



第十届中国IDC产业年度大典  
The 10<sup>th</sup> Internet Data Center Conference



# SDN与CDN “在一起”

蜗牛游戏  
李嘉俊  
202016/1/11





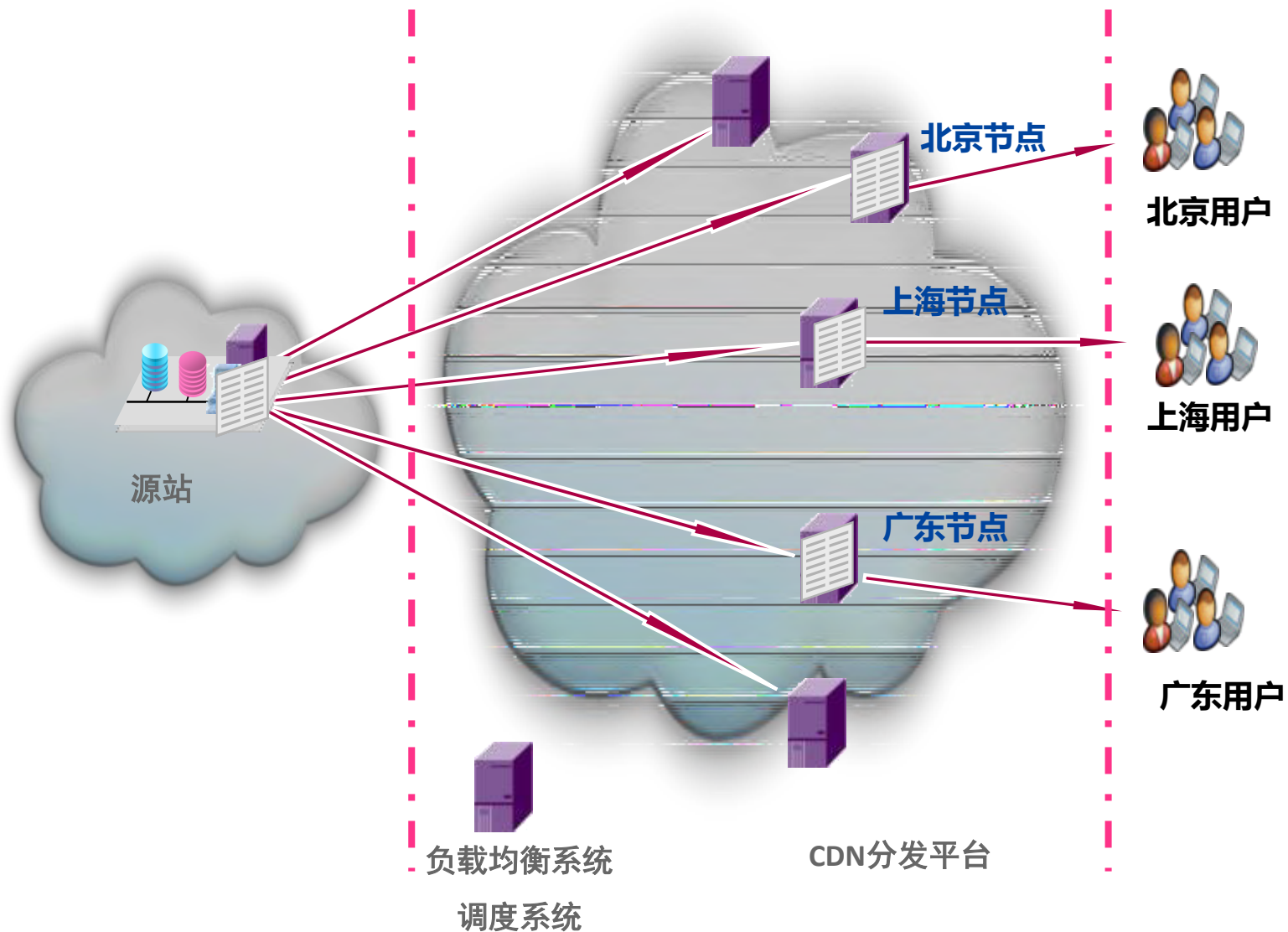
IDCC

175CM

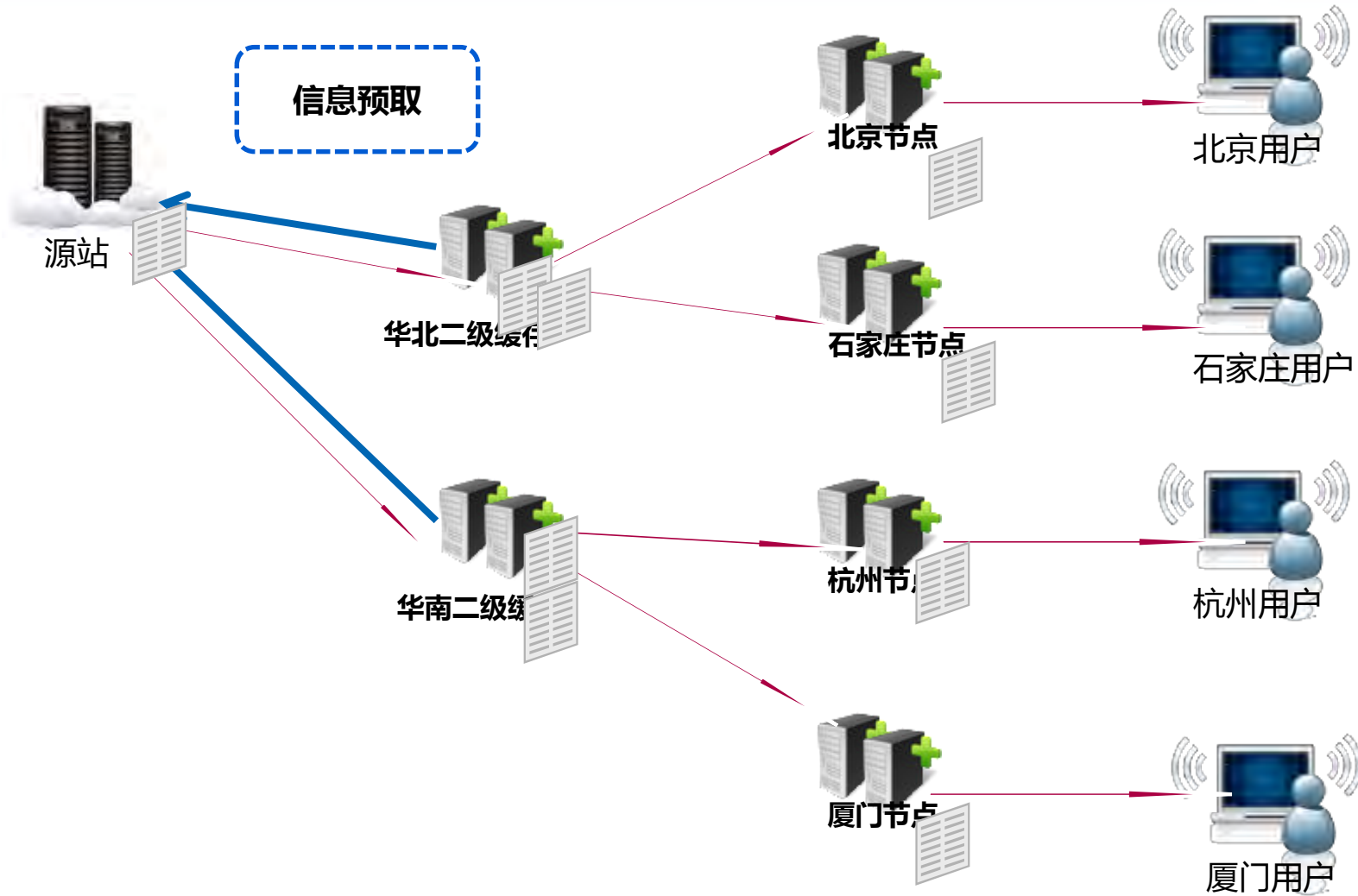
163CM



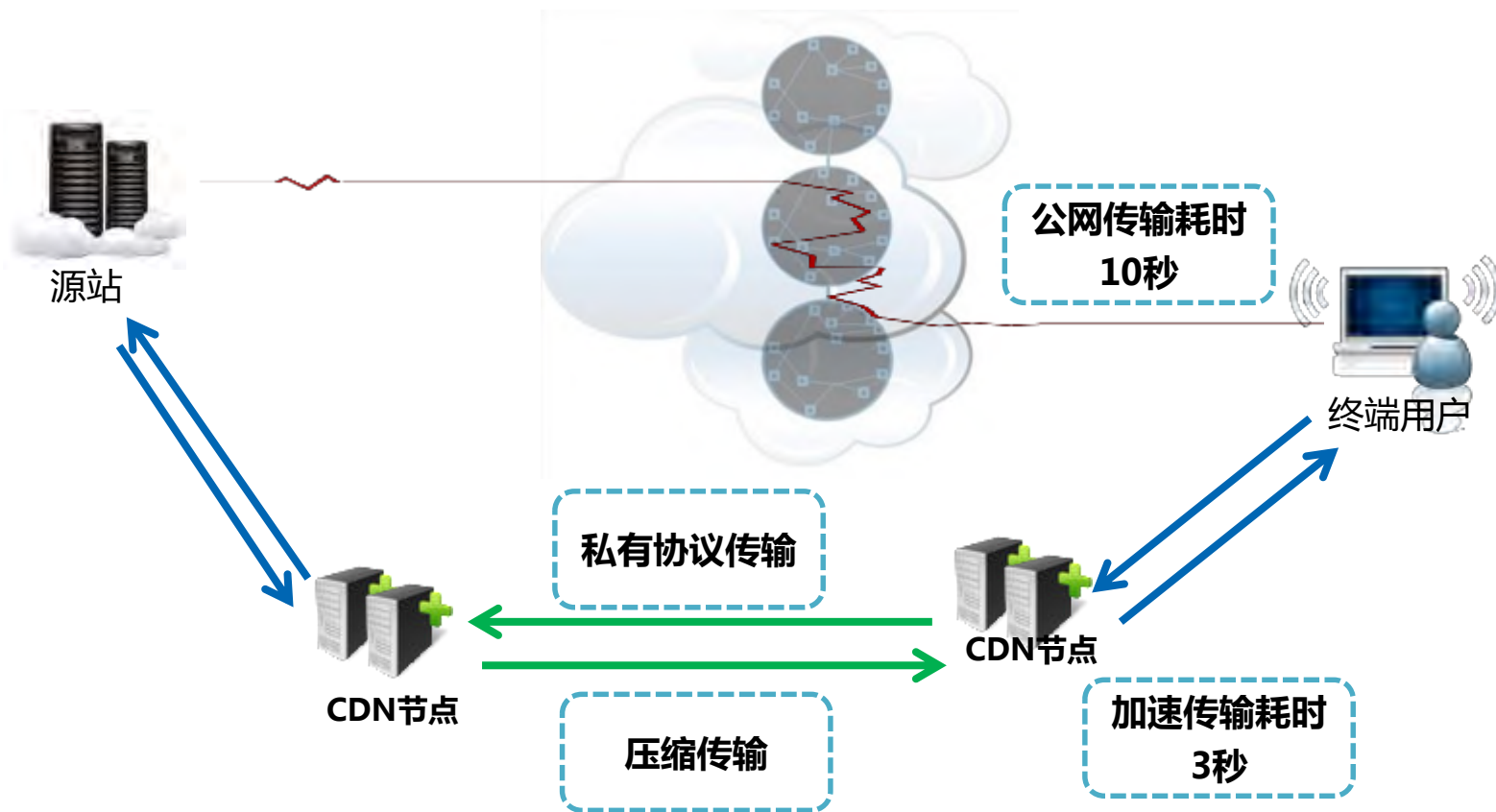
## 分发系统基础架构



提高服务响应和加速效果

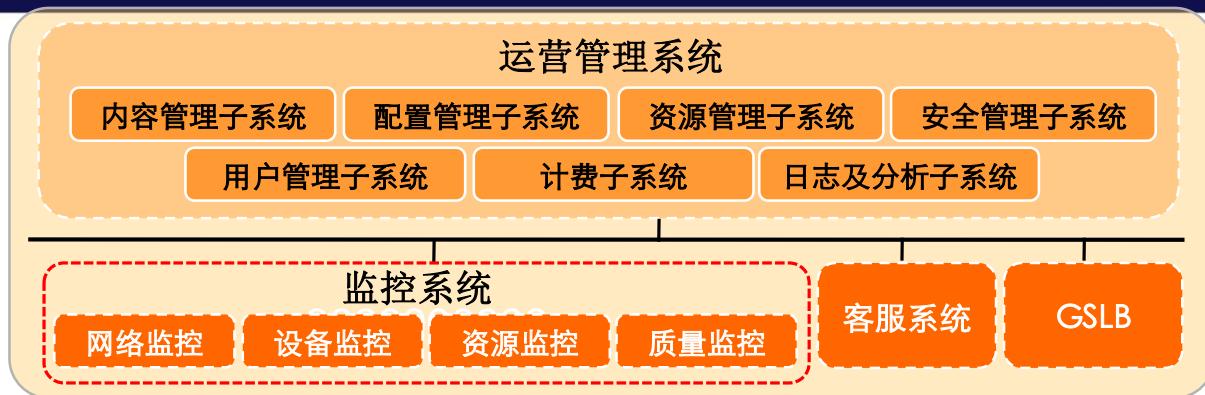


## 提高传输效率

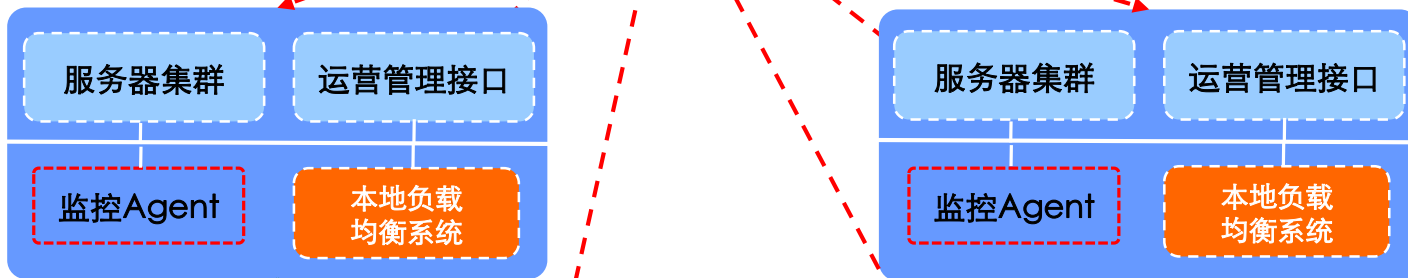


## 分发系统逻辑架构图

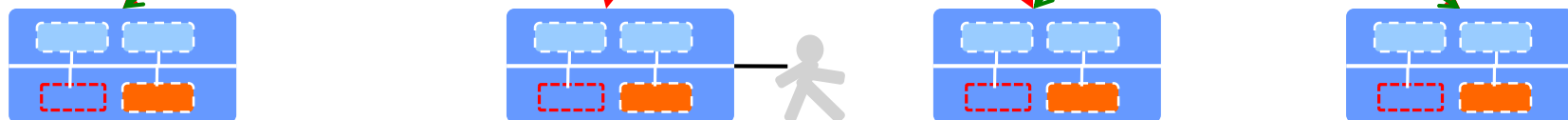
全国  
管控  
中心



内容  
中心



服务  
节点





## SDN的特征

### 集中控制

- ✓集中控制使得全局优化成为可能，比如流量工程、负载均衡
- ✓集中控制使得整个网络可以当作一台设备进行维护，设备零配置即插即用，大大降低运维成本，类似的技术：

### 开放接口

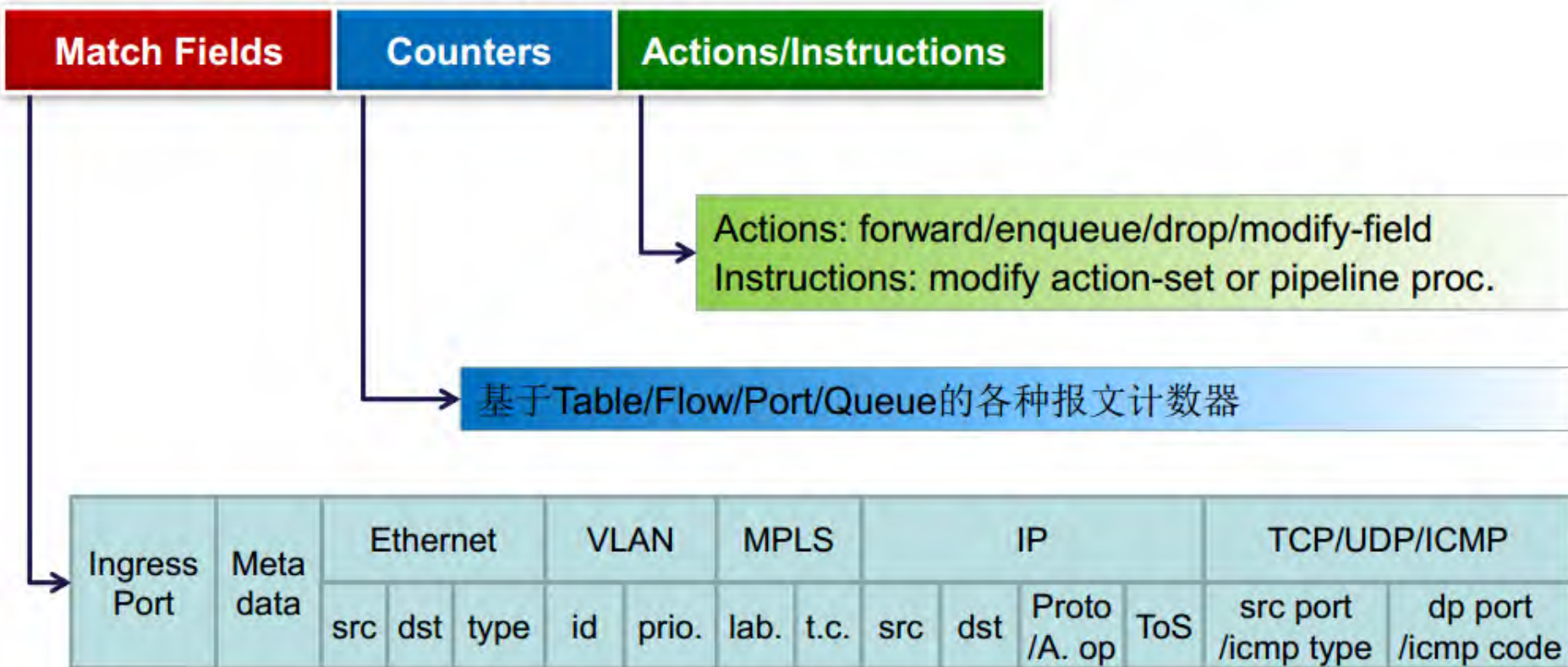
- ✓应用和网络的无缝集成，应用告诉网络如何运行才能更好地满足应用的需求，比如业务的带宽、时延需求，计费对路由的影响等
- ✓用户可以自行开发网络新功能，加快新功能面世周期
- ✓理论上NOS和转发硬件的开放标准接口使得硬件完全PC化

### 网络虚拟化

- ✓逻辑网络和物理网络的分离，逻辑网络可以根据业务需要配置、迁移，不受物理位置的限制
- ✓多租户支持，每个租户可以自行定义带宽需求和私有编址



## Flow Table的组成

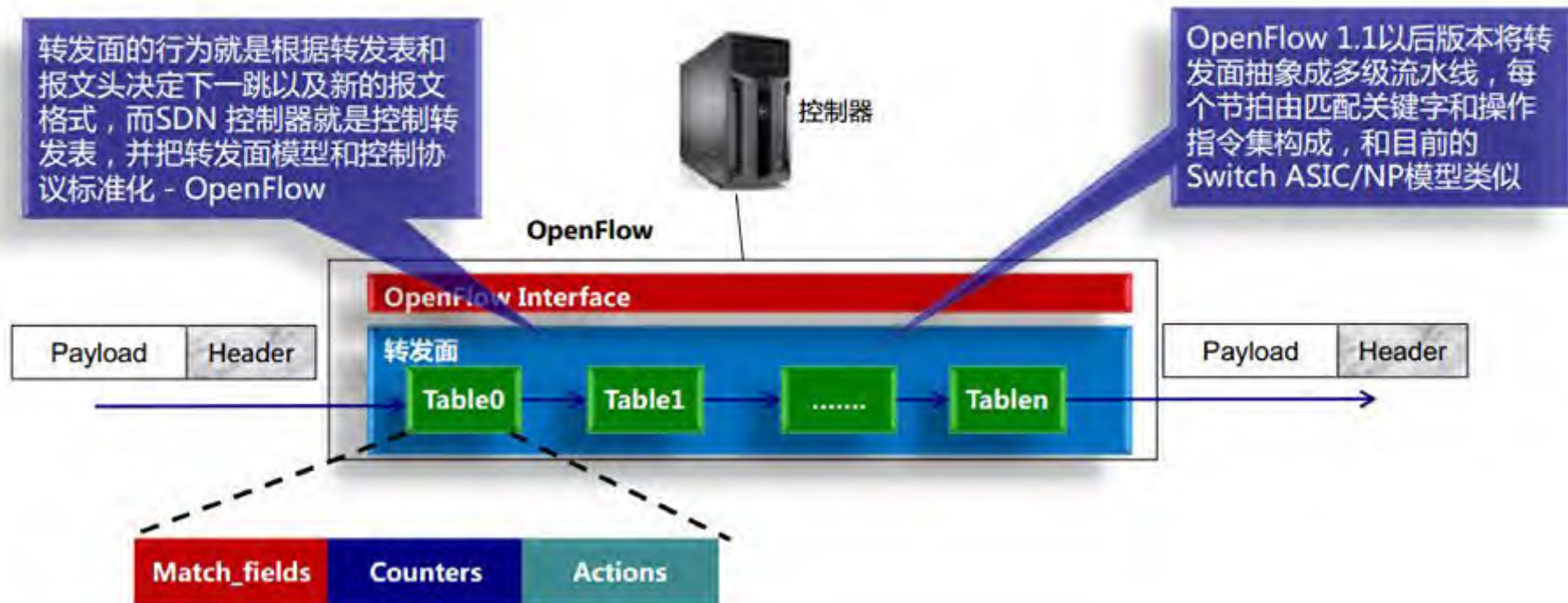


**Actions/Instructions**决定了OpenFlow对转发面行为的抽象能力，例如：

- 修改报文头部各个字段值、封装/去封装
- 将TTL值在内/外层头部之间进行复制
- 输出到一个端口或一组端口，实现组播、多路径转发、负载均衡等功能

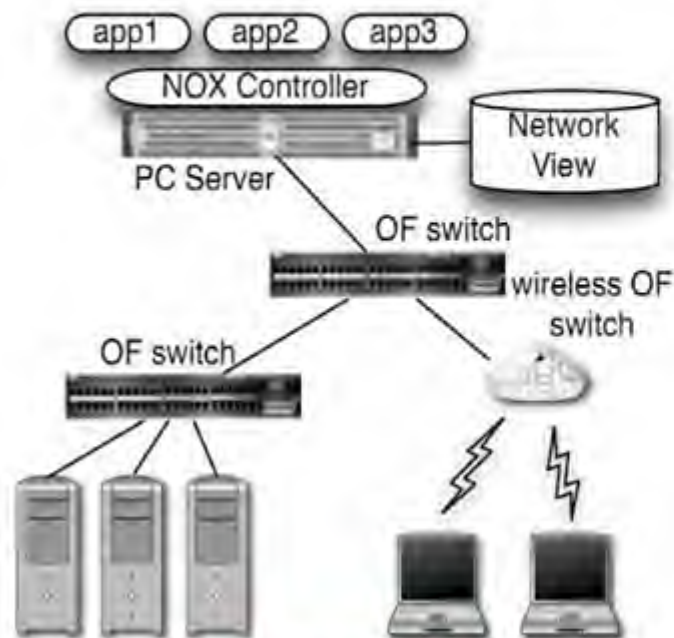


SDN转发层的关键技术是对转发面进行抽象建模。针对SDN转发面抽象模型，ONF标准组织提出并标准化了OpenFlow协议，在该协议中转发面设备被抽象为一个由多级流表(Flow Table)驱动的转发模型:



- SDN控制层的关键是SDN控制器，也可以称为网络操作系统（NOS）或网络控制器
- 网络的所有智能、核心均在SDN控制器中，由SDN控制器对转发面进行转发策略的调度和管理，通过无智能的快速转发面设备，支持运行在SDN控制器之上的不同业务
- 目前已公开的NOS源码和架构有：NOX、FloodLight、Onix等

■ NOX架构：是针对软件定义网络架构下网络控制器的一种开源实现，支持C++/Python，主要由Nicira公司支持。



NOX架构图

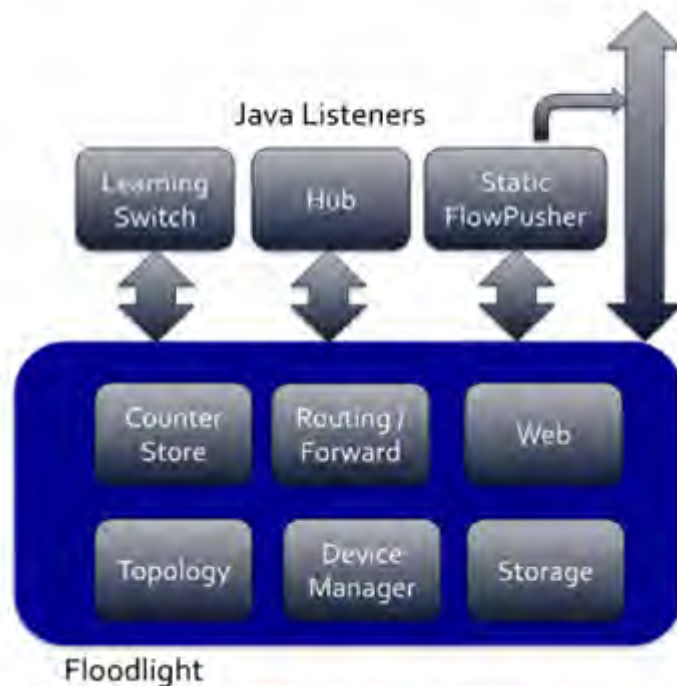
➢ **NOX controller:** 作为NOX的核心模块，为上层APP提供平台服务，例如应用加载以及维护、消息服务、事件注册以及回调机制、网络拓扑发现等。

➢ **Network View:** 在SDN架构下，网络的拓扑以及相应的链路状态信息统一保存在网络控制器中，也即由网络控制器进行拓扑发现、链路状态检测。

➢ **APP:** 基于controller以及network view，APP提供传统交换机/路由器控制器所提供的功能，包括基本功能：维护转发流表(二层)、路由表(三层)；高级功能：QoS/VLAN设置、ACL；网管功能：SNMP agent, Web Service等。

- SDN控制层的关键是SDN控制器，也可以称为网络操作系统（NOS）或网络控制器
- 网络的所有智能、核心均在SDN控制器中，由SDN控制器对转发面进行转发策略的调度和管理，通过无智能的快速转发面设备，支持运行在SDN控制器之上的不同业务
- 目前已公开的NOS源码和架构有：NOX、FloodLight、Onix等

■ Floodlight架构：由BigSwitch公司赞助支持的一个开源OpenFlow控制器实现，支持Java，其中蓝色部分为网络控制器所涵盖的功能。



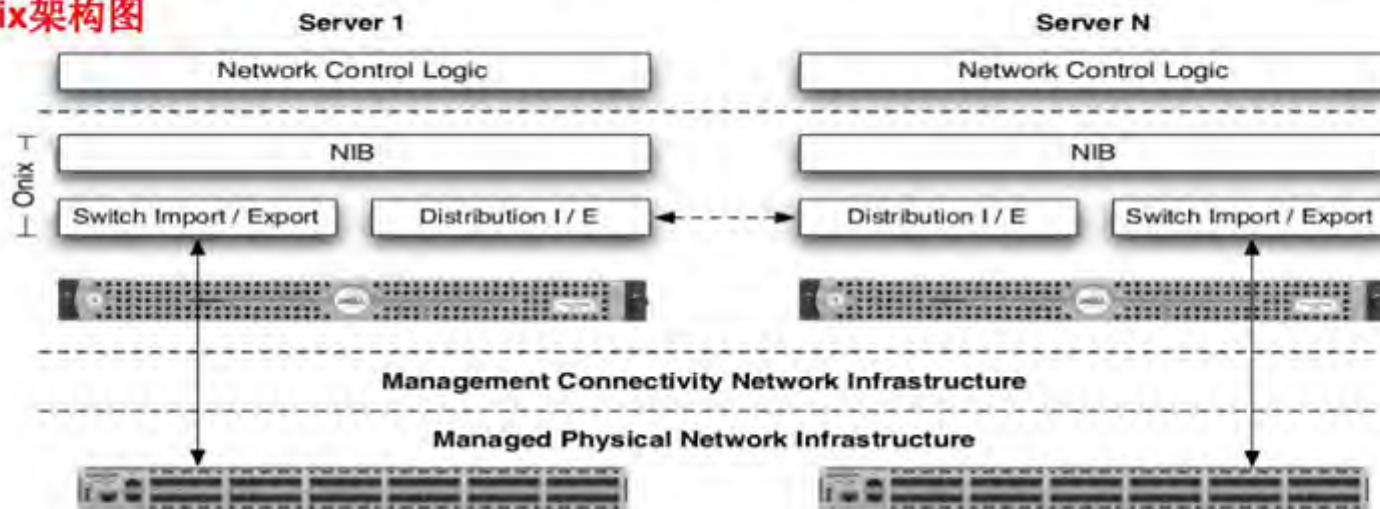
**Floodlight架构图**

- **拓扑管理**：维护整网的拓扑数据，并对外提供访问拓扑数据的服务。
- **设备管理**：维护并跟踪网络中设备的信息，例如OpenFlow交换机(端口、ID等)，主机/虚拟机(例如IP地址、MAC地址等)。
- **存储**：用于控制器核心模块对相关数据进行保存。
- **Counter Store**：负责流表管理、状态管理。
- **路由/转发**：计算路径以及下发流表的核心模块。
- **Web**：Floodlight内支持HTTP Restful接口的核心模块。



- SDN控制层的关键是SDN控制器，也可以称为网络操作系统（NOS）或网络控制器
- 网络的所有智能、核心均在SDN控制器中，由SDN控制器对转发面进行转发策略的调度和管理，通过无智能的快速转发面设备，支持运行在SDN控制器之上的不同业务
- 目前已公开的NOS源码和架构有：NOX、FloodLight、Onix等
- Onix架构：由Nicira, Google等公司提出的一个NOS架构，可以被视为NOX的增强版，主要在可靠性、性能、安全、扩展性等方面做了大量的优化。
  - NIB(Network Information Base)为Onix的核心，存储网络拓扑等网络相关数据。NIB通过Distribution I/E实现了数据的分布化。所有的网络业务逻辑(Network Control Logic)基于分布化的NIB数据，进行相关流程的控制和决策。

## Onix架构图





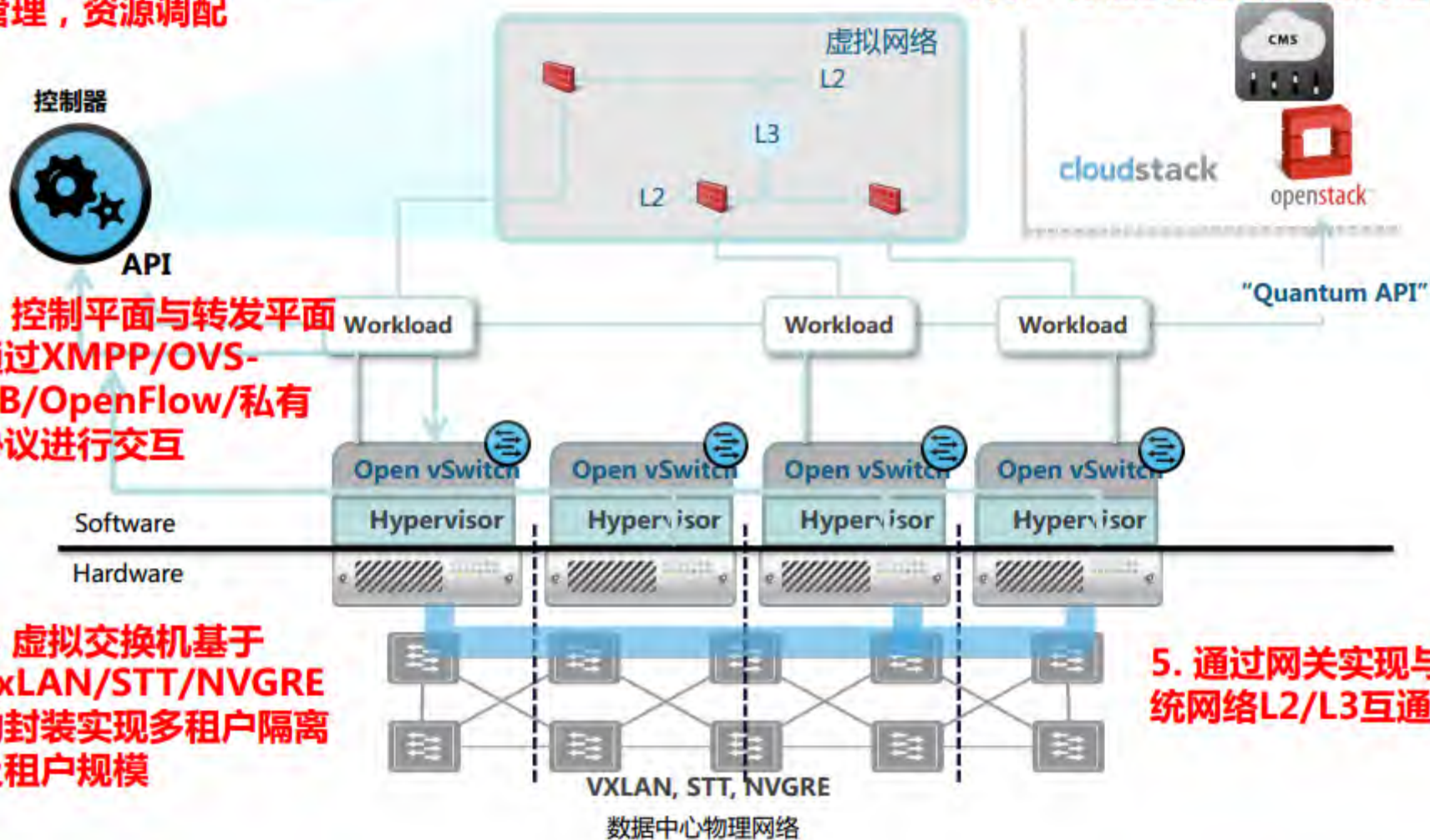
3. 控制器实现统一网络管理，资源调配

2. 控制平面与转发平面通过XMPP/OVS-DB/OpenFlow/私有协议进行交互

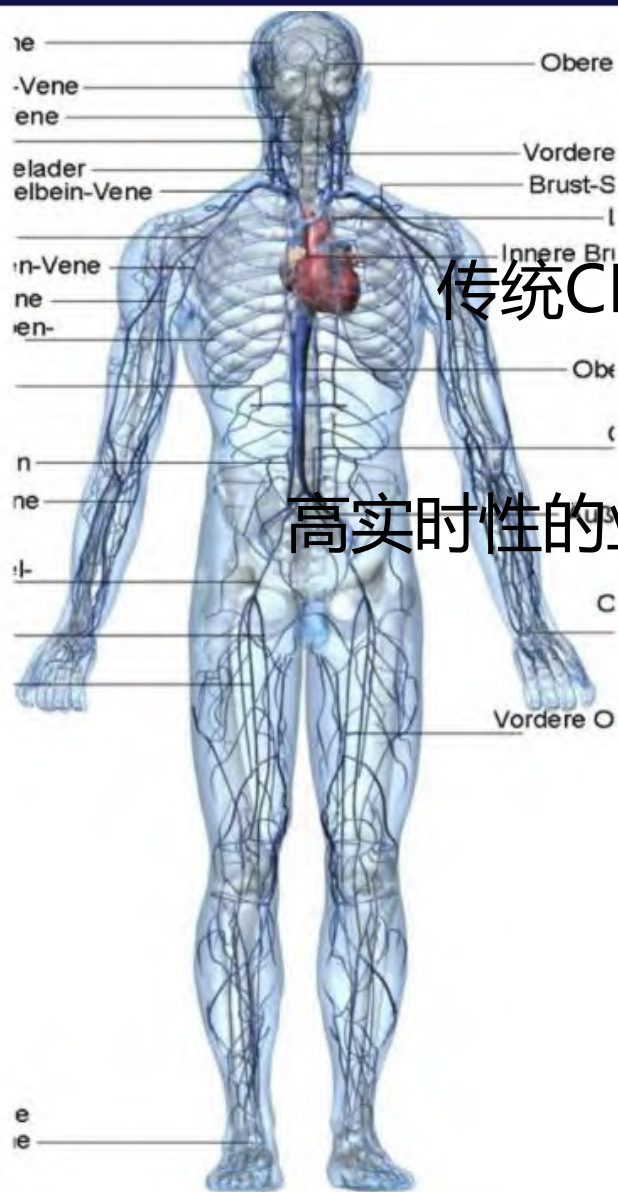
1. 虚拟交换机基于VxLAN/STT/NVGRE的封装实现多租户隔离及租户规模

4. 与云平台对接，实现灵活的网络+计算+存储的端到端云服务提供

5. 通过网关实现与传统网络L2/L3互通



# 神经网络



传统CDN 解决静态问题 无监督的学习

超强的自愈能力  
高实时性的业务交付如何更好的解决？

- 对损伤有冗余性
- 处理信息的效率极高
- 并行处理



CDN  
GSLB

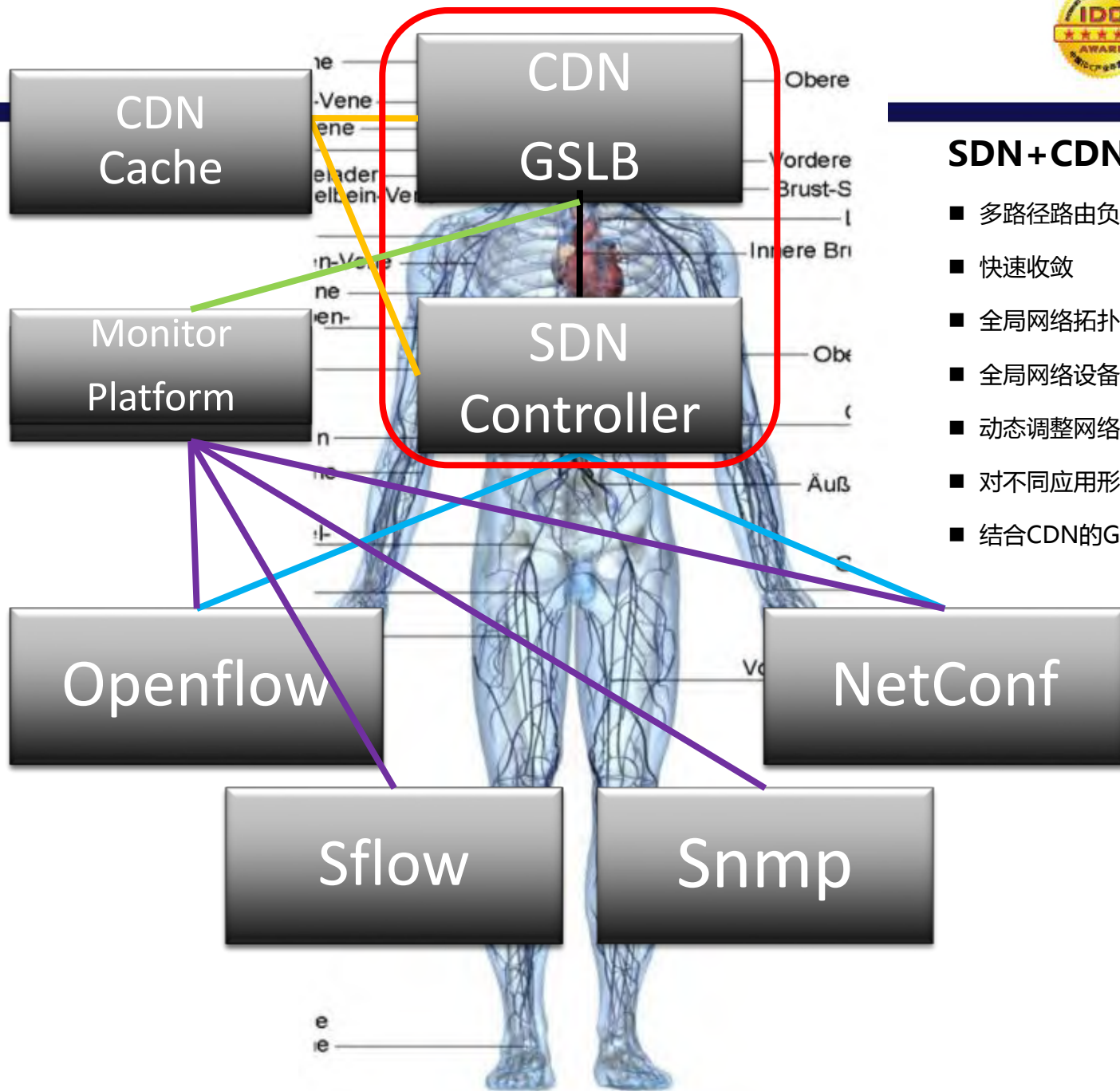
CDN  
Cache

SDN  
Controller

Openflow

Sflow

NetConf



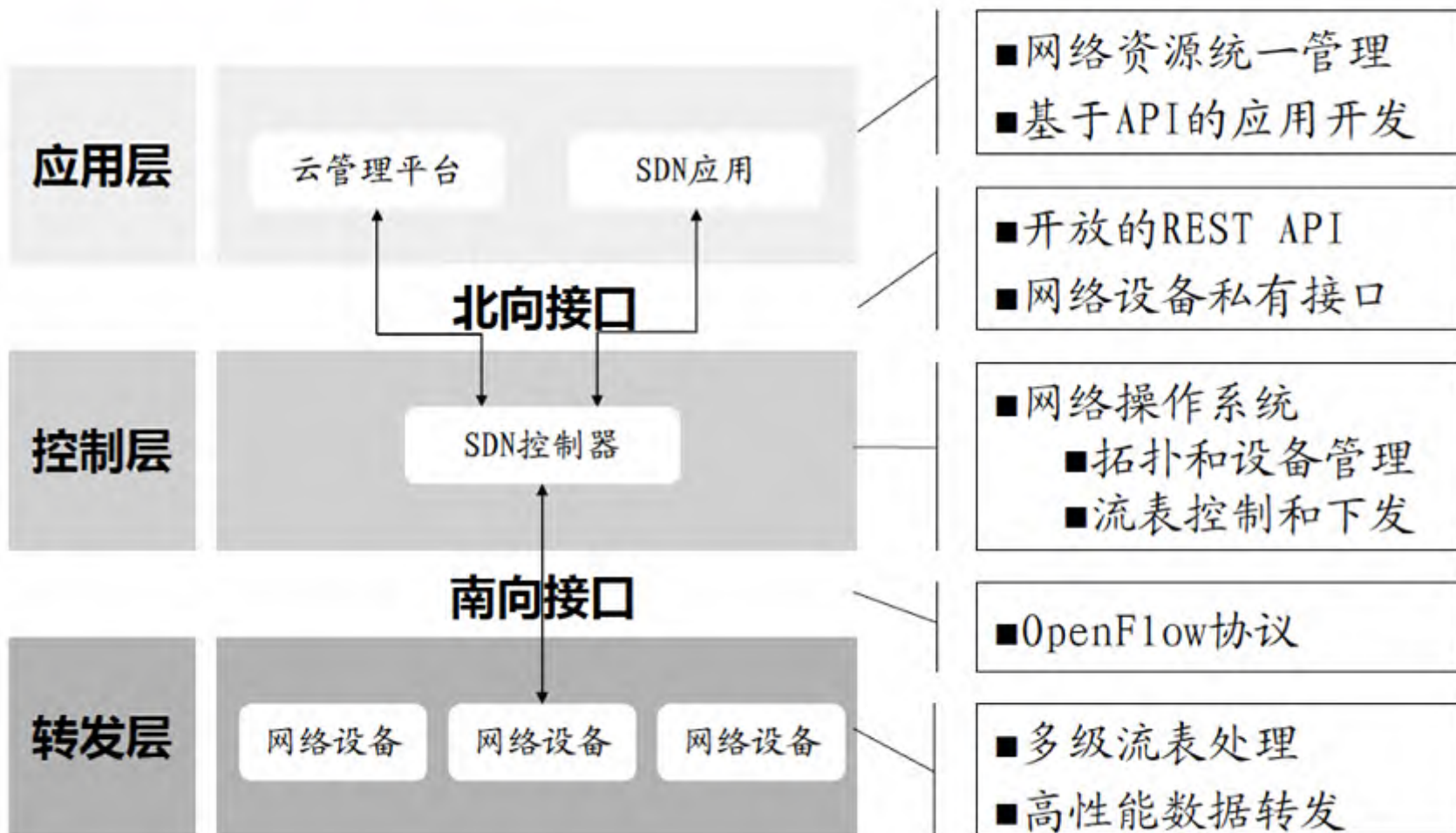
## SDN+CDN

- 多路径路由负载分担
- 快速收敛
- 全局网络拓扑
- 全局网络设备自动下发配置
- 动态调整网络
- 对不同应用形成各自不同的网络逻辑层
- 结合CDN的GSLB进行全局网络调度

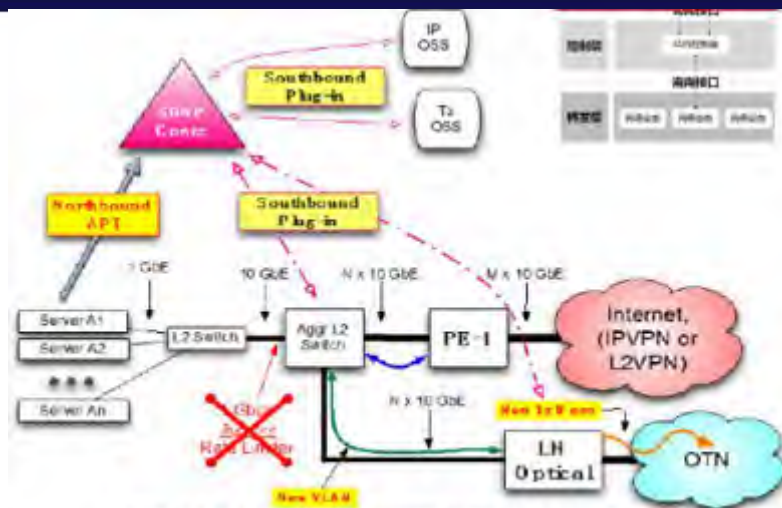




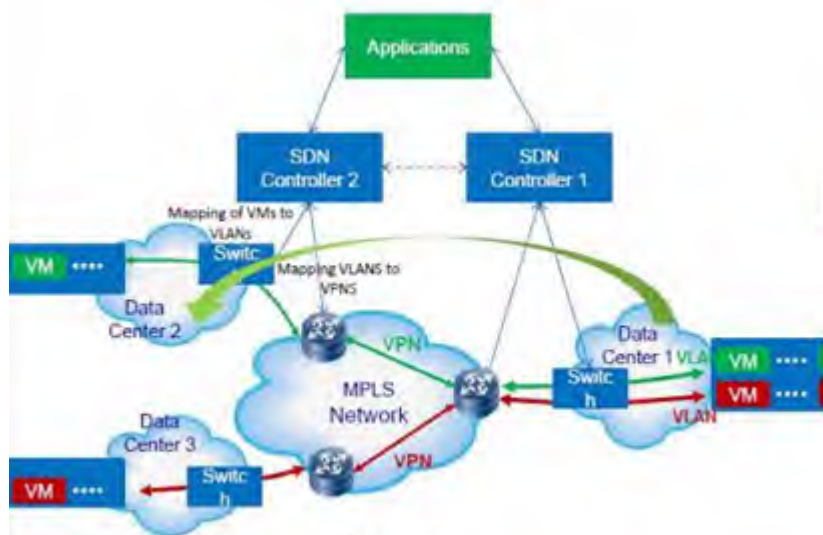
## SDN关键技术体系



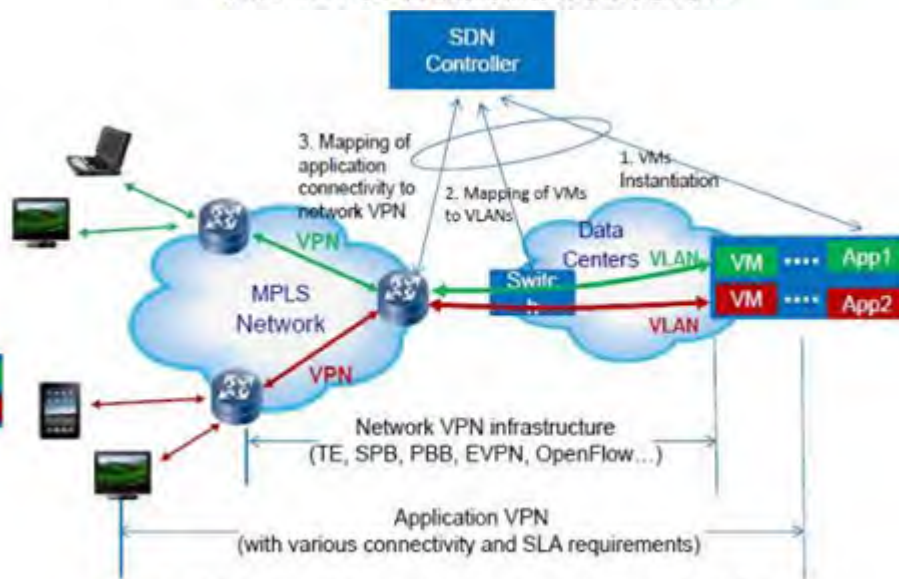
- 利用开放的北向API，调用网络能力
- 编程加速网络创新业务实现
- 制定灵活的配置和管控策略
- 定义高级别的安全性和可靠性



通过SDN控制器实现按需带宽调整



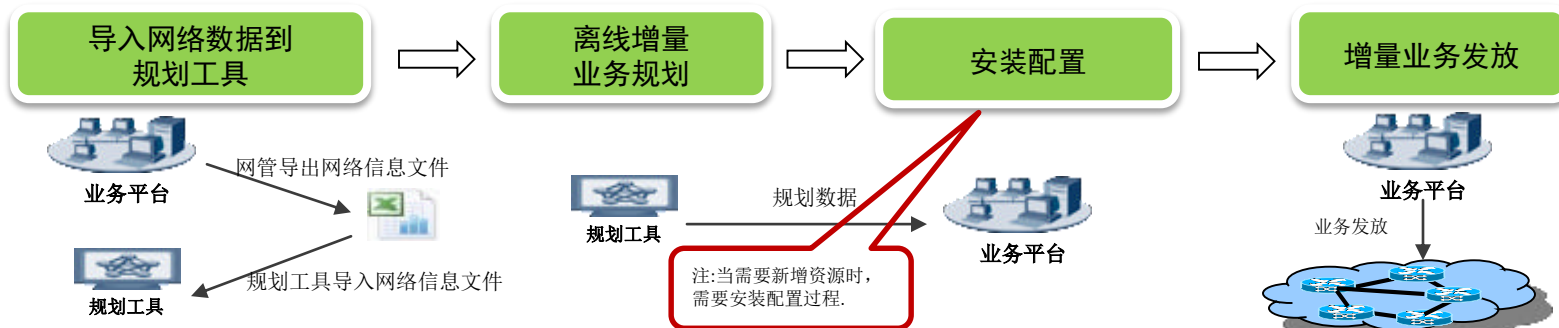
通过SDN控制器协助虚拟机迁移



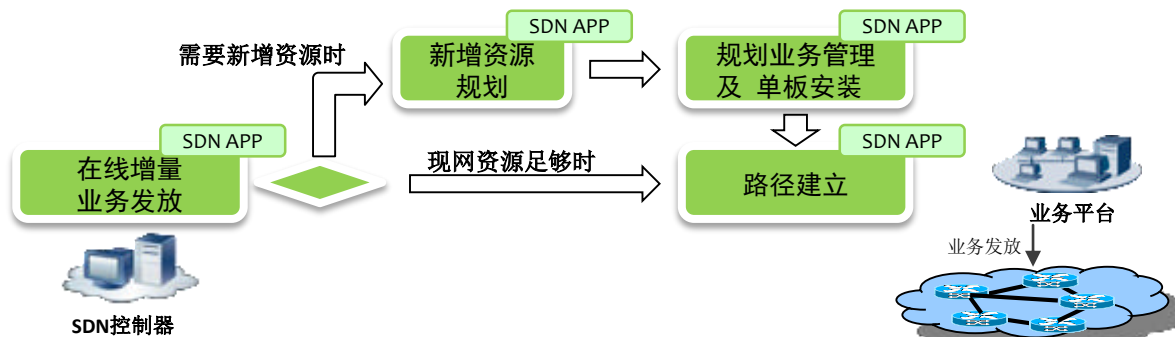
通过SDN控制器实现应用到VPN的映射

## 快速业务部署

引入SDN控制器  
之前  
(小时级)

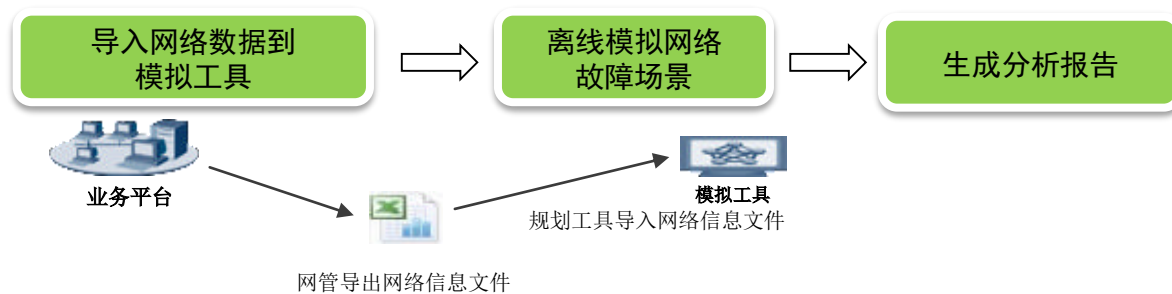
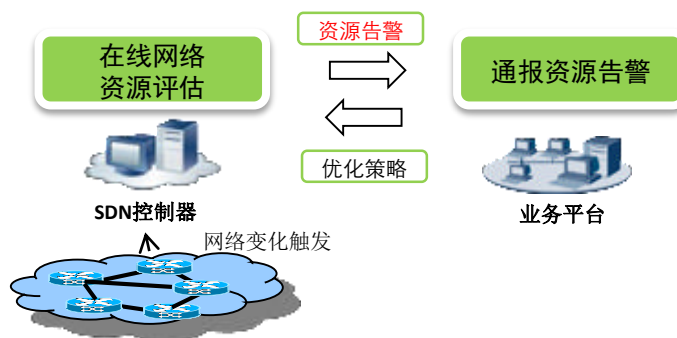


引入SDN控制器  
之后  
(分钟级)



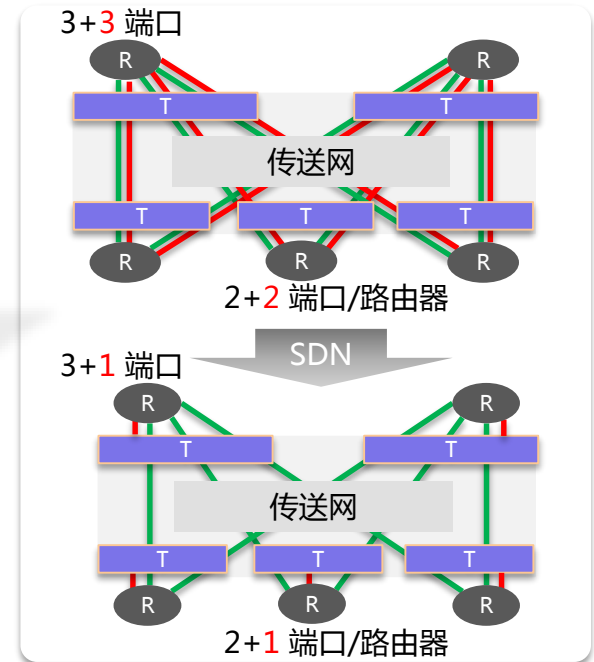
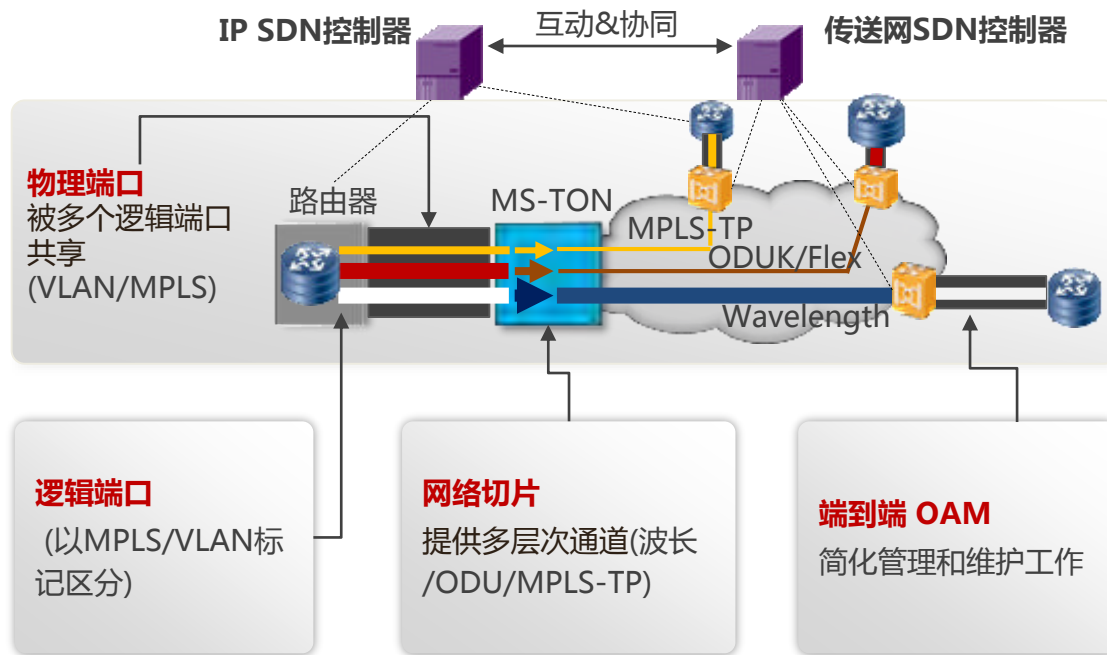
- 引入SDN后，增量发放新业务可以通过SDN控制器在线统一控制管理，操作简单、快速。
- SDN能够自动为业务计算路径，并判断现网资源是否足够，包括恢复资源是否足够。确保新增业务对网络生存性无影响。

## 网络状态实时感知

引入SDN控制器  
之前  
小时级引入SDN控制器  
之后  
分钟级

- 引入SDN后，可以直接在线对现网资源进行评估，避免了繁琐的数据导入导出操作，也解决了离线评估不能及时适应现网变化的问题。
- 自动根据网络变化评估现网资源是否满足业务生存性要求，如果不能，则主动上报告警提示进行修复或扩容操作，主动运维防患于未然。

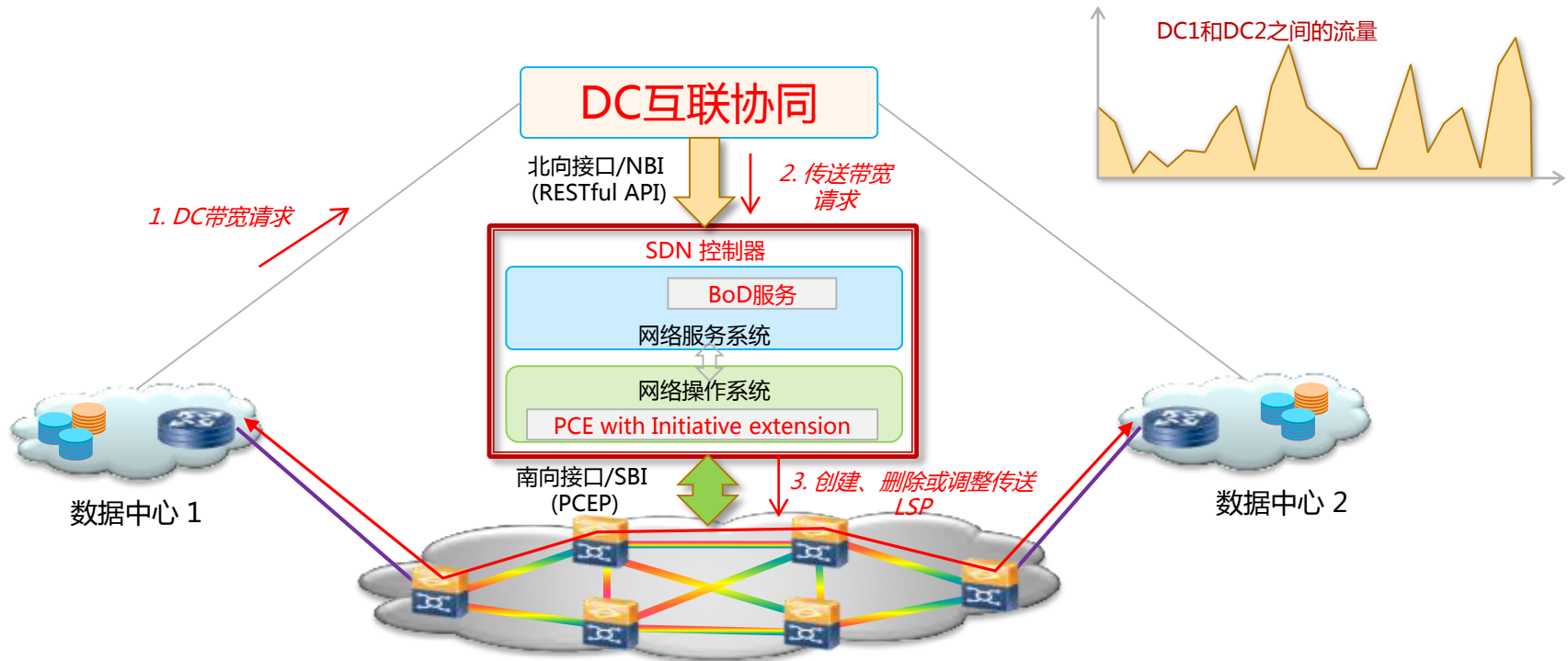
IP与光跨层协同，提升网络效率和资源利用率



**IP + 光多层协同：**

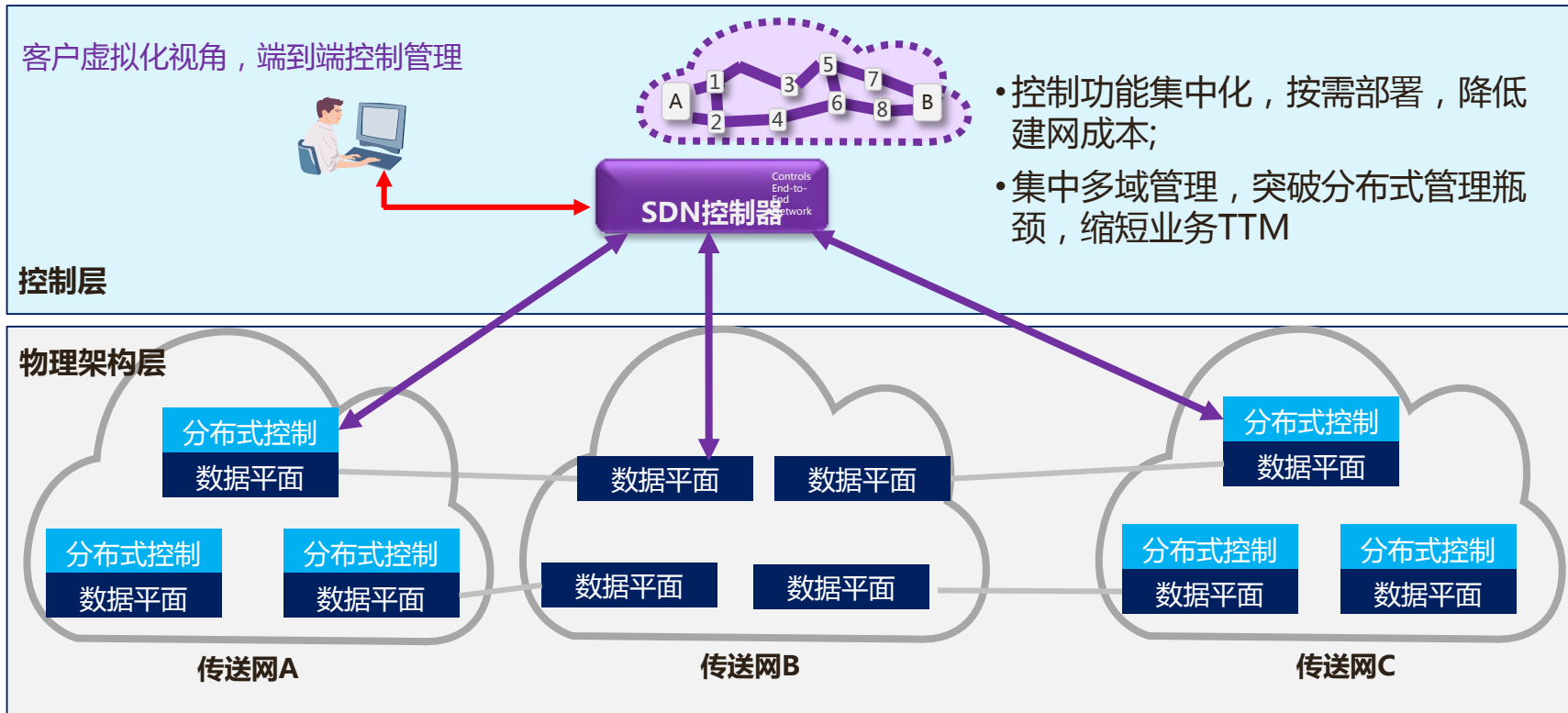
1. 通过多层网络优化提升网络效率；
2. 通过协同保护提升网络可靠性和保护效率；
3. 通过协同运维提升运维效率。

## 基于SDN调度为数据中心协同提供优化



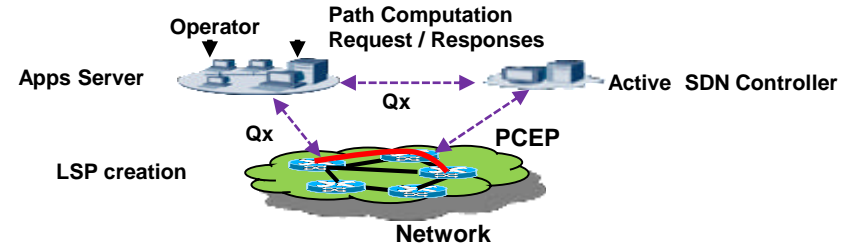
- 通过开放接口的传送带宽资源池方案，传送带宽与数据中心流量自动联动调度，提升资源利用率，降低CapEX。
- 通过自动BoD，提供自动实时带宽服务，避免人工干预，降低OpEX。

## 多区域大网统一调度



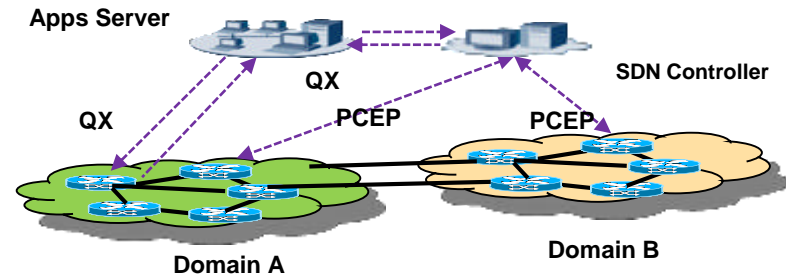
## 业务自动发放和修复

- 在线增量业务部署
- 业务集中重路由计算



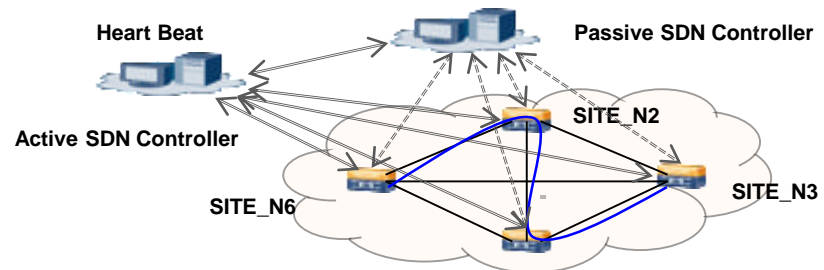
## 网络自动运维

- 大网、多域（子场景：端到端业务建立、删除、域内/域间故障重路由）
- 生存性分析（在线链路等各种故障模拟，提前对业务的生存性进行评估分析）

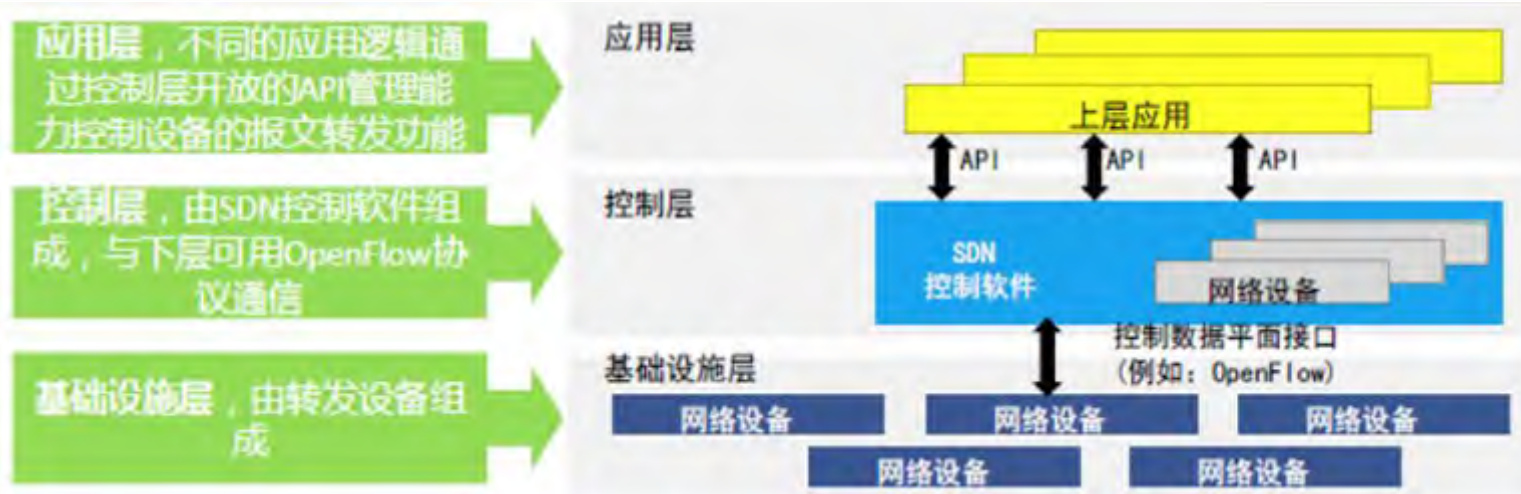


## SDN 架构

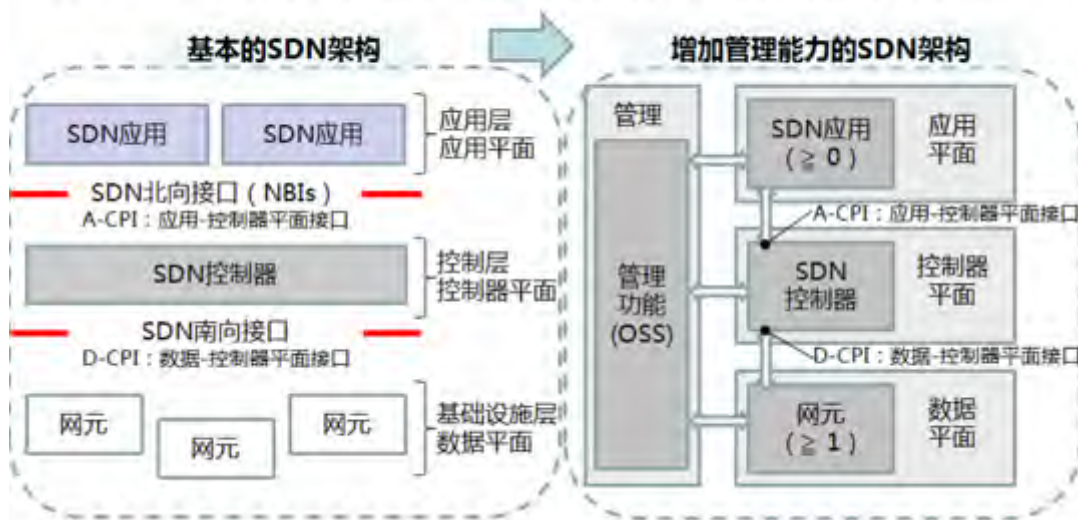
- 控制器 1+1、N+M
- 网络资源自动发现
- 业务集中路径计算
- Stateful PCE







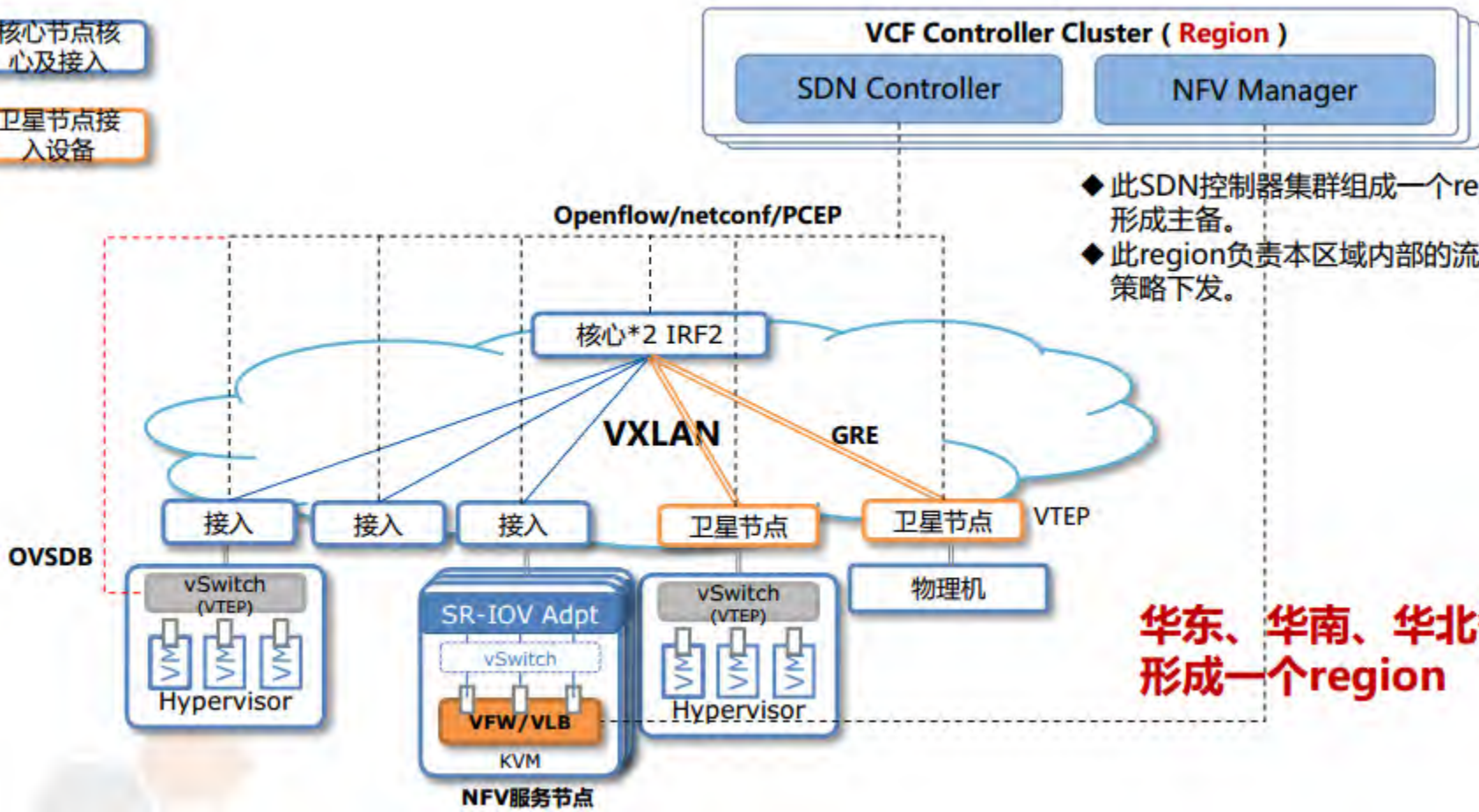
- SDN = 控制与转发分离 + 逻辑集中控制 + 网络资源能力状态开放





核心节点核心及接入

卫星节点接入设备

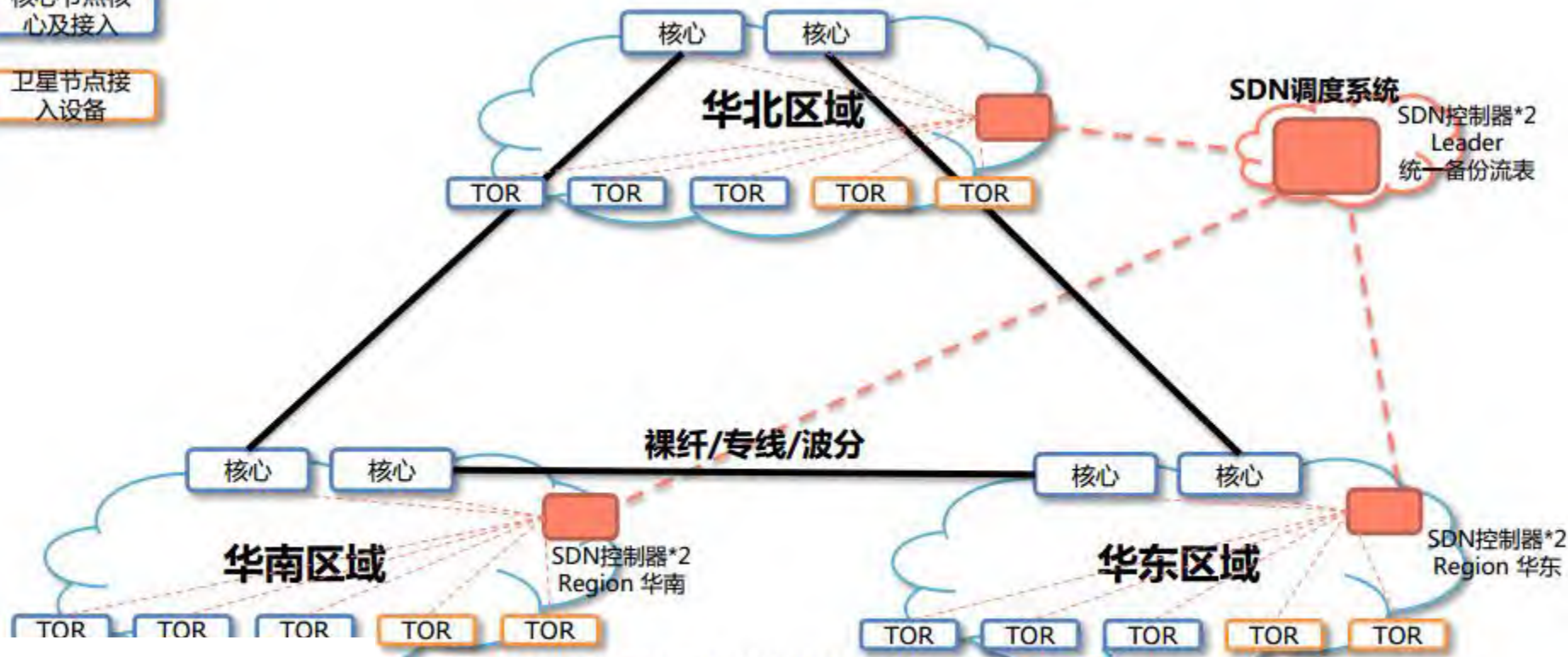


- ◆ 此SDN控制器集群组成一个region，形成主备。
- ◆ 此region负责本区域内部的流表及策略下发。

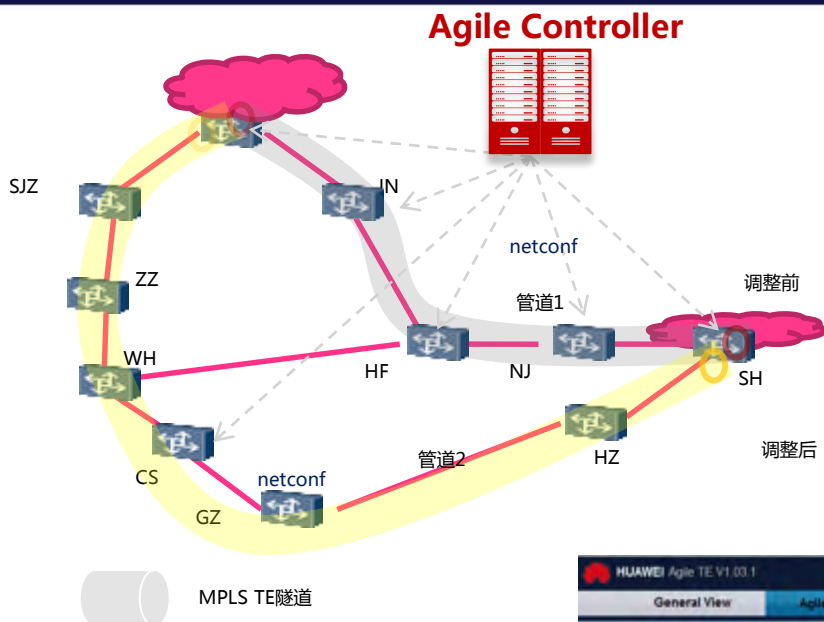
华东、华南、华北各形成一个region



- 核心节点核心及接入
- 卫星节点接入设备



- ◆ 软/硬件网络设备packet-in至控制器的流量可选带内或带外
- ◆ Packet-in流量可跨三层
- ◆ 本地控制器内未匹配流表/策略，上查Leader控制器
- ◆ Leader可统一管理所有控制器；Leader控制器提供北向接口



## 业务智能分配

- 高性能AC控制器，管控全局隧道（MPLS TE）拓扑
- 智能计算，为新业务选择最优隧道路径

## 业务智能调优

- 实时监控业务带宽和时延，动态调整业务隧道
- 实时故障监控，智能隧道切换

## Agile TE实现业务智能分配

- **隧道自由定制，简便快捷**  
1分钟即可在任意两地之间创建主备 TUNNEL隧道。
- **快速业务发放**  
新业务开通运行只需几分钟



## 业务智能调优，均衡流量，提升资源利用率

1) 发现当前管道已处于重载状态，比较拥堵



2) 在控制器GUI界面上，选择某条业务流，通过调整流管匹配策略，调整它的路径；



3) 调整完成，该条业务流跑到另一条轻载管道上



### ■ 带宽拥塞时，优化业务路径

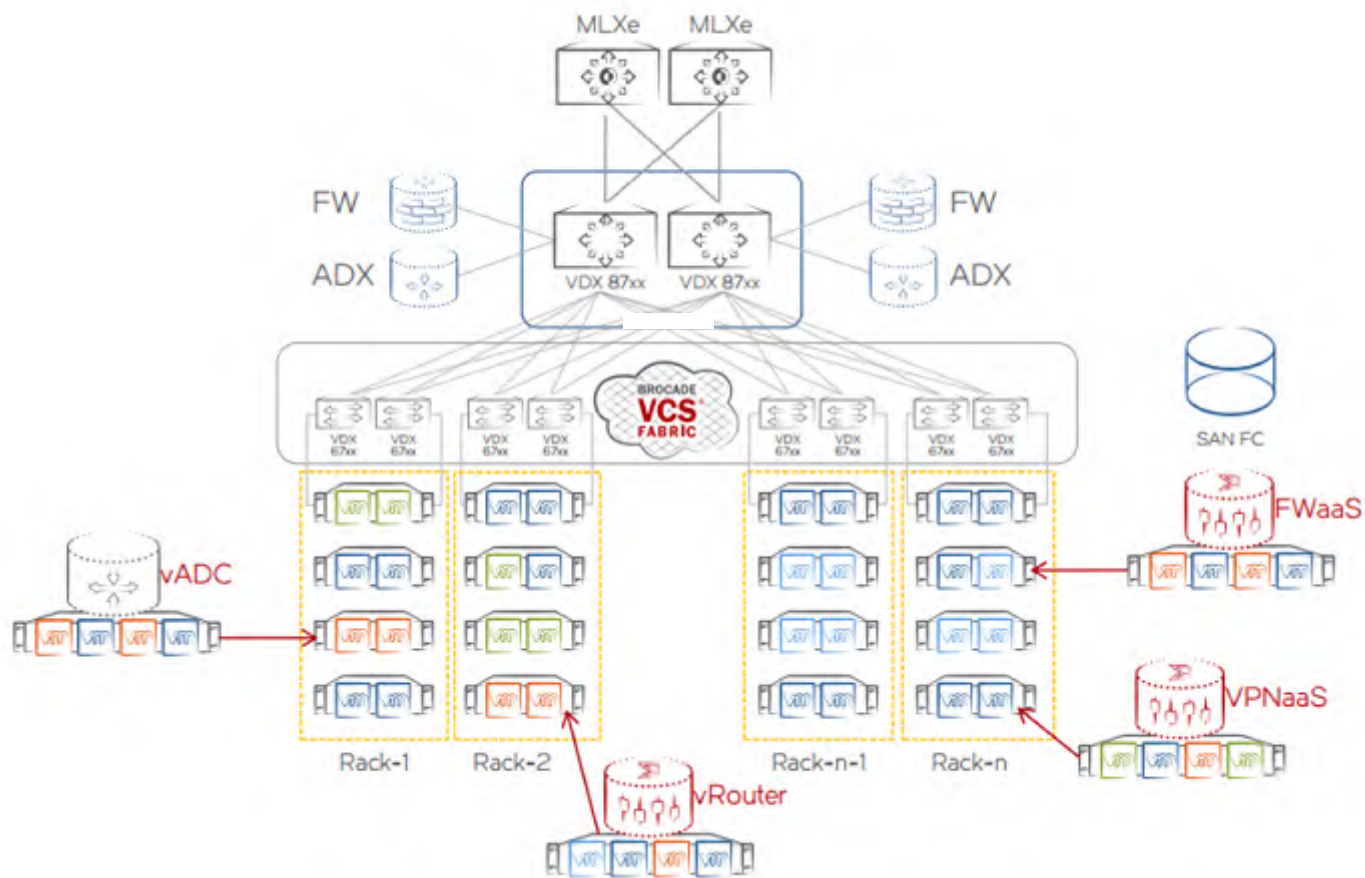
- (一) **保证高优先级的业务流量带宽**，当管道拥塞时把高优先级的业务流调整到其他非拥塞的/优质的管道上；
- (二) **日常均衡流量**，把中、低优先级的业务流量调整到局部轻载管道上，提升轻载管道的带宽利用率。

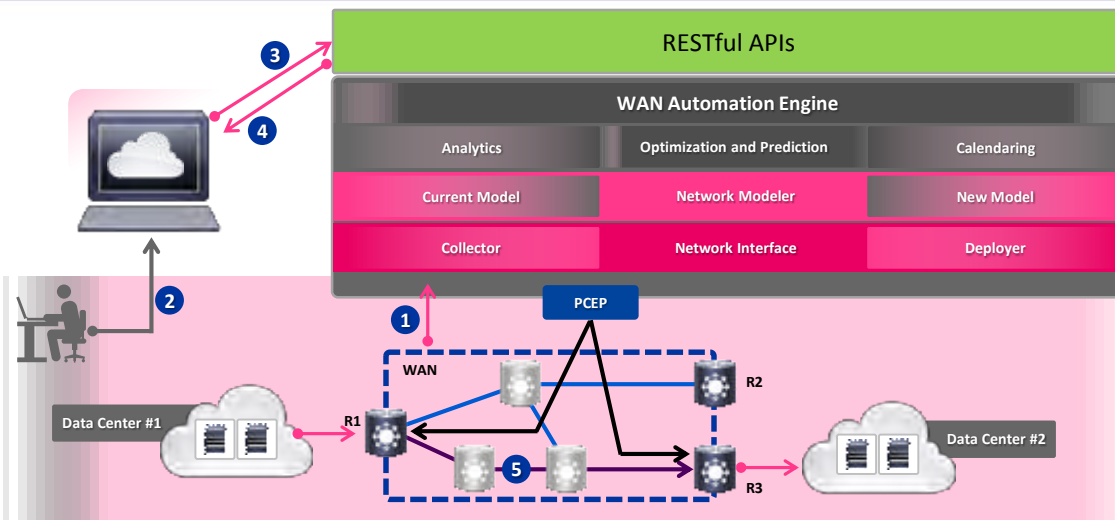


# IDCC



博科 SDN 控制器





- 1 Network conditions reported to collector consistently
- 2 Customer requests DC #1 – DC #2 bandwidth at a future date
- 3 Demand admission request: <R1-R3, B/W, future date>
- 4 WAE returns booking confirmation as the future date nears
- 5 On the future date, NS-OS places customer demand on IGP or explicit path (TE tunnel)

**Problem**

XX时间有在线实时业务，需要网络SLA保障及响应

**Solution**

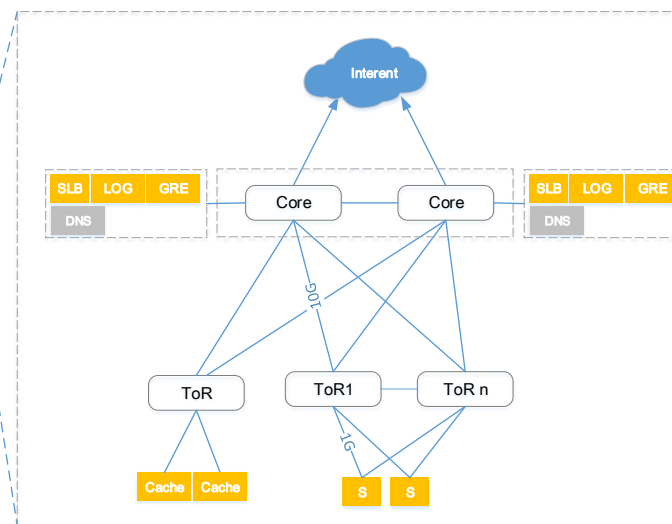
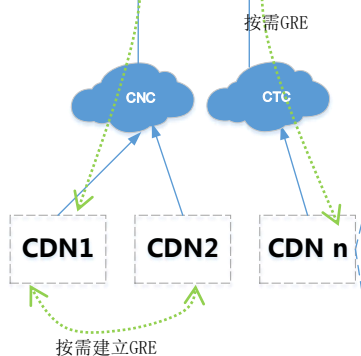
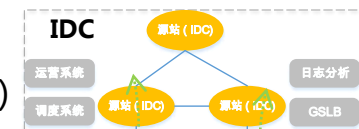
WAE评估未来事件网络情况，按需自动交付网络需求





## 设计建议：

- 借助裸纤/波分/DCI实现“音频、视频、游戏、CDN”业务
- 业务可以通过MPLS或VXLAN实例隔离
- L3层SW；
- 支持硬件GRE隧道、MPLS/L2 VPN
- 可靠性：冗余电源、风扇
- 小于2:1收敛比（出口带宽决定）



分发服务系统

负载均衡系统

运营管理系统



## 系统实现复杂

- 综合考虑用户位置、服务器负载、链路质量、带宽成本等多元素，造成中央调度系统重载，算法复杂
- 整体调度控制层面过于扁平，分层不够清晰

## 动态内容回源延时不可控

- 针对服务节点直接回源，中间路径可能遇到跨运营商、高峰拥塞、运营商网络故障等情况
- 回源时延不可控，无法均衡用户体验

## 运营商网络黑洞不可控

- 节点之间及区域与节点之间的流量调度（横向及上行），中间路径只能寄托运营商网络的质量—不可控的“黑洞”

## 网络设备能力没有充分利用

- 软件+服务器方式的流量调度，导致网络设备的一些功能性及灵活性的能力被浪费—传统设备

基于SDN建立一个独立的**网络控制面**

基于网络控制层面，实现业务加速的**自动化调度**



## 互联网企业案例-Google



- Google采用OpenFlow技术，通过10G网络链接分布全球的12个数据中心；
- 方案目标：提高网络的可用性和容错能力，通过周密的流量工程和优先次序工作，将链路使用率从平均的30%-40%提升至接近100%；
- 故障处理：通过没有严格传输时间限制的弹性流量来保护高优先级流量，通过全球范围的网络拓扑和动态地改变通信特征（使用非最短路径转发来路由绕过故障链路）；

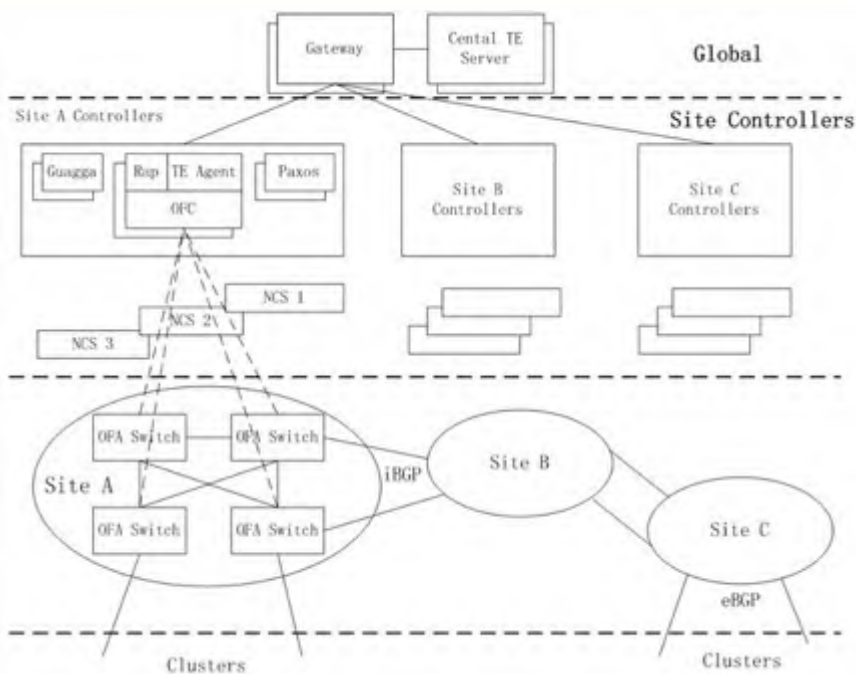




## 运营商案例-NTT Communication

- NTT全球数据中心实现OpenFlow/SDN部署和互连，并基于此提供虚拟私有云服务，实现了网/云集成部署、拓扑灵活自主设计等功能，解决了VLAN数限制问题
- 自主研发了Virtual Network Controller
  - 用于多个数据中心的统一服务和按需配置，已在其欧洲、美国和日本的数据中心进行了虚拟数据中心的部署
- 推出多项基于基于SDN的创新网络服务，改善用户体验
  - 基于SDN的网关互联自动化：通过门户调用API实现VPN

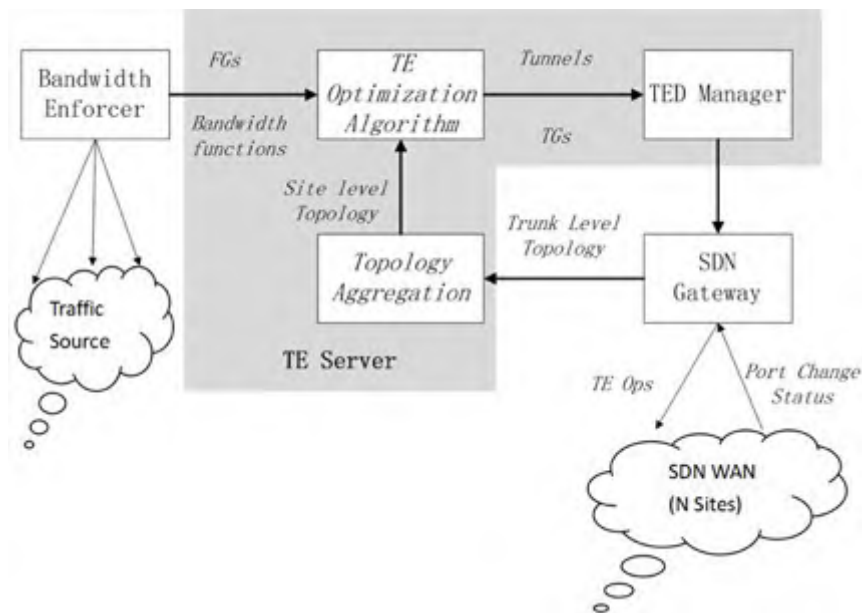




1. 用户数据备份，包括视频、图片、语音和文字等
2. 远程跨数据中心存储访问，例如计算资源和存储资源分布在不同的数据中心
3. 大规模的数据同步（为了分布式访问，负载分担）

TE和BGP都可以为一条流生成转发路径，但TE生成的路径放在ACL表，BGP生成的放在路由表（LPM），进来的报文如果匹配到ACL表项，会优先使用ACL，匹配不到才会用路由表的结果。一台交换机既要处理从内部发到别的数据中心的数据，又要处理从别的数据中心发到本地数据中心内部的数据

对于前者，需要使用ACL Flow表来进行匹配查找，将报文封装在Tunnel里面转发去，转发路径是TE指定的，是最优路径。而对于后者，则是解封之后直接根据LPM路由表转发。还有路过的报文（从一个数据中心经过本数据中心到另外一个数据中心），这种报文也是通过路由表转发





平台承载：容纳1000万用户，100+IDC节点，1T以上带宽容量

平台能力：可加速用户上传不同类型视频、图片、数据文件，可处理各种数据流，并实现动态加速、https加速

网络情况：通过SDN、动态路由等技术优化核心专网，利用网络技术提高平台安全系数



IDCC



不懂武功的两伙人打架靠什么赢？

会武功的两伙人打架靠什么赢？





第十届中国IDC产业年度大典  
The 10<sup>th</sup> Internet Data Center Conference

**Thank you!**

