

实现高性能、高可用的公有云网络

阮有明

Date 2016.12.01



京东云
JCLOUD.COM

C 目 录 contents



一

从传统网络说起

二

京东云网络架构

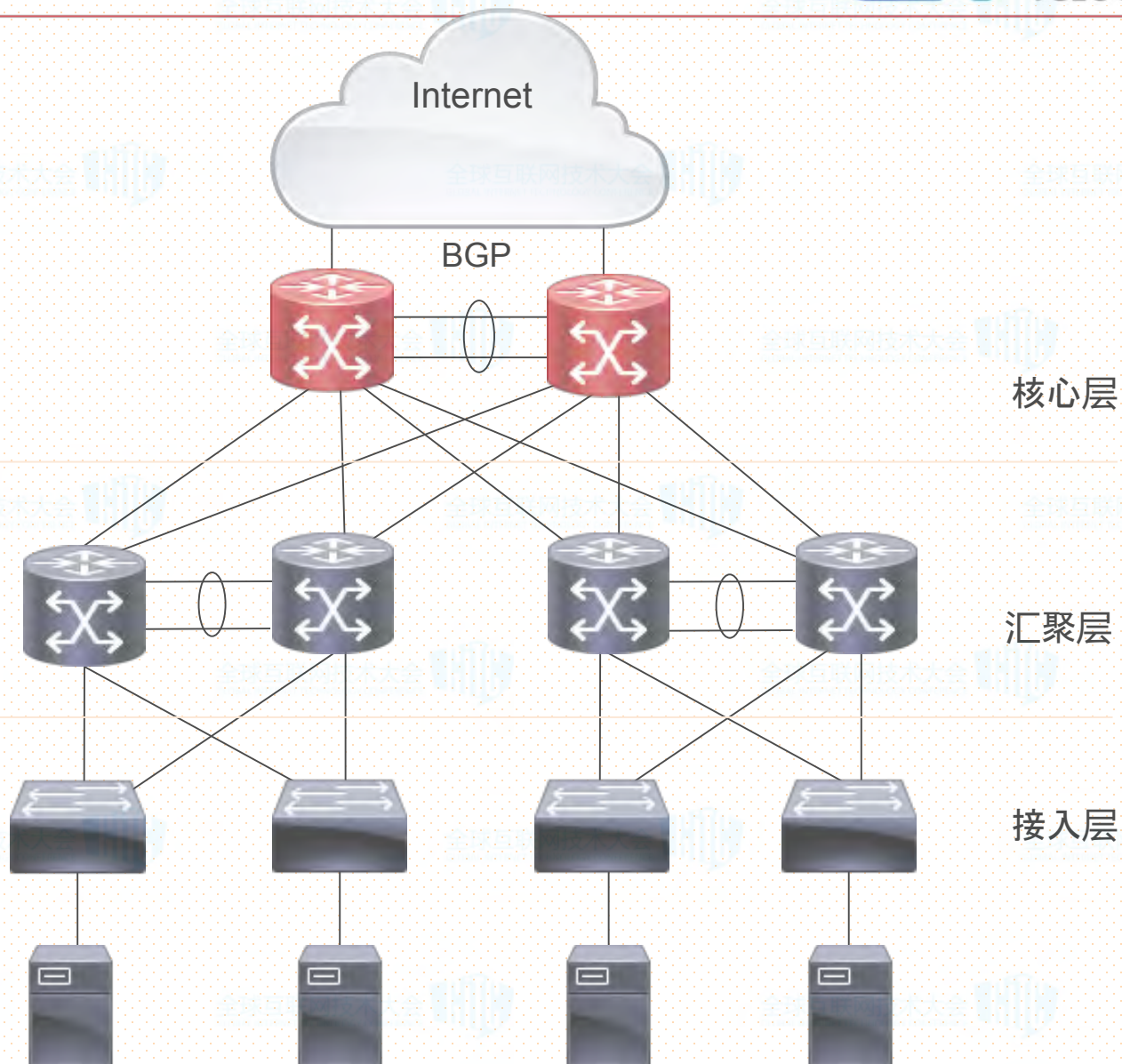
三

高可用性设计

四

vPFE和性能设计

传统数据中心网络结构



云计算在传统网络数据中心的困境


- 计算、资源、软件和网络紧密耦合
- 难以运维
- 难以扩展
- 成本高昂



SDN / NFV 的出现开启了数据网络的新时代，数据中心网络软件化，虚拟化，但NFV同时也有一些挑战

- 可靠性问题
 - 通用服务器的可靠性要低于专业网络设备
- 网络性能问题
 - 通用服务器和芯片在单机性能上无法和专业网络设备、芯片比较
- 运维问题
 - 公有云数据中心数万台服务器规模的运维
- SDN / NFV标准化问题

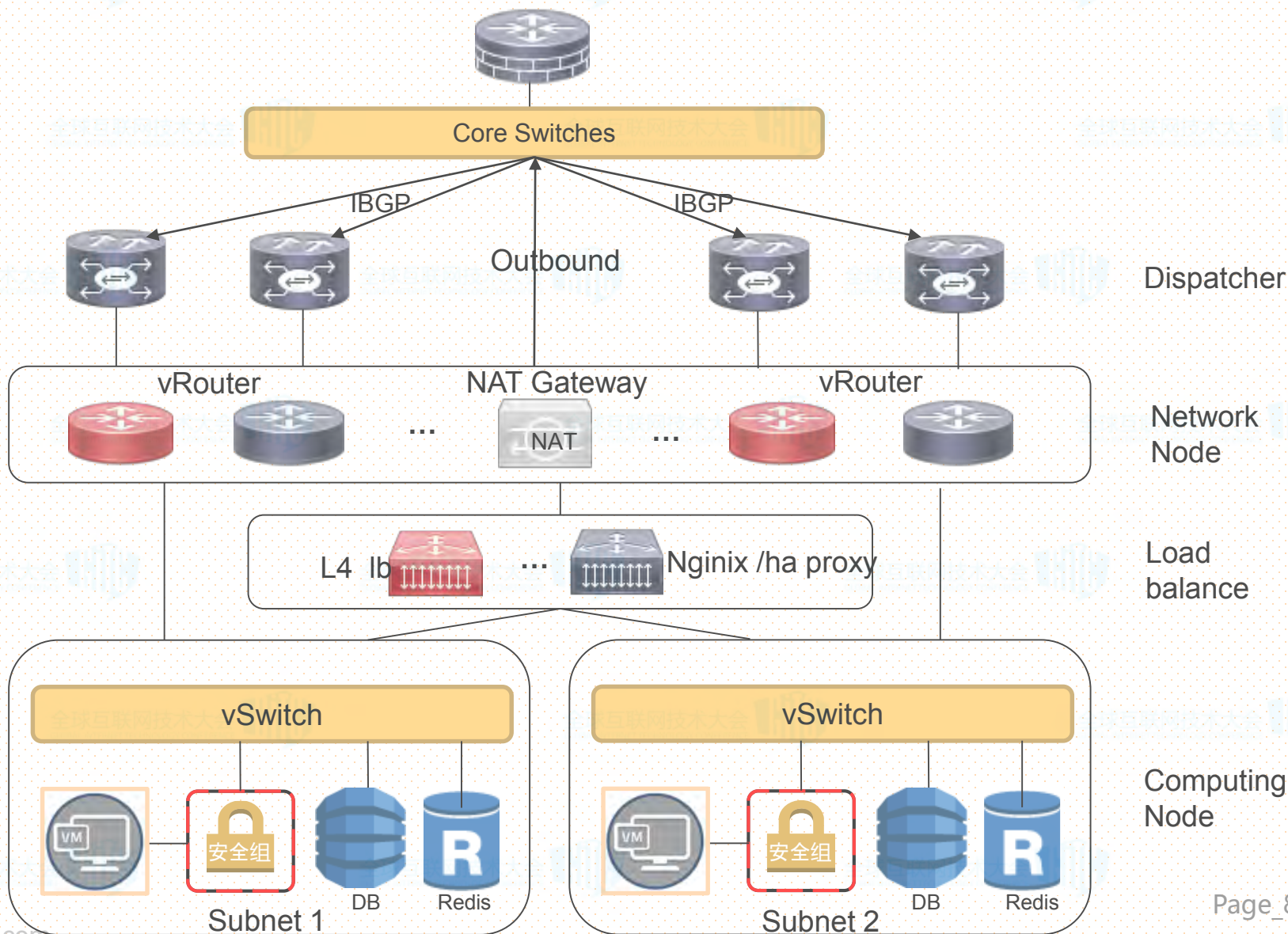
京东云网络架构

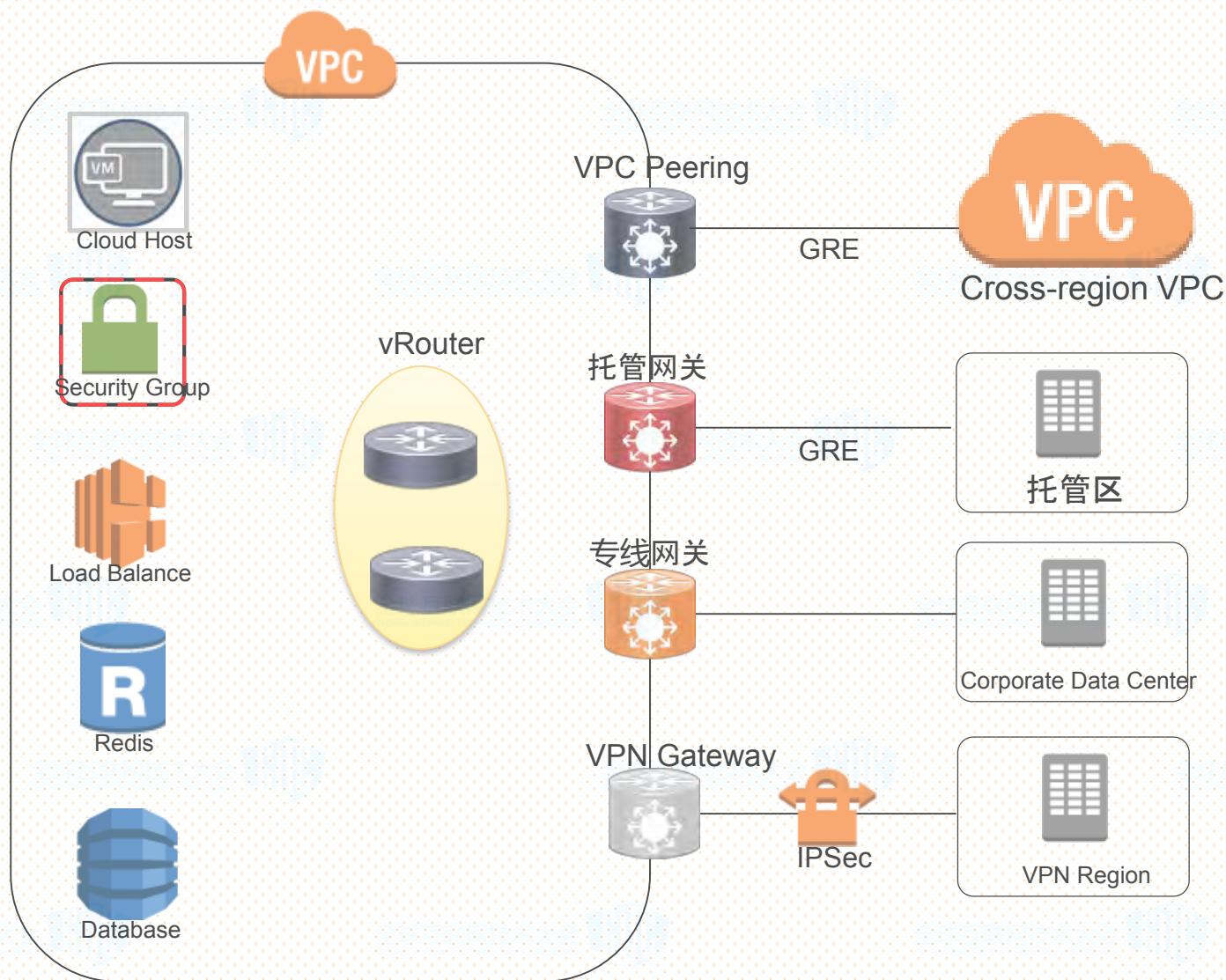


京东云在SDN/NFV做了
哪些探索？


- 私有网络VPC
- 同地域和跨地域VPC Peering
- 任意子网划分（无CIDR限制）
- 子网路由和ACL
- 1:1 NAT和NAT网关
- 安全组
- 高可用Load Balance
- 灵活的混合云接入方式

VPC 网络架构





高可用性

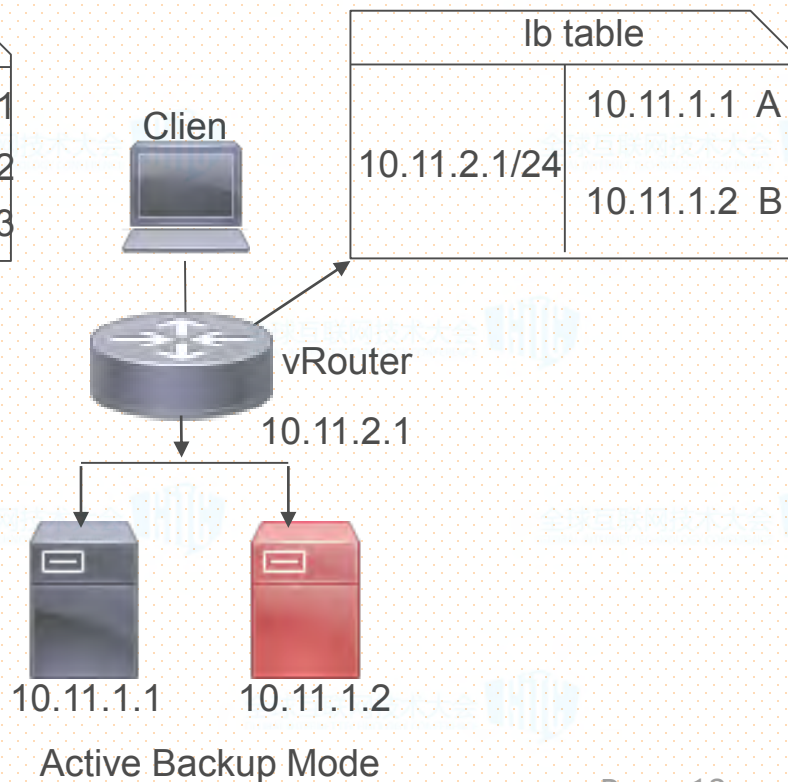
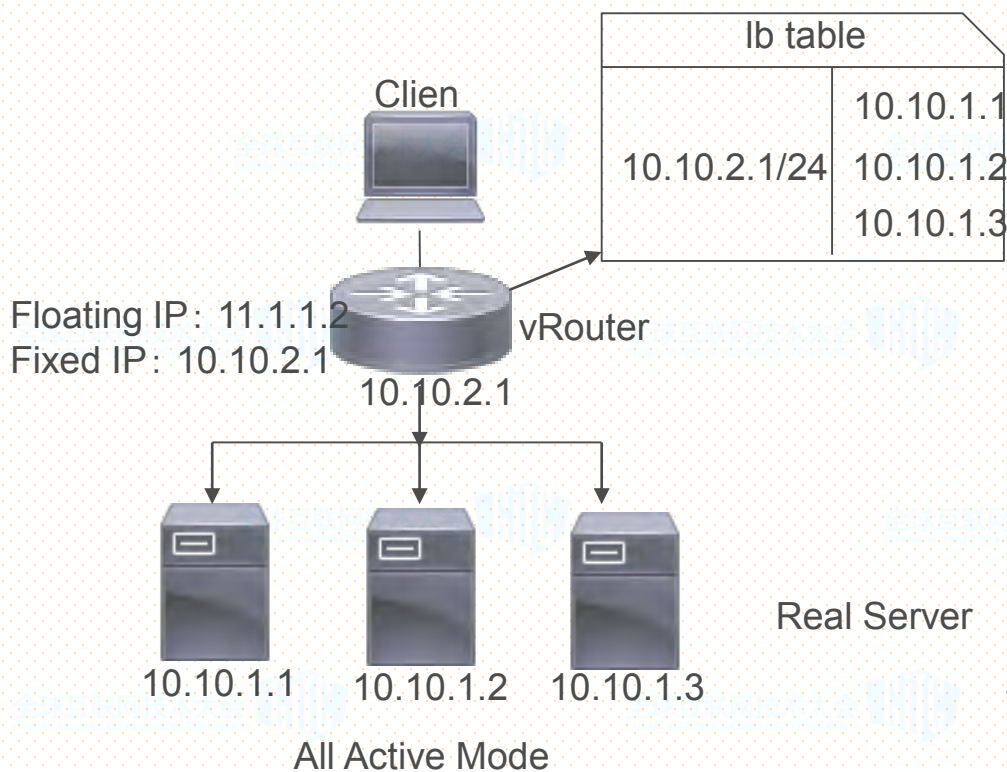


京东云在网络各个节点、层次都做到了灵活全方位的高可用！

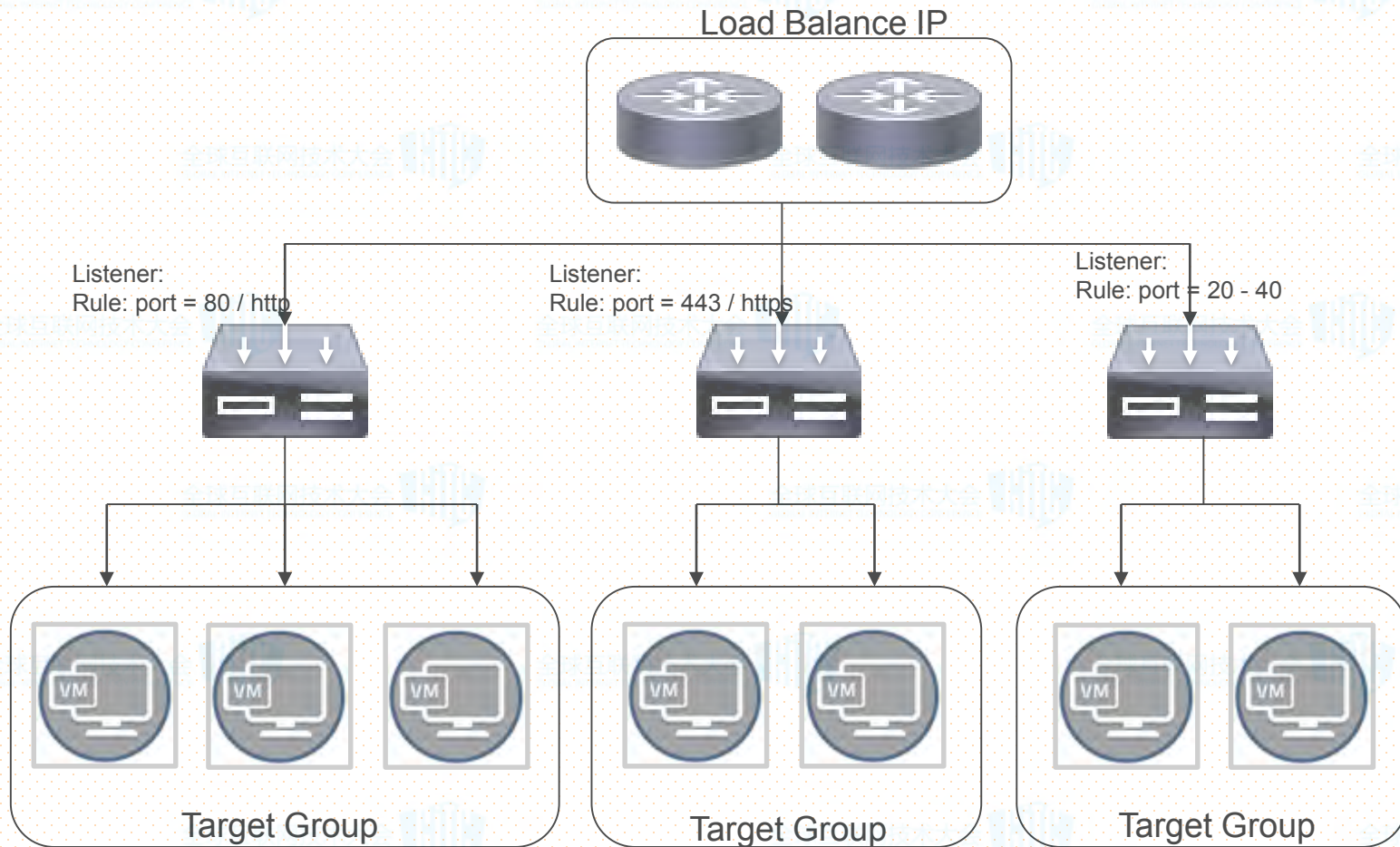
- 链路聚合 (Link Aggregation) , LVS DR
- ECMP, VIP
- L4 load balance
- Ngnix和L7 Health Check
- 全网链路监测

高可用VIP

- 流量分发，一个对外的IP可以分发到多个Backend Real Server
- 绑定Floating IP，对外只需要一个公网IP
- Active - Active模式、Active - Backup模式



L4 – L7 负载均衡



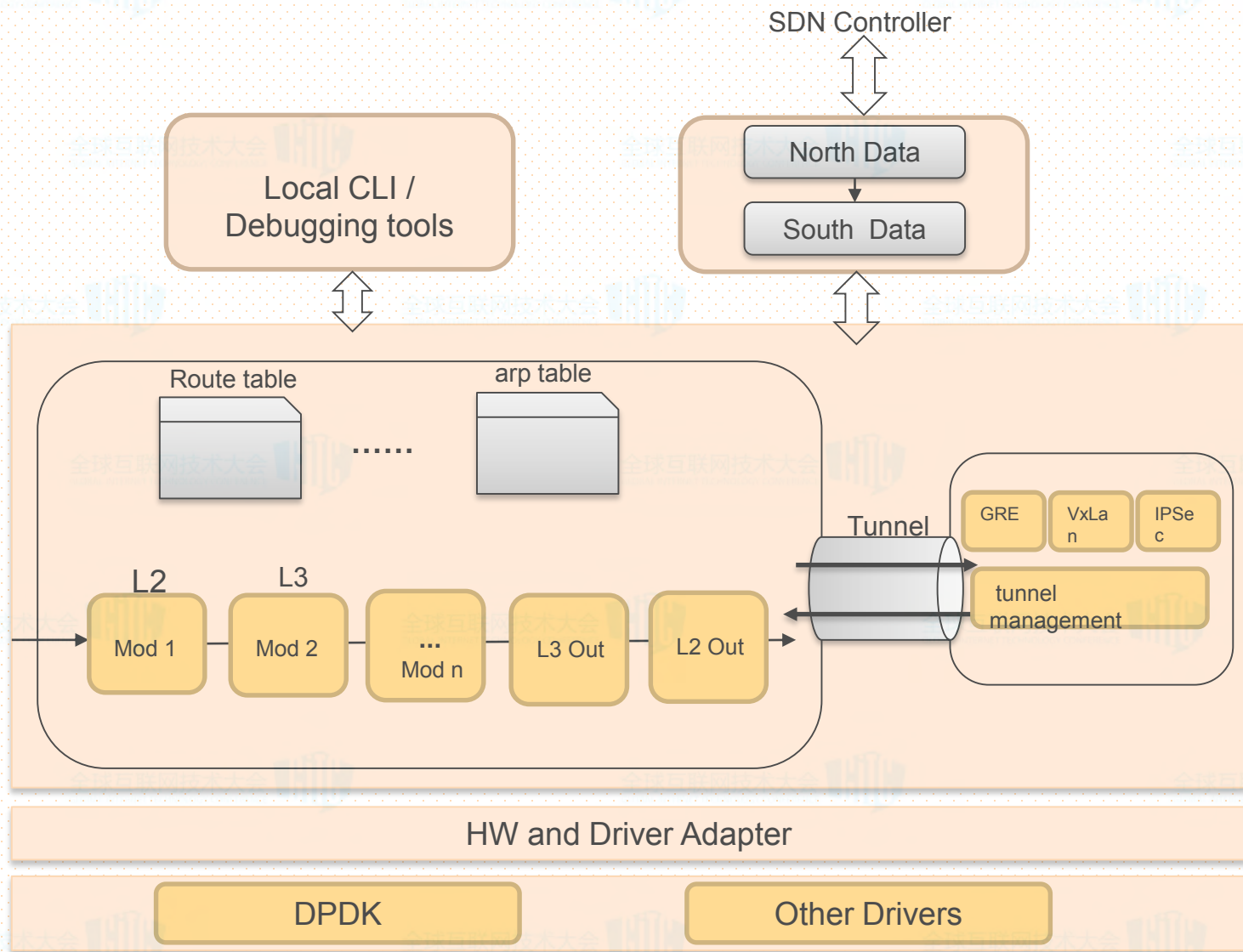
数据转发和高性能



vFE – Virtual Forwarding Engine

- 京东云数据面转发平台
- 运行于用户态，可以基于DPDK或者其它驱动之上
- 设计目的：
 - 高度可扩展性
 - 未来可以支持交换、路由、LB、防火墙等数据面
 - 高性能

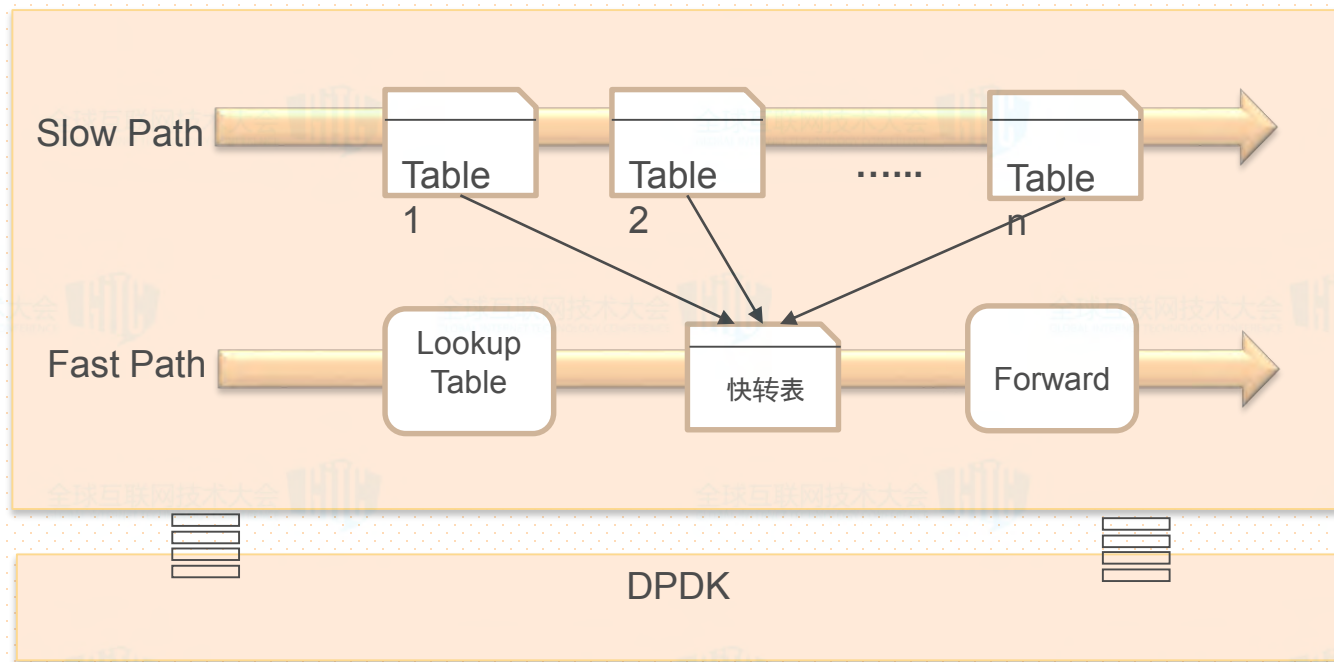
- 在vFE的data path上可以灵活增加和删除各种处理单元，输入和输出参数通过统一的接口进行传递
- 支持传统路由、交换等主要特性，将来支持防火墙
- 支持Overlay和Underlay等虚拟网络特性
- 支持公有云等特定的需求：如Overlay下，路由、流表、NAT、ACL等具有子网属性
- 支持VxLAN、GRE、IPSec等隧道特性



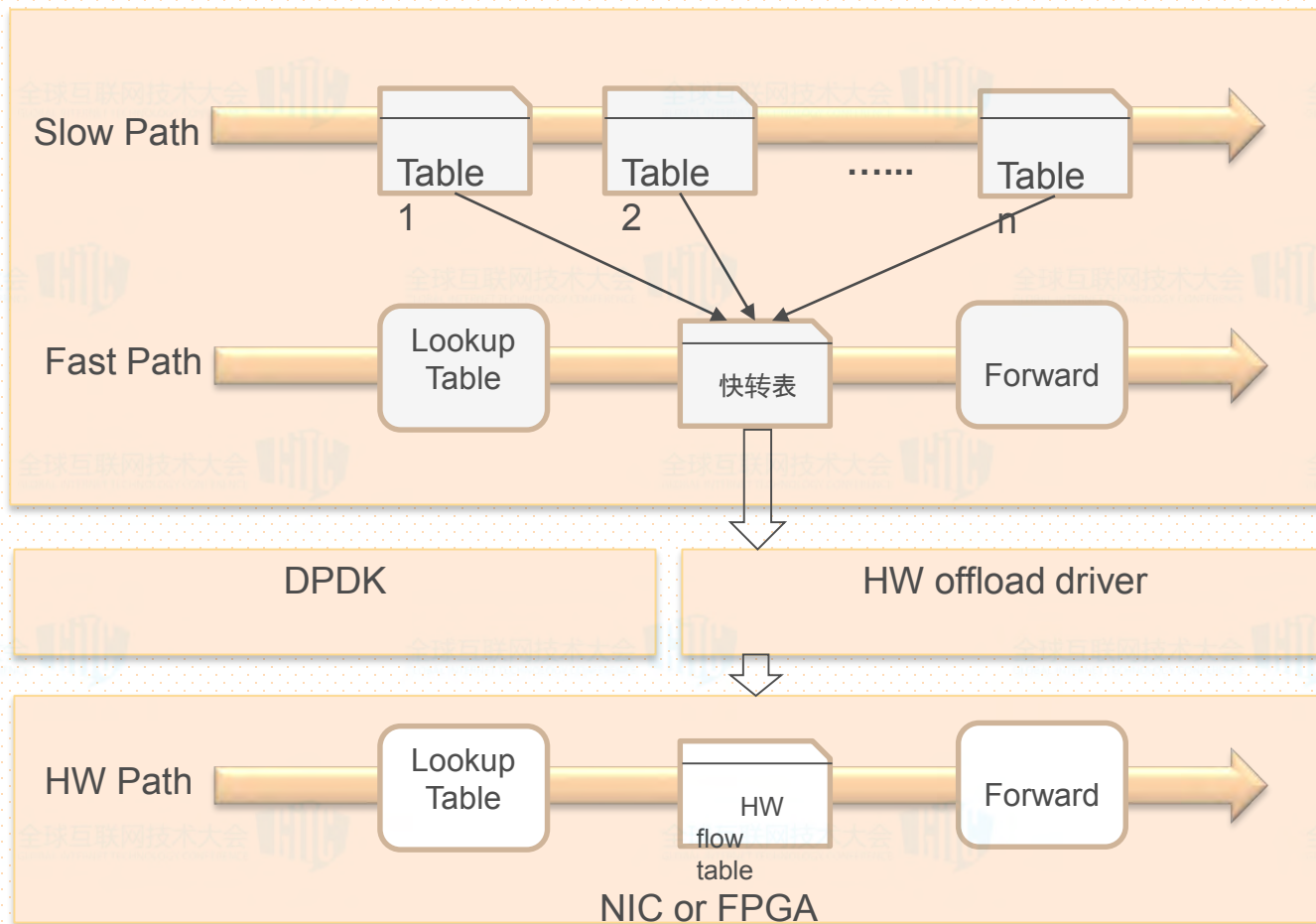
- 根据业务场景选择Run-to-complete
- Data Path无锁转发
- 表项查找算法优化， $O(1)$ 级查找
- 提高Cache的利用率
- 核心表项预分配连续内存
- Key Path保持精简的设计，新增节点不会影响Key Path
- 编写高效率的代码

快转发路径 (Fast Path)

- 相较Slow Path各个节点有自己的表项，Fast Path生成一张快速转发表，后续报文根据快转表转发

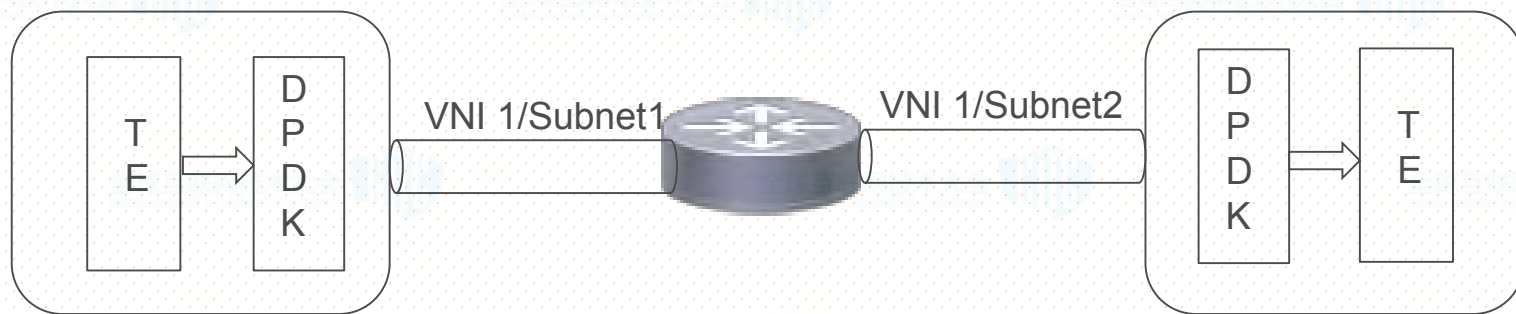


- 通用**CPU**的网络性能将会遇到瓶颈
- 价格低廉的智能网卡、**FPGA**芯片是硬件加速的很好选择
- 硬件实现**Forwarding Plane**在传统网络技术非常成熟，可以**Offload**到硬件，利用硬件的高性能优势包括：
路由、交换、**ACL**、基本防火墙的**session**、加解密、校验和、基于统计的**DDOS**功能等等



Testbed: Intel (R) , CPU E5-2698 v3 @2.3GHz

Packet size: 96bytes, 10GE, 40GE Intel NIC



vFE单核: 9.8M pps

OVS over DPDK四核: 7.6M pps

Thanks

感谢您的聆听

We're hiring !

ruanyouming@jd.com
chenfeng@jd.com