. SNAT/DNAT

3.0 EVS混合网络

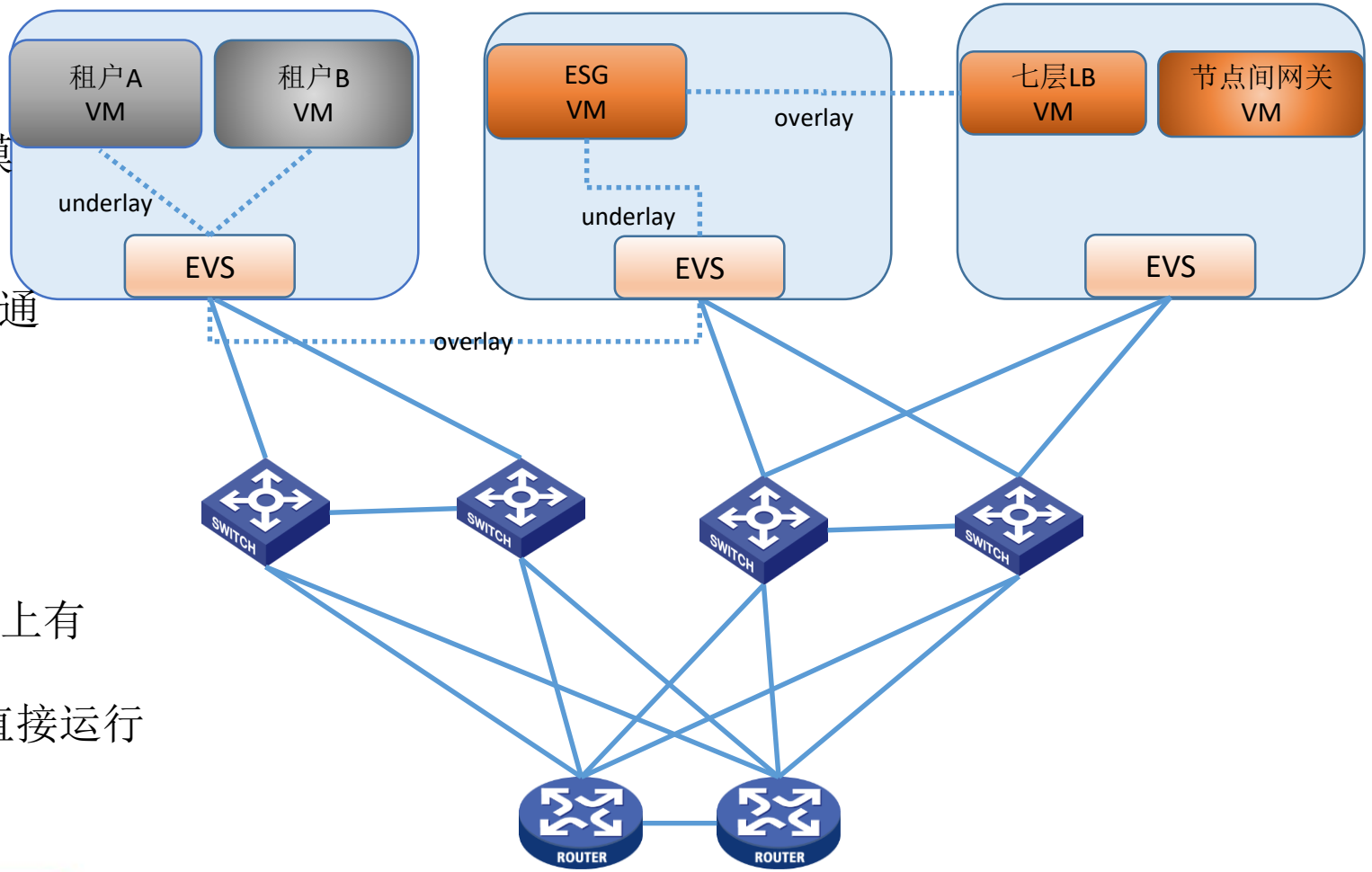
实际实现

1. 物理交换机采用去堆叠模式支持百台规模
2. EVS 支持overlay与underlay混合运行
3. 节点网关支持四层负载均衡，SNAT/DNAT
4. 节点间网关支持公网加速，跨节点VPC互通
5. 网关虚拟化，横向扩展，弹性伸缩

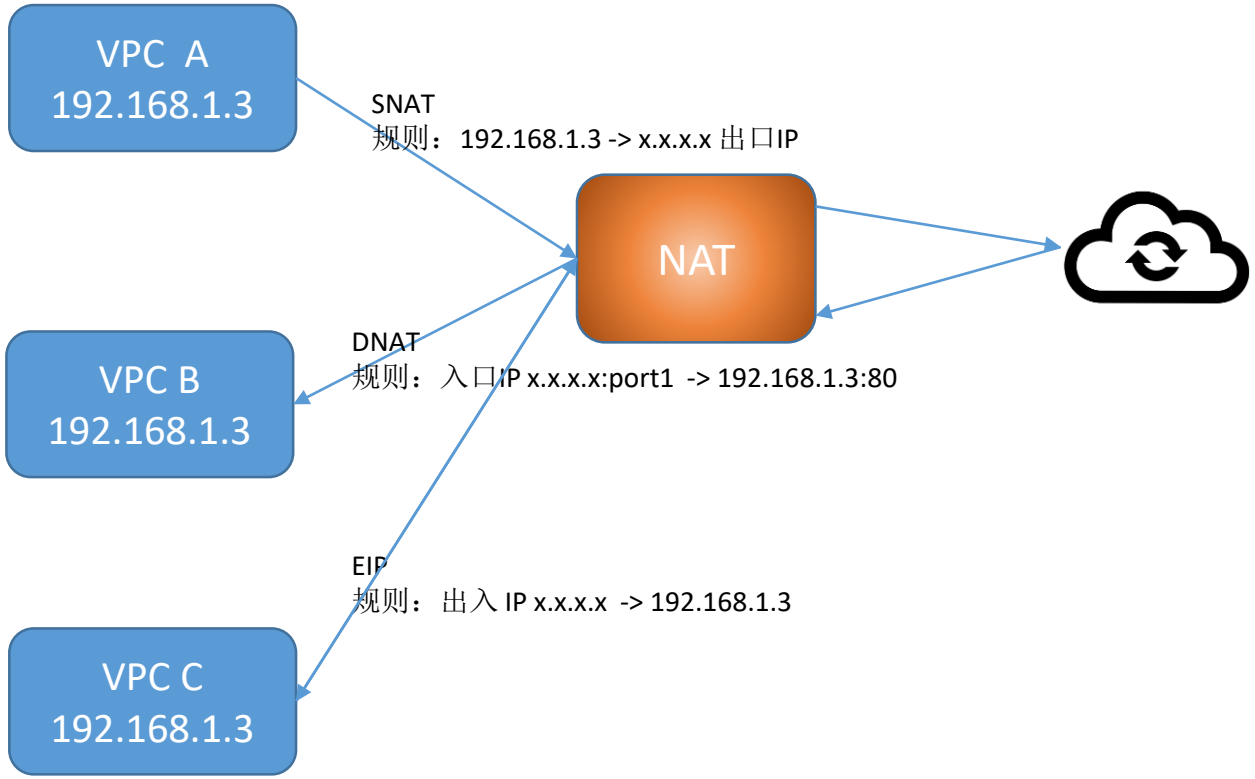
缺点

1. 网关虚拟化之后性能一般
2. Overlay与underlay混跑之后，对用户使用上有一定理解成本，逐渐收敛到overlay

3. 对传统的客户比如CDN场景，LVS DR无法直接运行



NAT



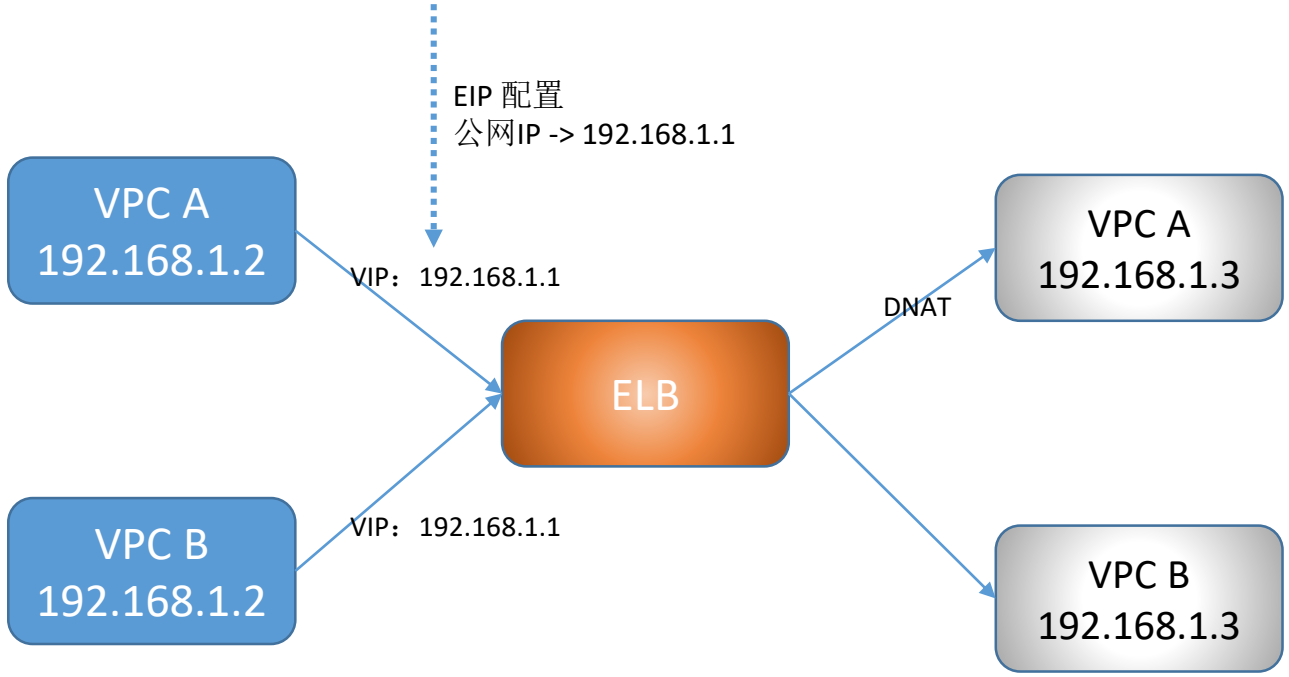
1. SNAT出口IP支持多个公网IP作为出口，采用轮询方式选择

2. DNAT可以配置多条规则分别映射到多个内网+Port

3. DNAT支持多个IP+port 映射到同一个内网+port

4. EIP只支持一对一映射

四层负载均衡-VPC内

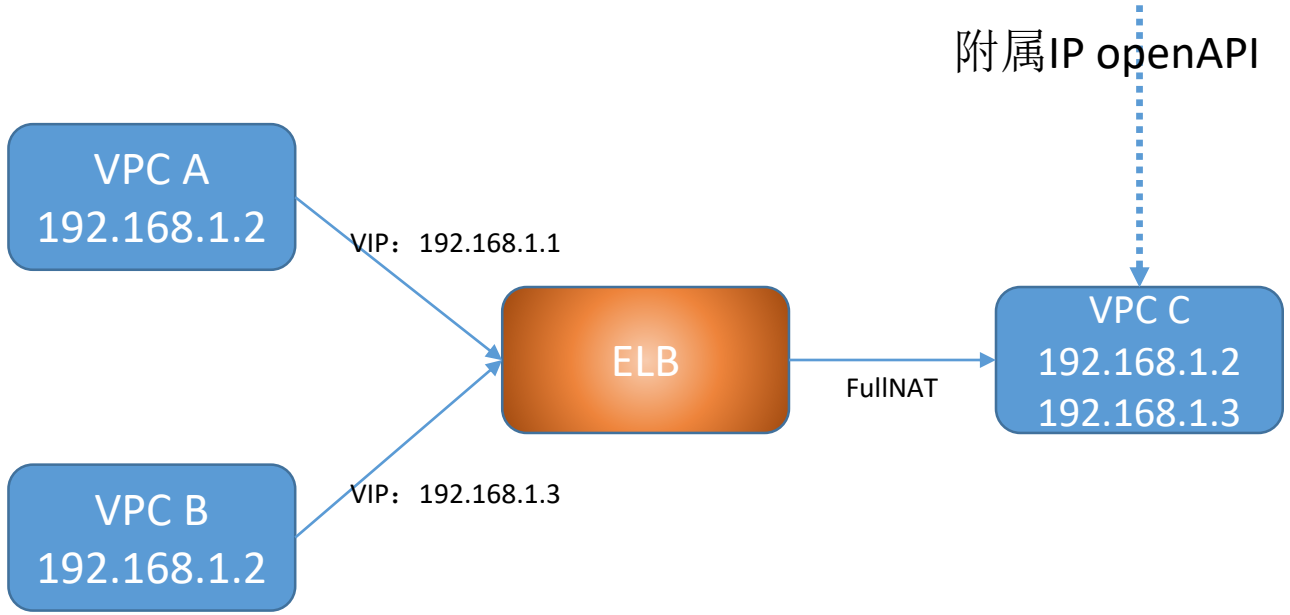


1. 通过给内网VIP配置EIP实现公网LB

2. 每个VPC内创建自己的LB实例，与其他VPC完全隔离

3. LB RS支持四七层健康检查，有hash，轮询权重策略进行rs选择

四层负载均衡-跨VPC



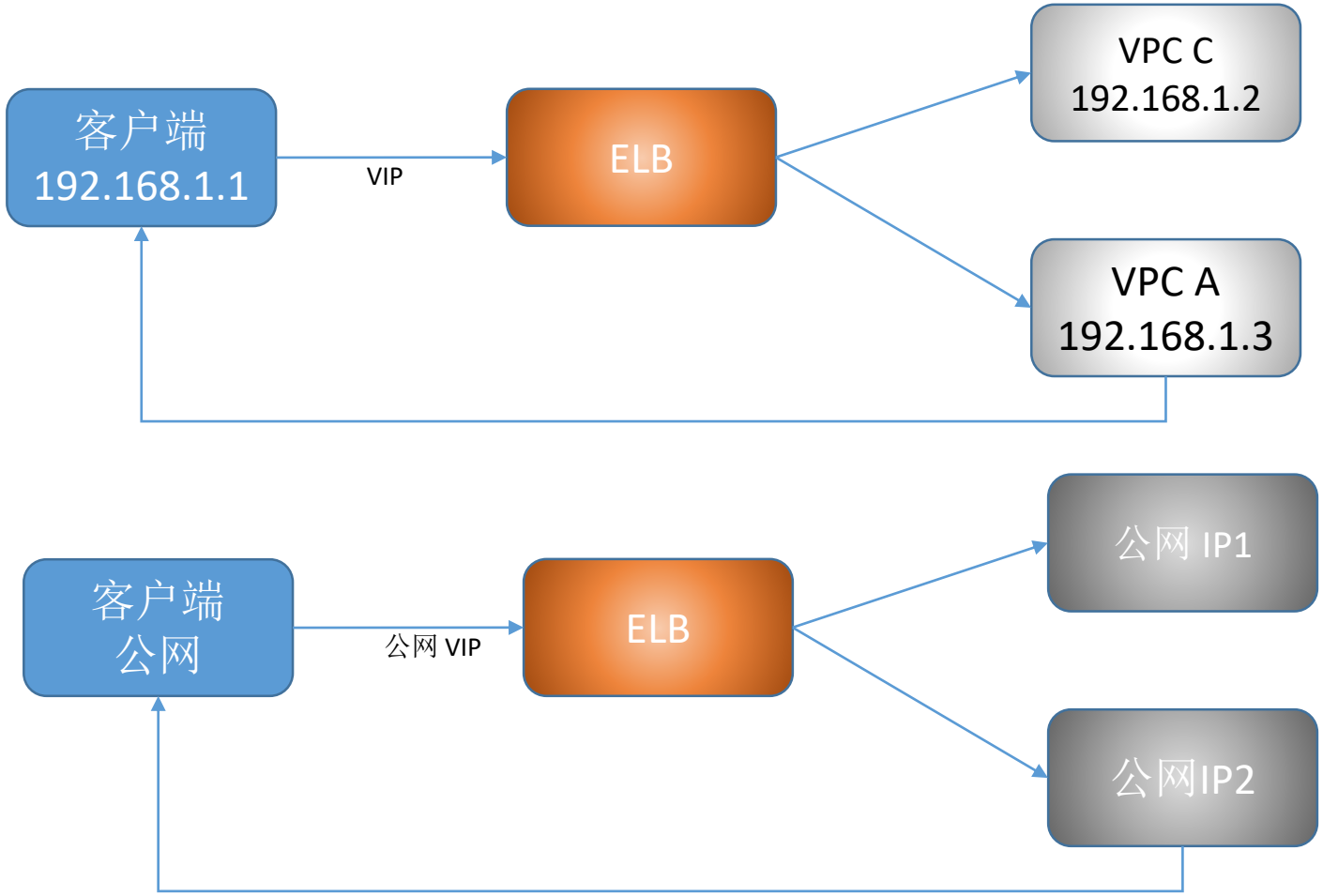
1. FullNAT的源IP为唯一IP段，不与VPC IP冲突

2. VPC C通过TOA获取客户端地址

3. VPC A与B的VIP可以不一致，都会占用VPC内一个IP资源

4. VPC C内可以通过不同的port或者不同IP区分

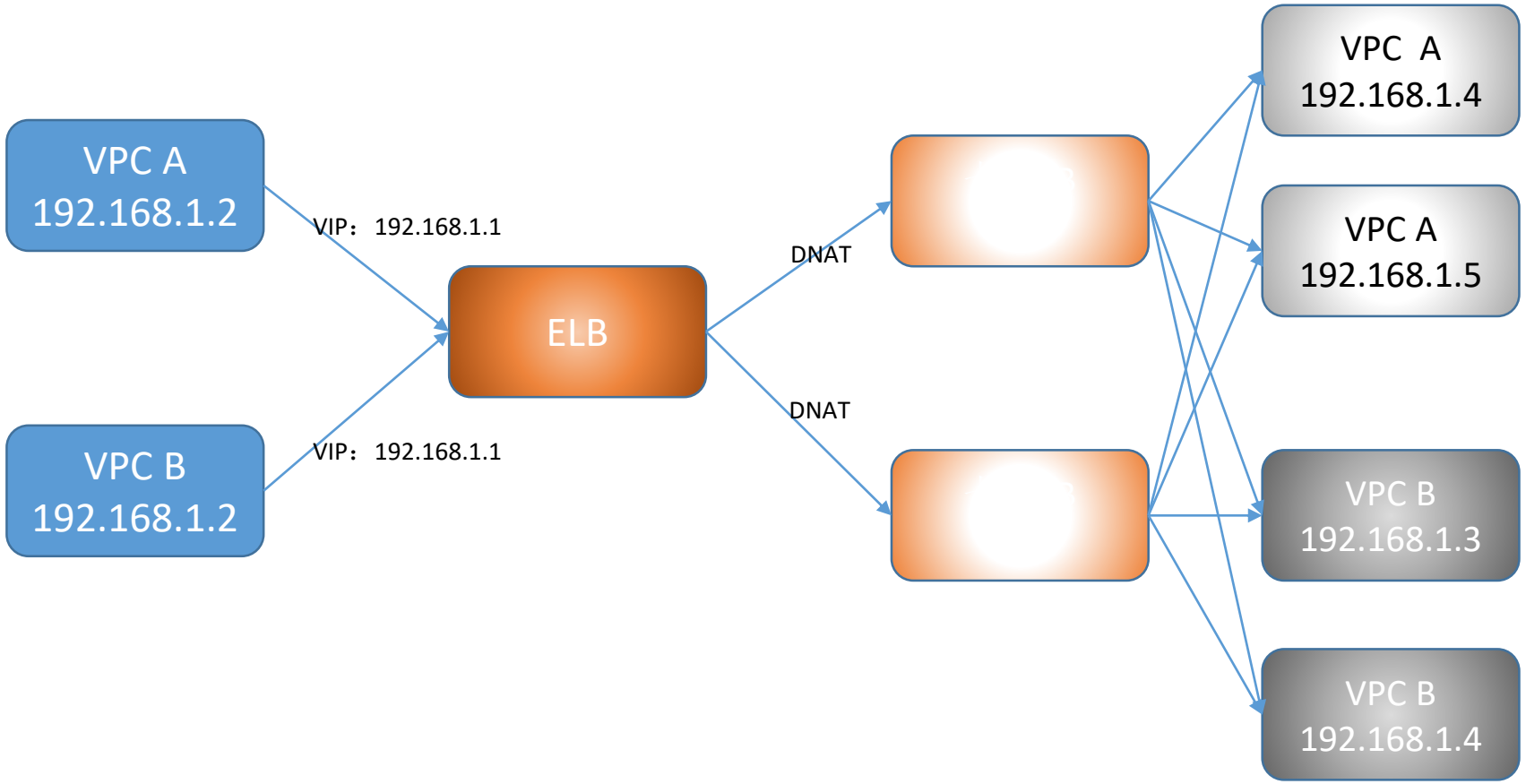
四层负载均衡-三角模式



1. 三角模式回程流量绕行ELB，直接发往客户端，提升性能

2. 客户端与后端RS可以是VPC内网IP也可以是公网IP

七层负载均衡



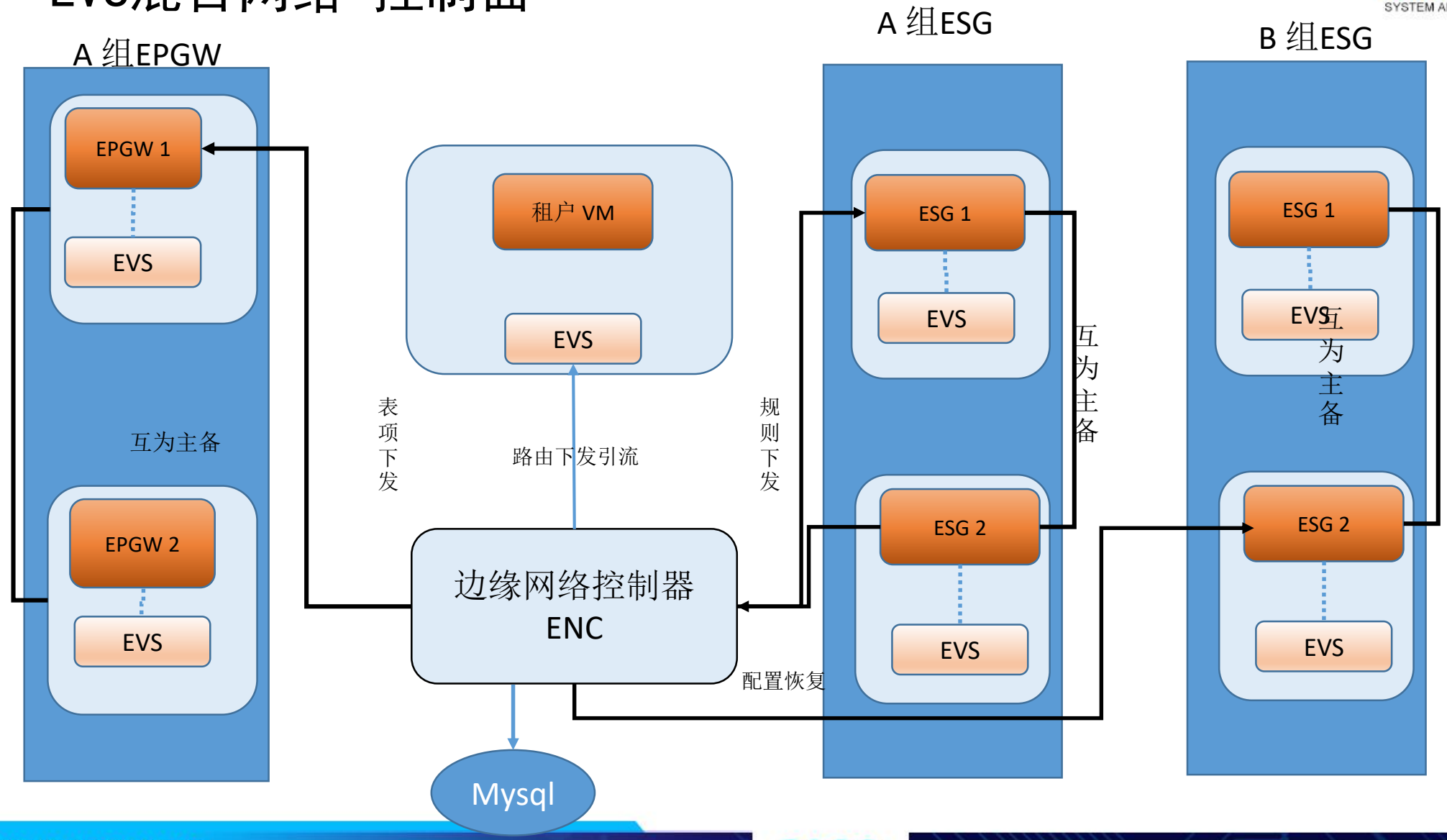
1. 七层LB挂在四层LB之后，复用四层能力

2. 七层LB也是虚机部署，网络与租户隔离

3. 支持HTTP/HTTPS

4. RS 通过HTTP-X-Forward获取客户端IP

3.0 EVS混合网络-控制面



3.0 EVS混合网络-控制面

- 所有网元管控统一由边缘网络控制器（ENC）调度处理
- ENC的配置存储在边缘MySQL，定时数据备份，并且支持一键从中心进行数据恢复和同步
- 网元支持横向扩容，手动与自动化扩缩容
- 由于限流问题，单个VPC的配置集中在一组内的单台机器
- 单台ESG处理EIP，采用VRRP协议负载均衡
- 网元不持久化配置，启动恢复配置，运行过程中巡检配置

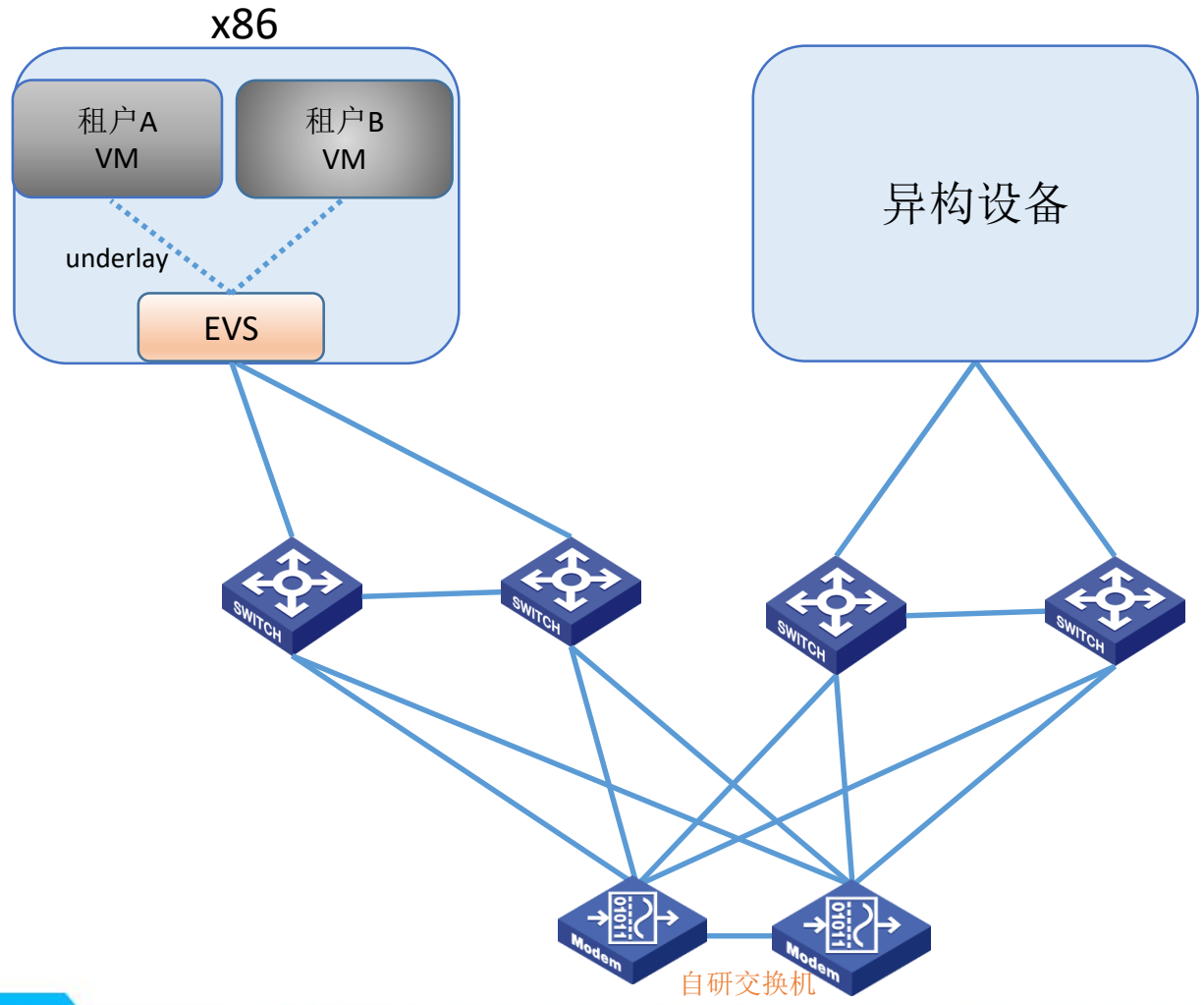
3.0 EVS融合网络

1. 自研可编程交换机，EVS ESG融合部署

2. 支持异构设备，无缝接入边缘云网络

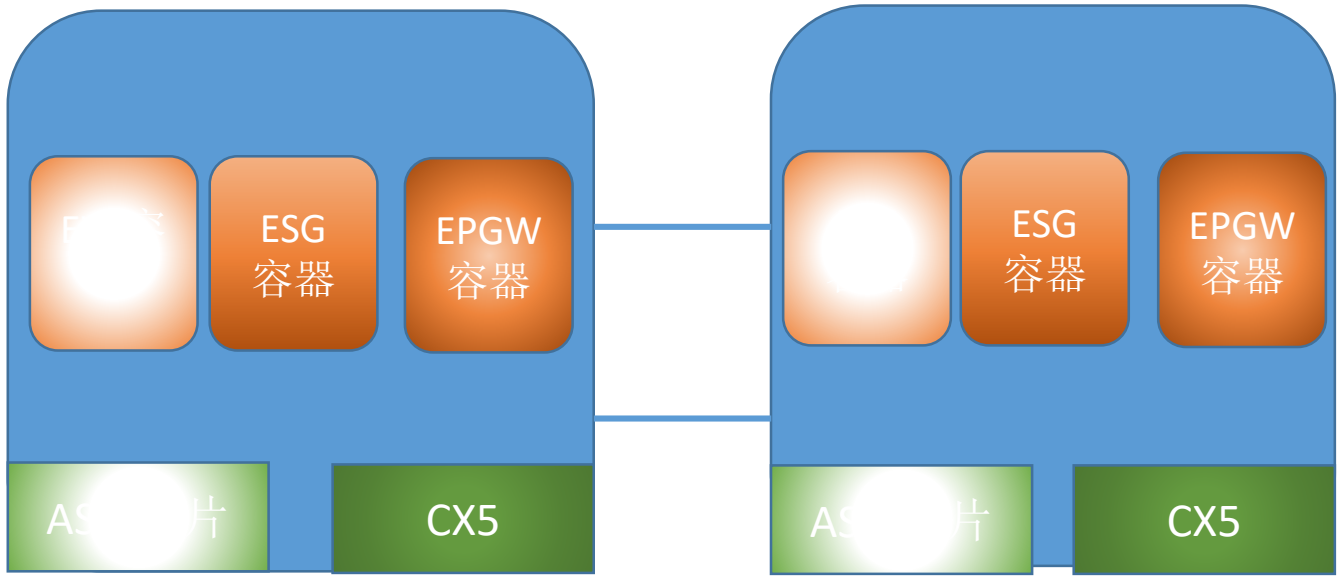
3. 支持虚机，裸金属混布

4. 完整支持VPC网络能力，单台LB，NAT能力200Gb/s



3.0 EVS融合网络

- 节点成本降至1/4
- 芯片T级安全防御
- 芯片精确流量统计
- 芯片硬件HGW加速



1. 支持系统难度增加
    1. 系统装配维
    2. 纳管上线流程
    3. 容器化流程
2. 网络复杂度
    1. 多网元分流策略
    2. 多网元资源分配
3. 稳定性挑战
    1. 集中式网关失效
    2. 热升级
    3. 扩容
        1. 单台性能提升
        2. 横向扩充服务器

网络-全链路监控

- 网元核心监控
- 狼烟
  - 图
- Etrace
  - 图

四 未来方向

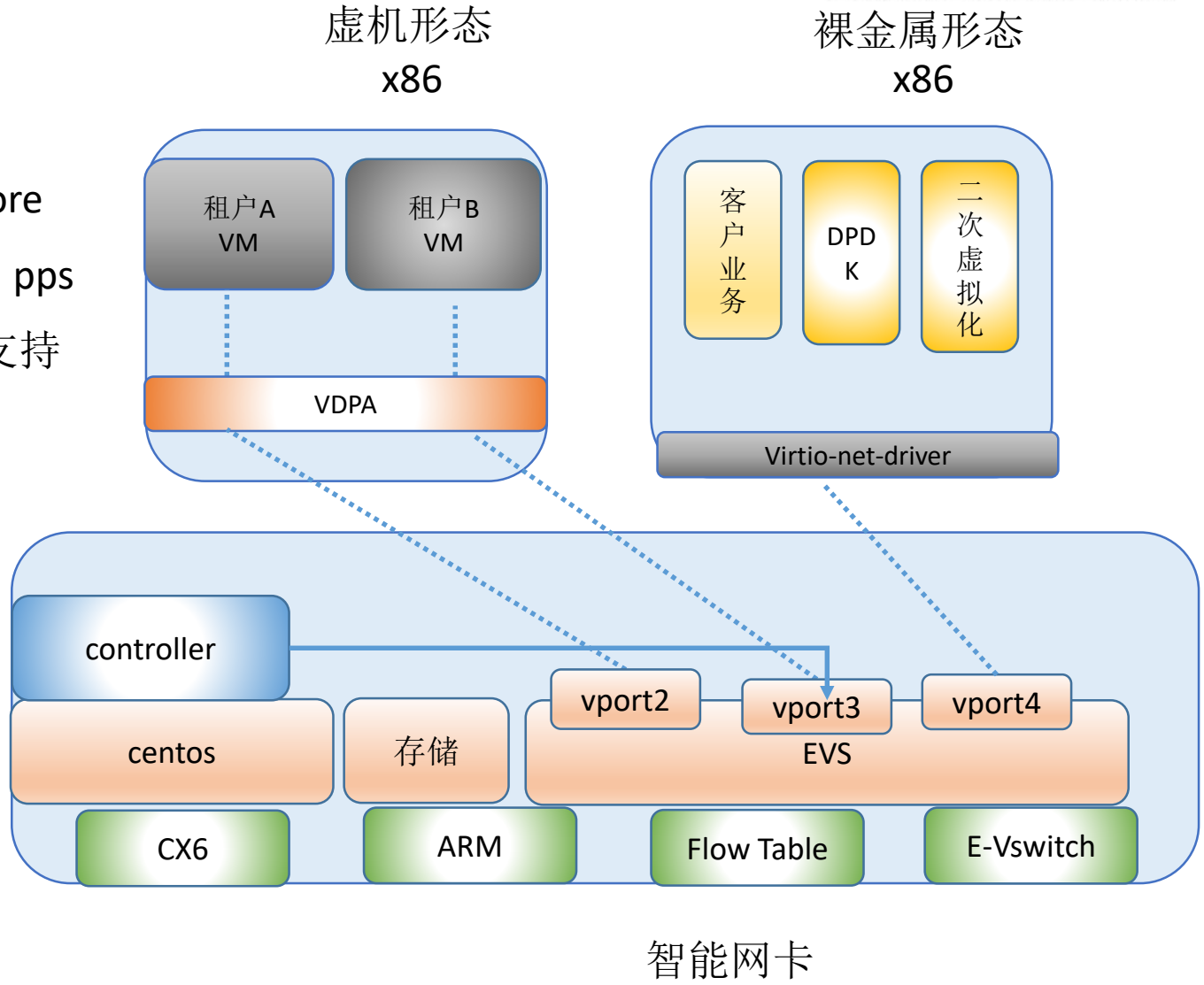
智能网卡

百G网卡

高性能网关

4.0 智能网卡

1. 智能网卡采用8core arm，16G内存。EVS占用4core
2. EVS卸载到智能网卡，性能由6M pps提升到 20M pps
3. 支持虚机与裸金属两种计算形态交付，裸金属支持二次虚拟化
3. 装机流程，纳管流程对接，最小化改动
4. 虚机采用VDPA方案方便热迁移流程
5. 云盘存储卸载到智能网卡
6. 解决大象流问题



THANKS