
WOTA
51CTO
World Of Tech 2017
全球架构与运维技术峰会
2017年4月14日-15日 北京富力万丽酒店
ARCHITECTURE


出品人及主持人：

# 张立刚

1号店技术部
电商云平台技术总监

## 云服务架构

# 京东云虚拟网络架构设计实践

WOTA

51CTO

# 陈峰

京东

专家架构师

分享主题：

## 虚拟网络架构设计实践

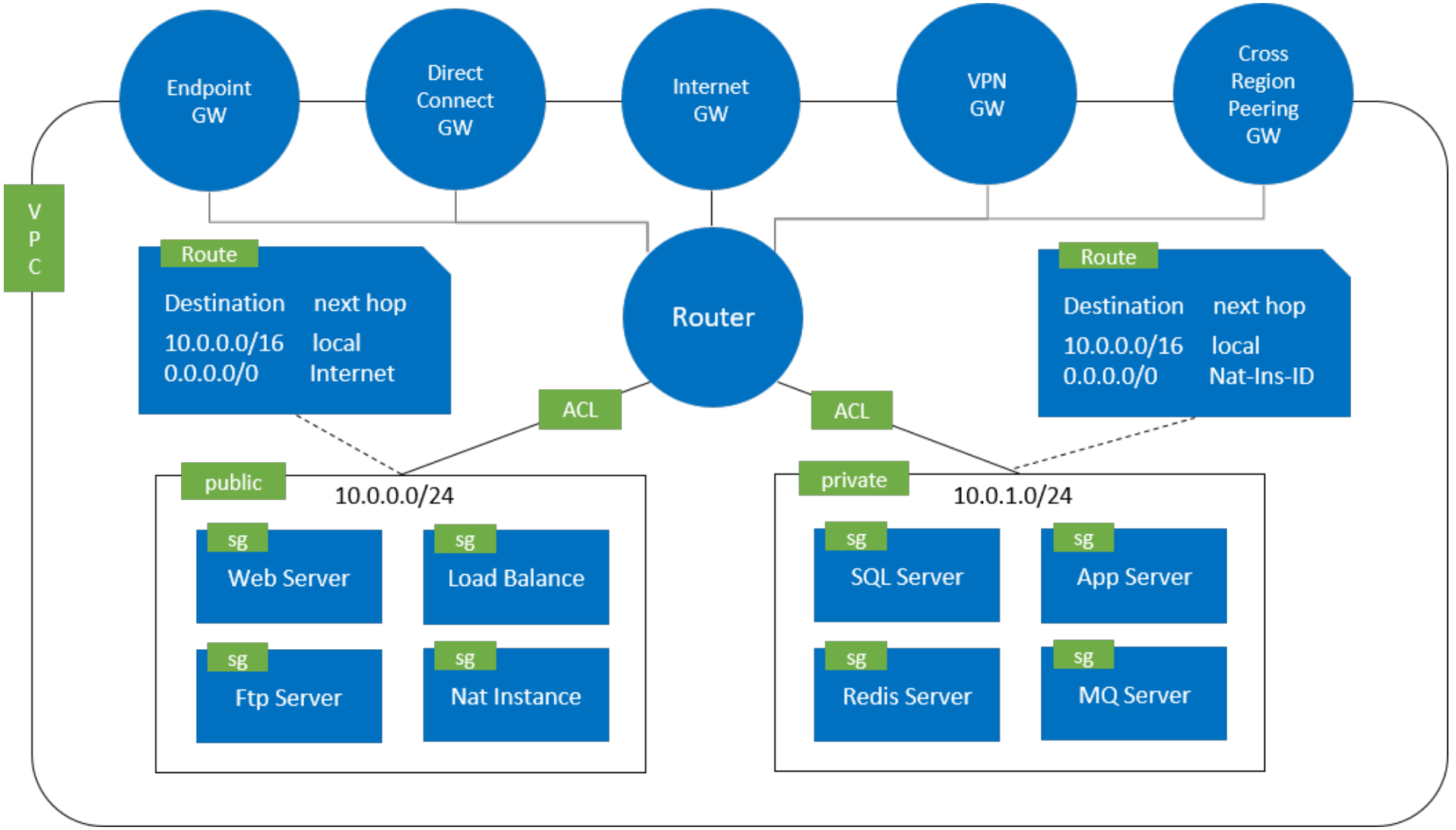
# 议题

- 虚拟网络简介
- 虚拟网络架构设计挑战
- 网络数据面架构设计实践
- 网络控制面架构设计实践
- 虚拟网络未来

# 虚拟网络功能一览

- 每个租户拥有1到多个完全隔离的虚拟网络
- 按业务需要灵活划分子网，无CIDR限制
- 按业务需要不同子网使用不同的路由策略
- ACL实现子网级别的安全控制
- 安全组实现实例级别的安全控制
- 多类型网关满足客户的各种混合云场景

# 虚拟网络典型场景


Endpoint GW
Direct Connect GW
Internet GW
VPN GW
Cross Region Peering GW
VPC
Route
Destination next hop
10.0.0.0/16 local
0.0.0.0/0 Internet
Router
Route
Destination next hop
10.0.0.0/16 local
0.0.0.0/0 Nat-Ins-ID
ACL
ACL
public
10.0.0.0/24
sg
Web Server
sg
Load Balance
sg
Ftp Server
sg
Nat Instance
private
10.0.1.0/24
sg
SQL Server
sg
App Server
sg
Redis Server
sg
MQ Server

# 虚拟网络架构设计挑战

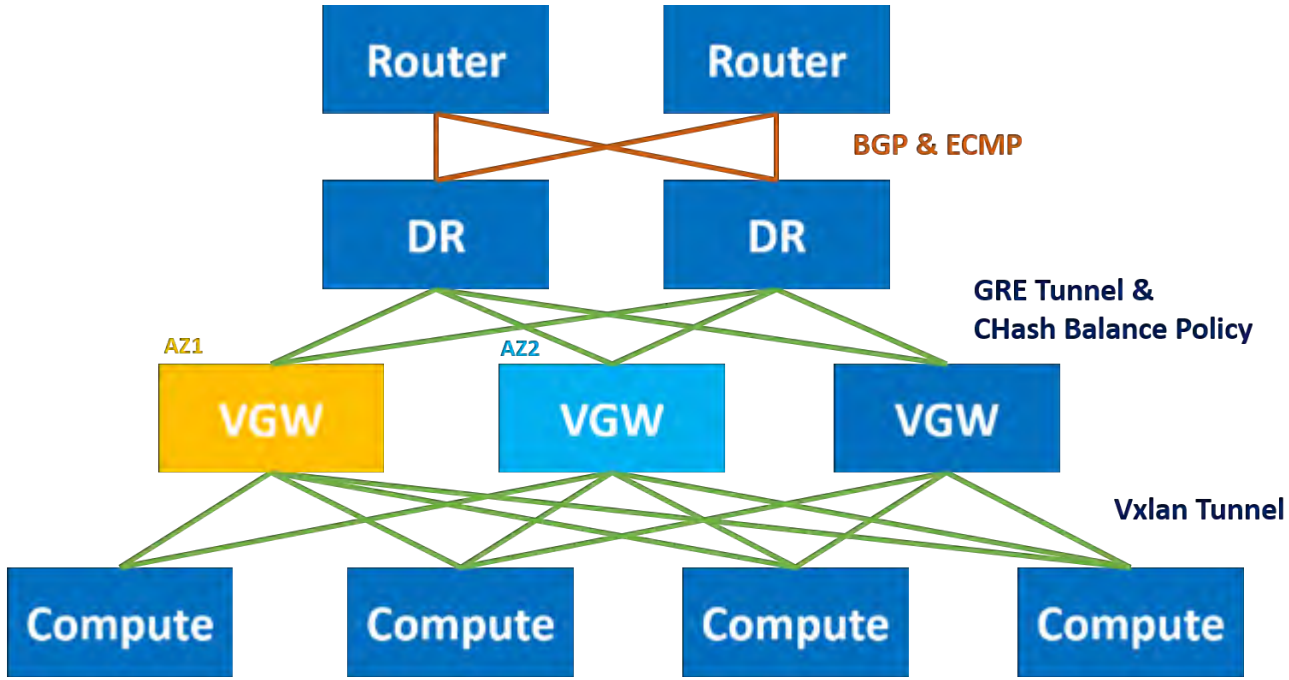
- 虚拟网络的子网和VIP可以跨数据中心和可用区
- 规模庞大的虚拟网络和子网路由表数目
- 快速迭代升级的虚拟网络组件，满足用户新的业务场景
- 用户的虚拟网络流量差异变化巨大，无法预先规划
- 共享虚拟网络组件保证每个客户网络性能SLA
- 数据中心数万台网络节点的SDN控制信息的准确快速下发

# 网络数据面架构要点

- 虚拟网络组件采用三层技术互联
- 虚拟网络节点的功能单一化和集群化
- 核心路由集群无状态，可动态横向扩展
- 核心路由和网关节点均为Active-Active模式
- 东西向跨子网采用分布式路由器和集中路由器结合

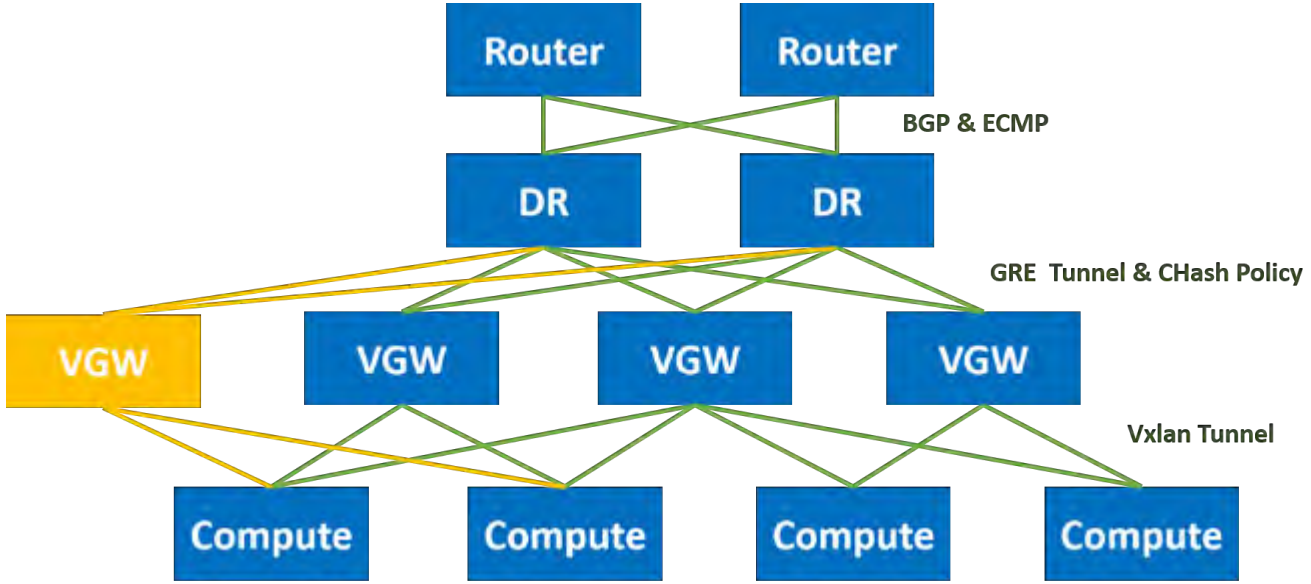
# 三层互联技术

- 物理网络采用Spine-Leaf架构
  - ➢基于VLAN技术对于部署和规模限制很大
  - ➢核心转发面去中心化
  - ➢核心交换机的私有双活协议 VS 成熟BGP协议ECMP
- DR和VGW采用GRE隧道技术 VS Direct Routing



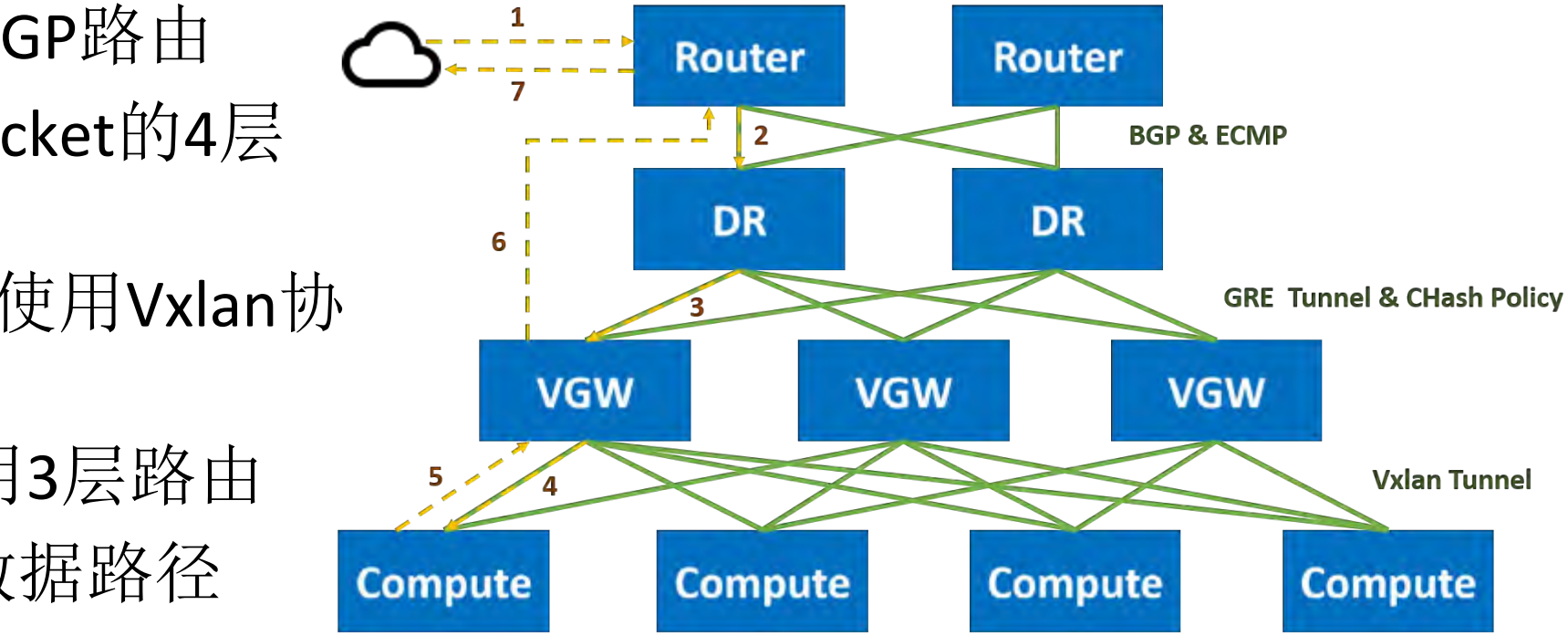
# 横向扩展架构

- 网络节点的功能单一和集群化部署，保证核心转发路径的稳定性，应对快速业务变更
- 网络节点采用Active-Active模式
- 核心转发路径只提供纯路由功能，保证无状态，快速扩展
- 有状态转发路径使用一致性HASH算法



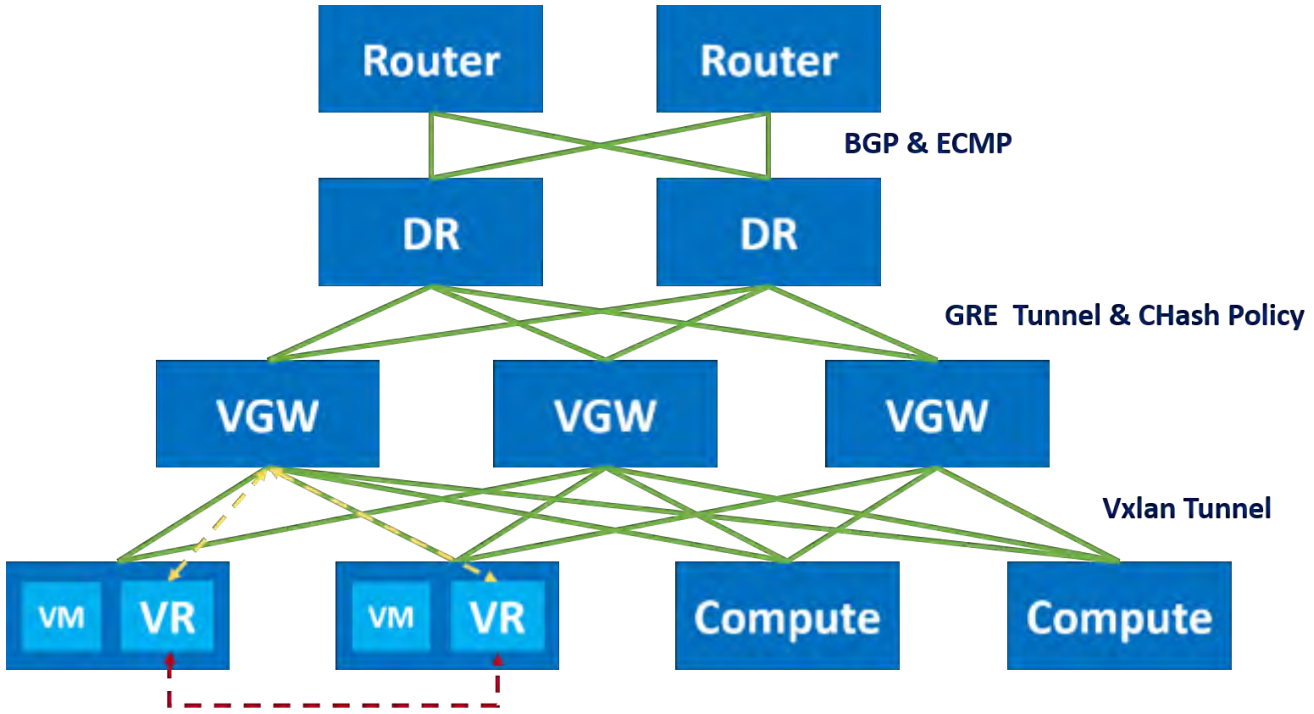
# 包流向分析

- Router -> DR使用BGP路由
- DR -> VGW 使用Packet的4层Forward
- VGW <-> Compute使用Vxlan协议
- VGW -> Router使用3层路由
- 南北向非对称的数据路径

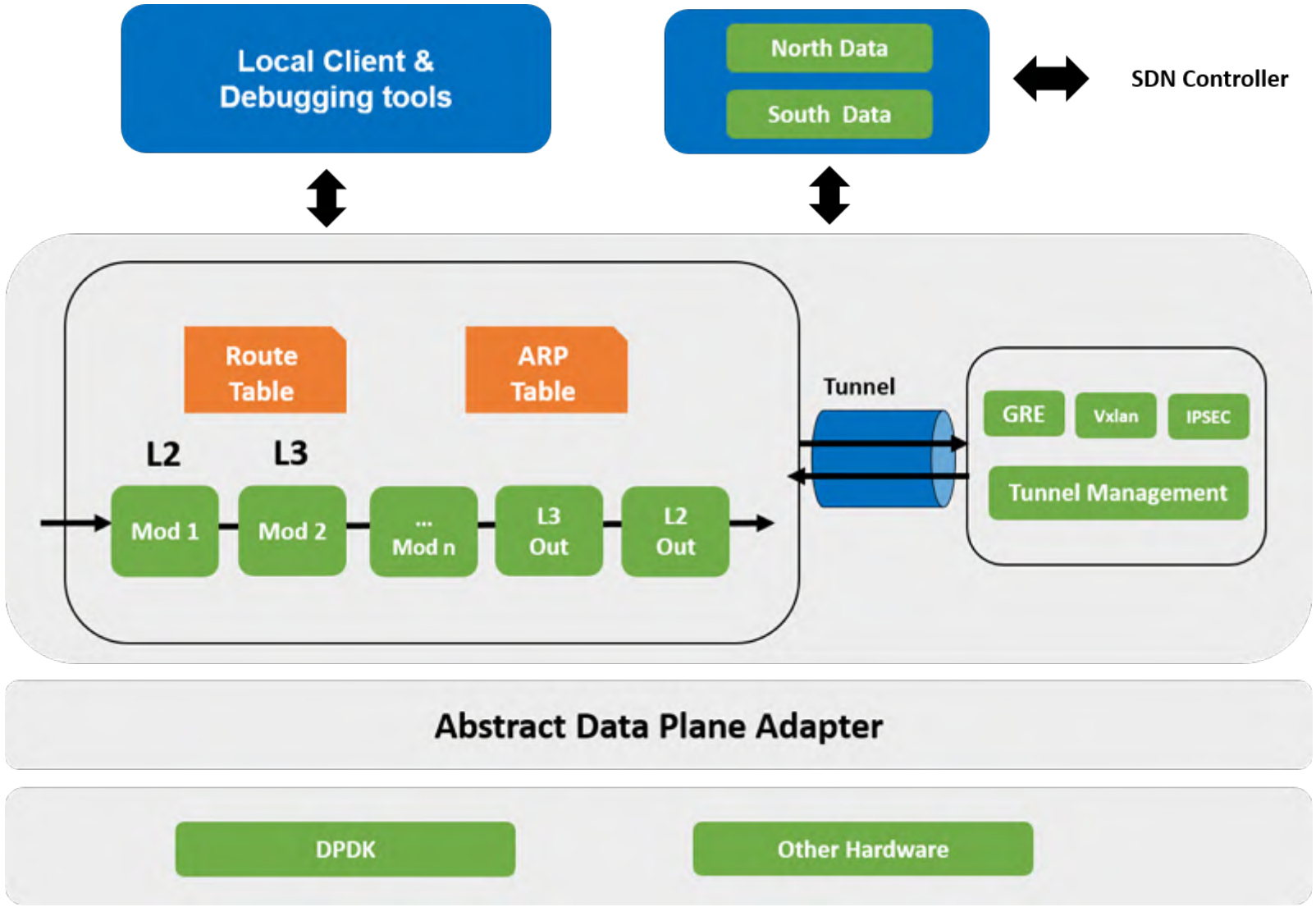


# 分布式和集中路由结合

- 东西向跨子网使用分布式路由
- 虚拟网络子网主机非常多并且需要支持动态路由协议时，考虑使用集中式路由
- 集中式路由不意味着单点，通过ECMP实现高可用

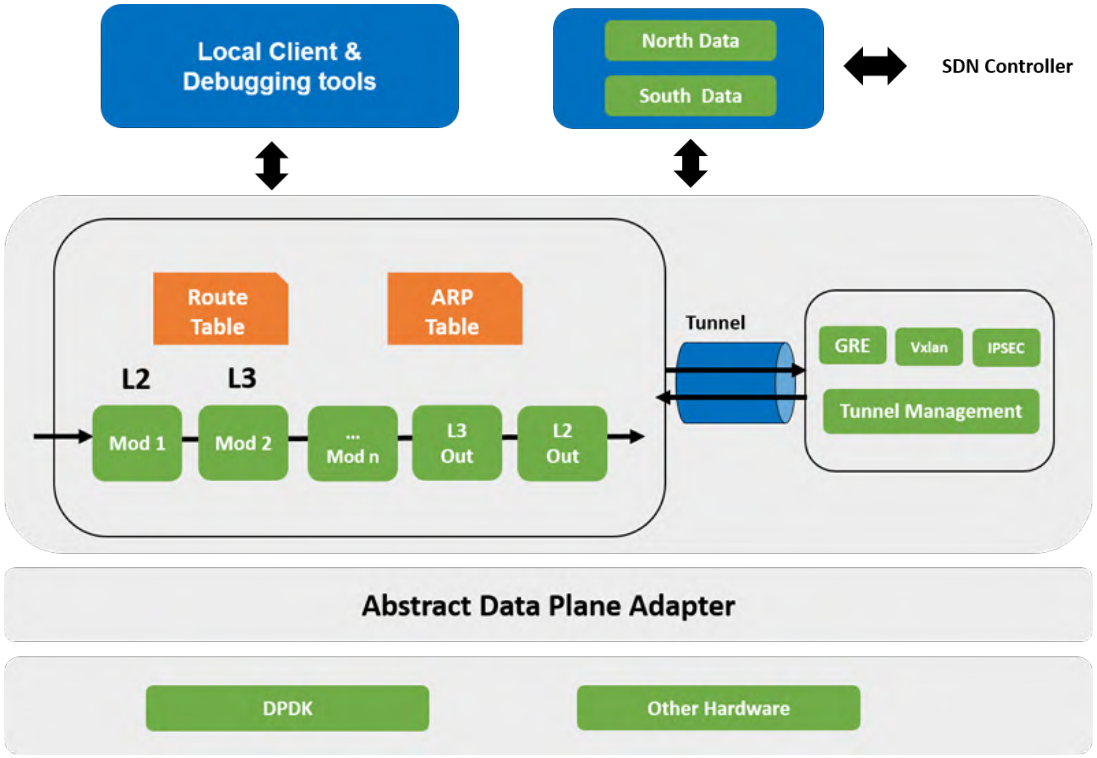


# 虚拟网络转发引擎vFE



# vFE特性

- 用户态数据面转发平台
- 虚拟路由和网关节点的基础
- 抽象数据面API，适配DPDK和者硬件驱动
- 高度可扩展，支持交换、路由、LB、防火墙等数据面
- 灵活增加处理单元，输入和输出参数通过统一的接口进行传递

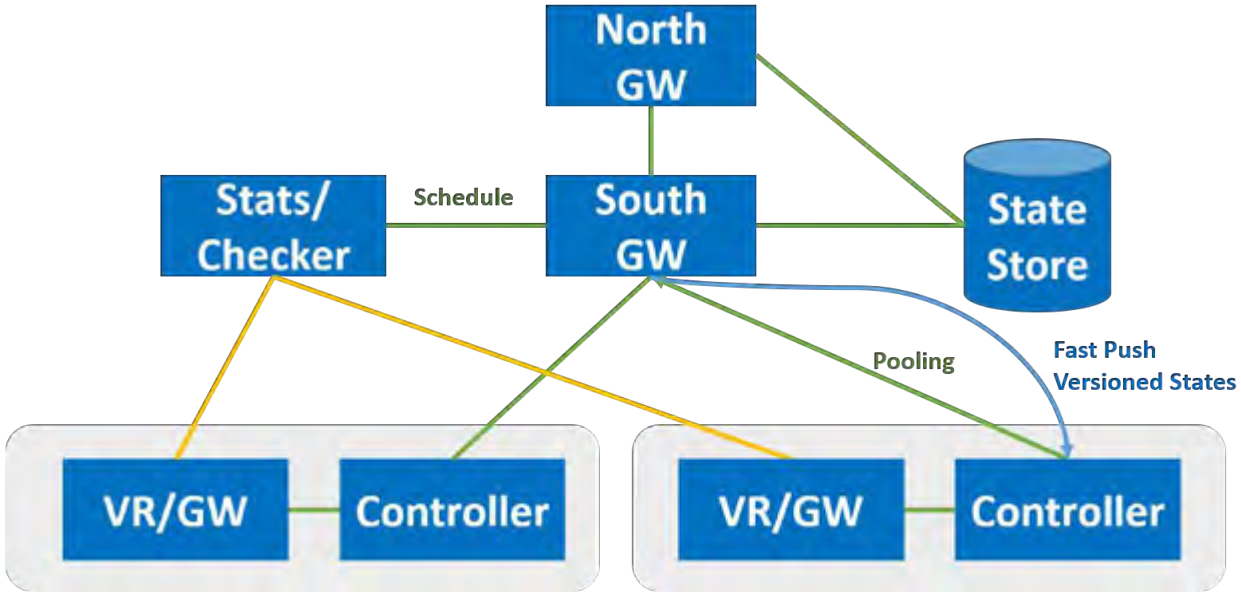


# vFE性能优化实践

- Run-to- complete和Pipeline模型结合
- 核心表项查找算法做到O(1)级查找
- 基于统计的有条件Qos实现
- 快速转发表优化查表速度
- 核心转发路径无锁，引用数据结构延迟释放
- 提高Cache命中率
  - 针对应用优化的内存池
  - Thread Local的数据结构
  - 核心表项预分配连续内存
  - 内存对齐

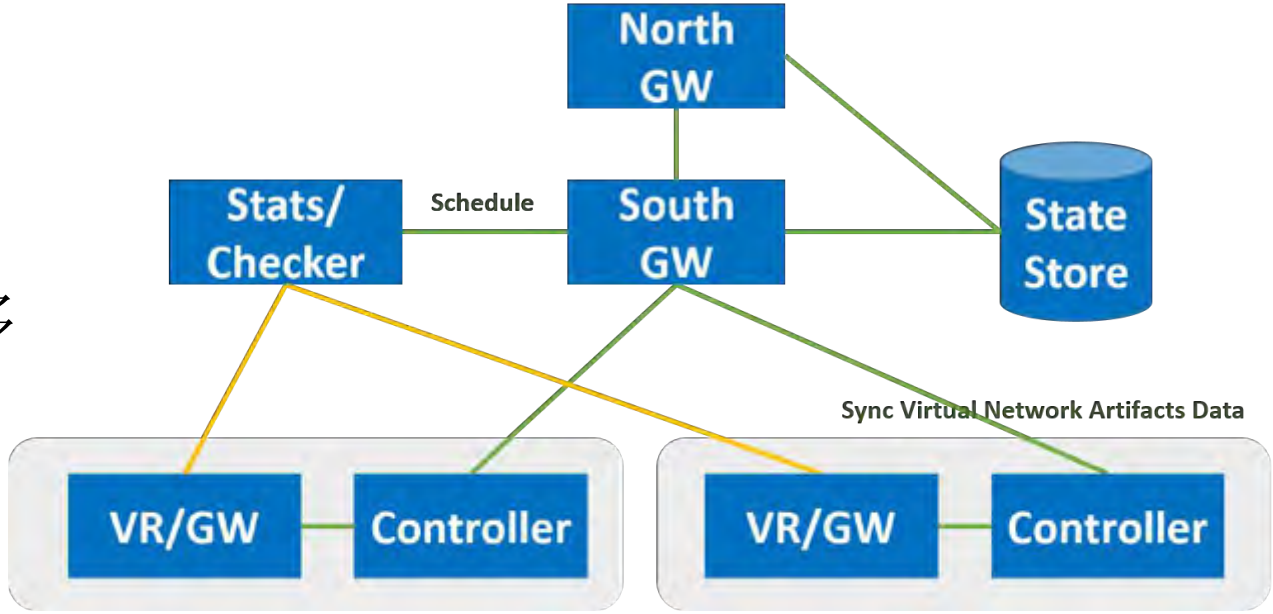
# SDN控制面架构实践

- 虚拟化网络期望拓扑状态和实时运行状态集中存储，便于集中决策和状态管理
- 网络状态的数据模型有全局版本，便于状态同步和校验
- 同步和异步的状态更新机制。异步的Fast Push不要求完全可靠，快速心跳同步最新的拓扑状态



# SDN控制面架构实践

- 南向网关和北向网关无状态，可横向扩展
- 网络状态同步按业务建模，与Data Path实现无关，通过Adapter 支持多种Data Path实现同时在线
- 根据实时监控数据触发调度；快摘除，慢恢复
- 给DEVOPS提供原生支持



# 虚拟网络未来

- 混合云的大规模实施
- 复杂的虚拟网络组网需求
- 网络硬件由黑盒向白盒演进
  - 智能网卡、ASIC、FPGA
  - 可编程交换机
- 实时的全网链路探测


WOTA
51CTO
Thank you !