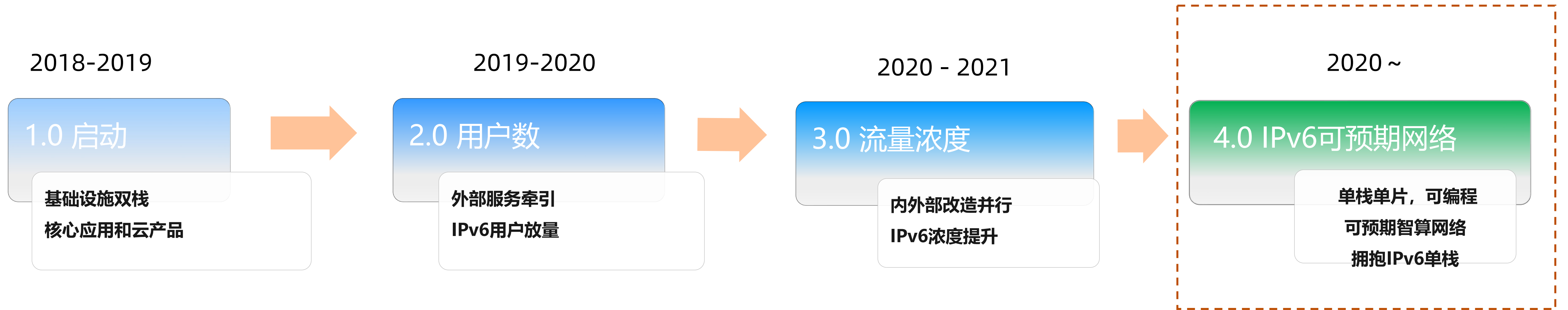


IPv6 创新赋能：可预期智算网络

蔡德忠，阿里云

IPv6下一代互联网技术创新与国际标准研讨会

阿里巴巴IPv6规模部署和创新应用路径



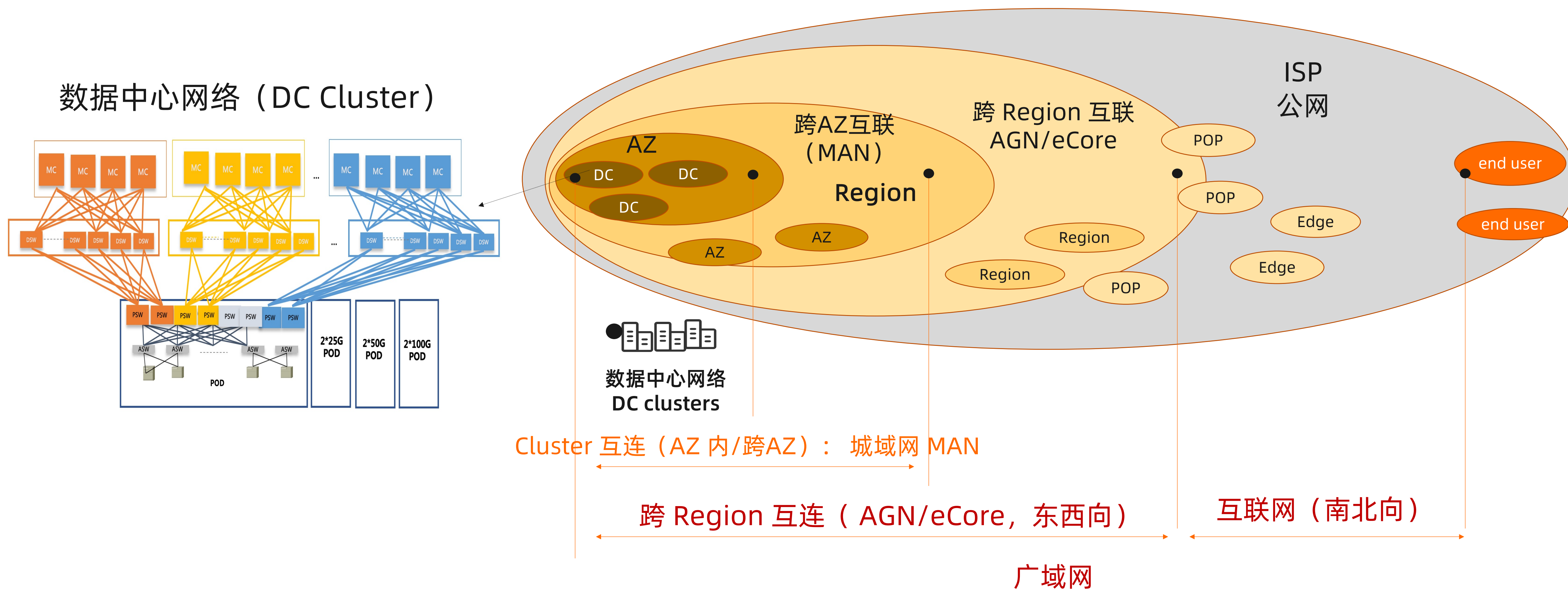
IPv6规模部署阶段

生态解耦，异步演进，保障商业连续

IPv6创新阶段

释放IPv6红利，支撑可预期智算网络

阿里云基础设施网络大图

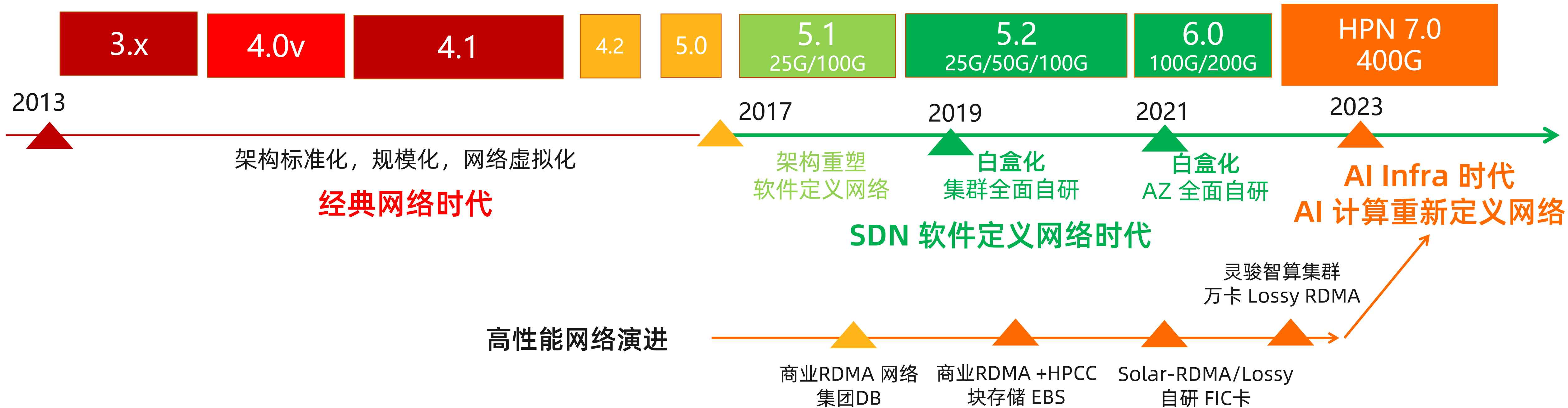


阿里云数据中心网络架构演进历程



数据中心网络架构演进

阿里云 HAIL 数据中心网络架构



2016

Google, AWS, FB, Azure 在 2013-2016 年间完成了数据中心网络的“软件定义和白盒化”架构升级, 阿里落后大概 3-5年 我们花3年时间赶上了 Big4, 在 2019 年到了同一个起跑线

网络架构演进：AI 计算重新定义数据中心

网络性能进入可预期时代：From Best-effort to the Predictable Network

第一个10年

经典网络

信息化，传统 IT → 互联互通



网络以协议为中心
设备黑盒，人肉运维
IETF 定义互联标准

第二个10年

SDN 软件定义网络

互联网应用 → 超大规模 & 弹性



互联网应用重新定义网络系统

软件定义网络 (SDN)
设备白盒解耦，运维自动化，
SONiC 为代表的网络开源生态

Now: AI Infrastructure

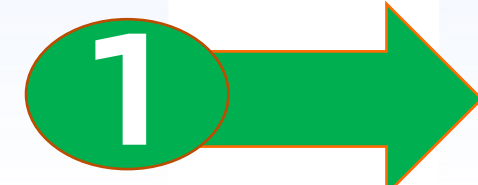
AI 计算定义网络

AI 大模型 → 大算力 → 可预期性能



AI 计算重新定义数据中心架构

高性能智算网络
端网融合，计算/存储和网络协同
生态：开放解耦与闭环并存



架构重塑（绿皮火车到高铁时代）
以协议/设备为中心 → 以软件为中心

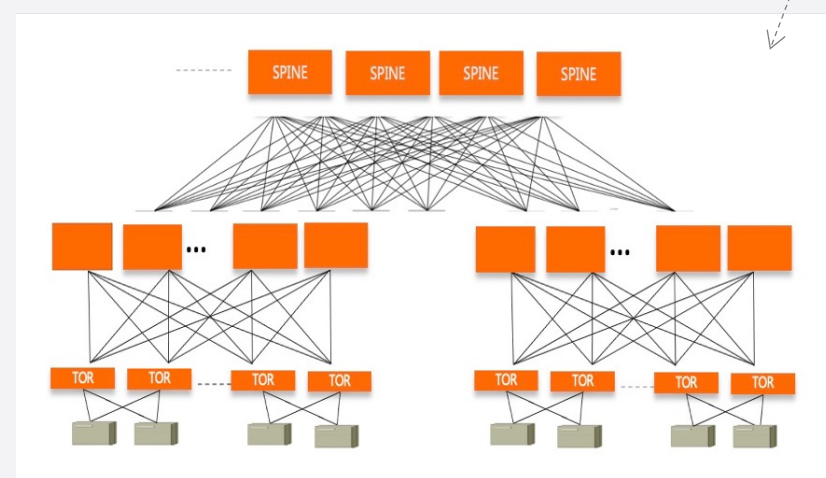


端网融合架构（计算和网络协同设计）
AI计算重新定义网络

网络研发 技术体系大图

2

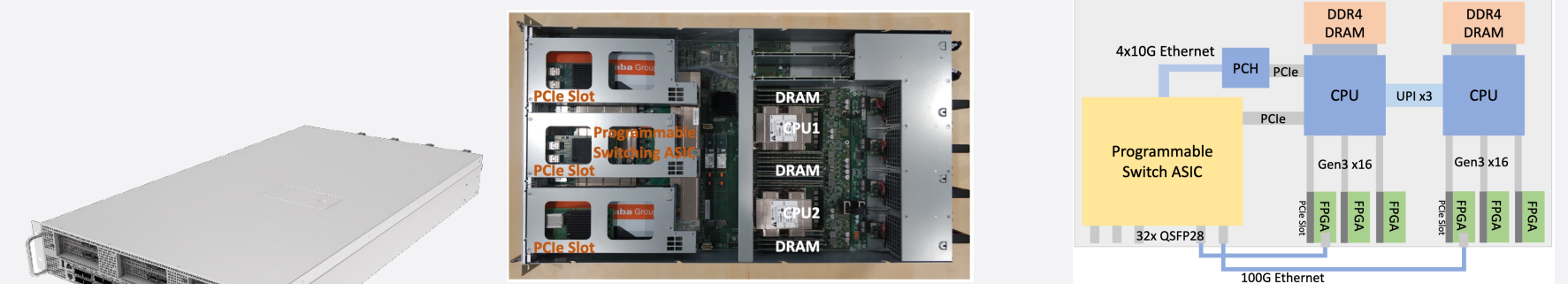
端网融合技术体系：端和网协同设计，应用和网络协同设计



超大规模数据中心网络



超高性能超大规模智算集群



边缘云超融合网络平台

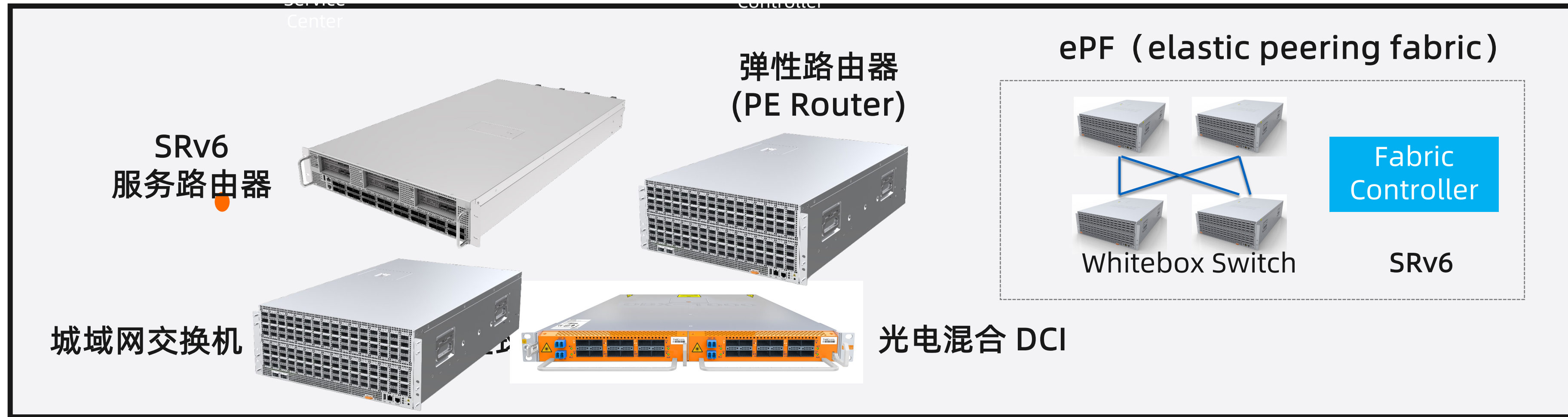
1



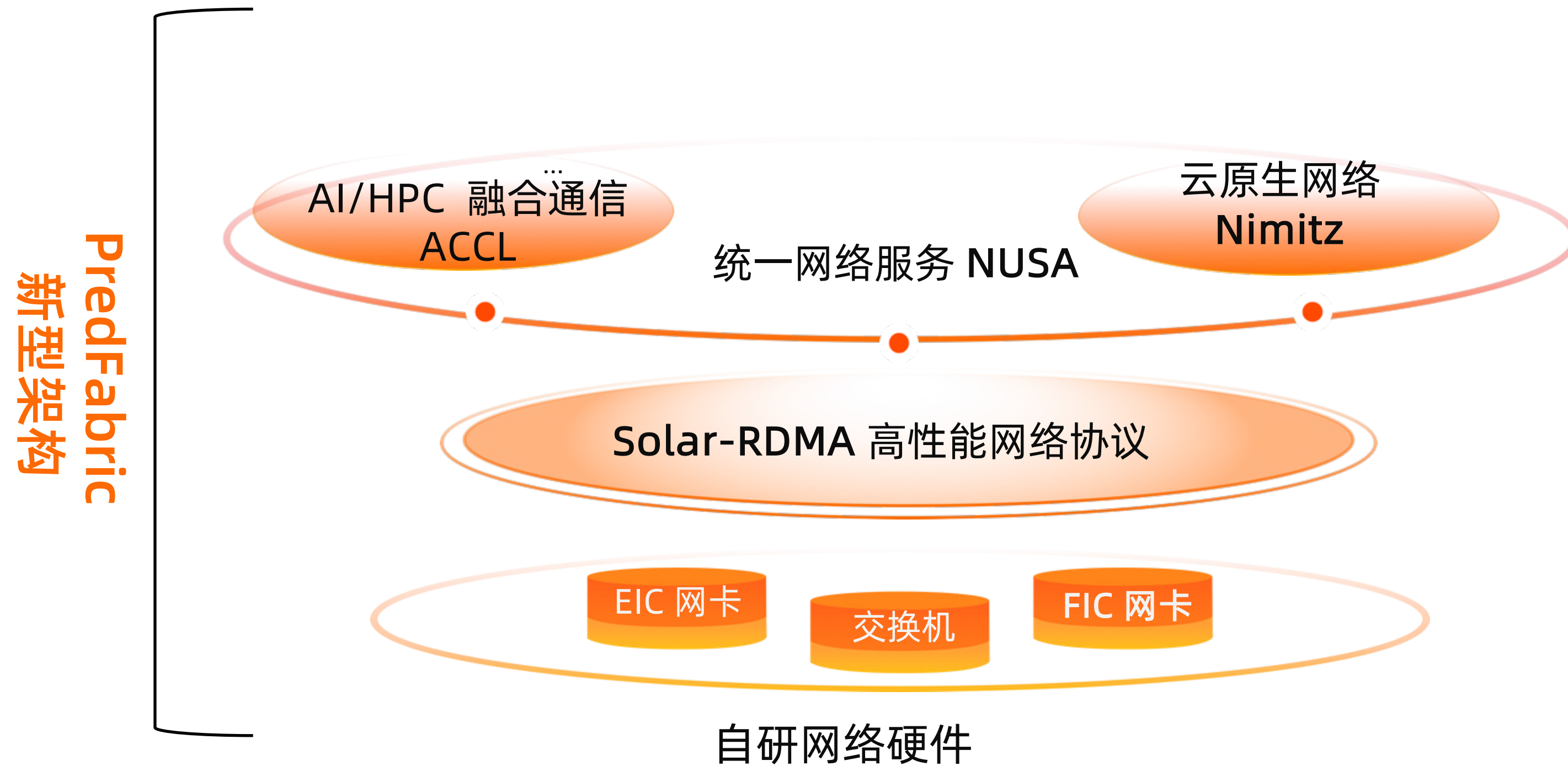
白盒化技术体系：单栈+单片+可编程
IPv6/SRv6 + Single-chip white box + Programmability

白盒化技术体系 - 网络全域可控

数据中心交换机（现网部署 10万+），边缘超融合网关， Metro/WAN/Peering 路由器，光电混合 DCI设备



端网融合技术体系 - 可预期网络



400G 高性能智算网卡



ACCL通信库 [HPCA'20, MICRO'20]

作业和网络 相互感知和利用, 拓扑感知调度最优资源匹配

Solar-RDMA, vFabric, [2xSIGCOMM'22]

超大规模、多路径、自适应路由、ReLSA低时延原子操作

HPCC流控 [SIGCOMM'19]

多路径下智能化网络流控, pfc-free, 不同模型flexibility

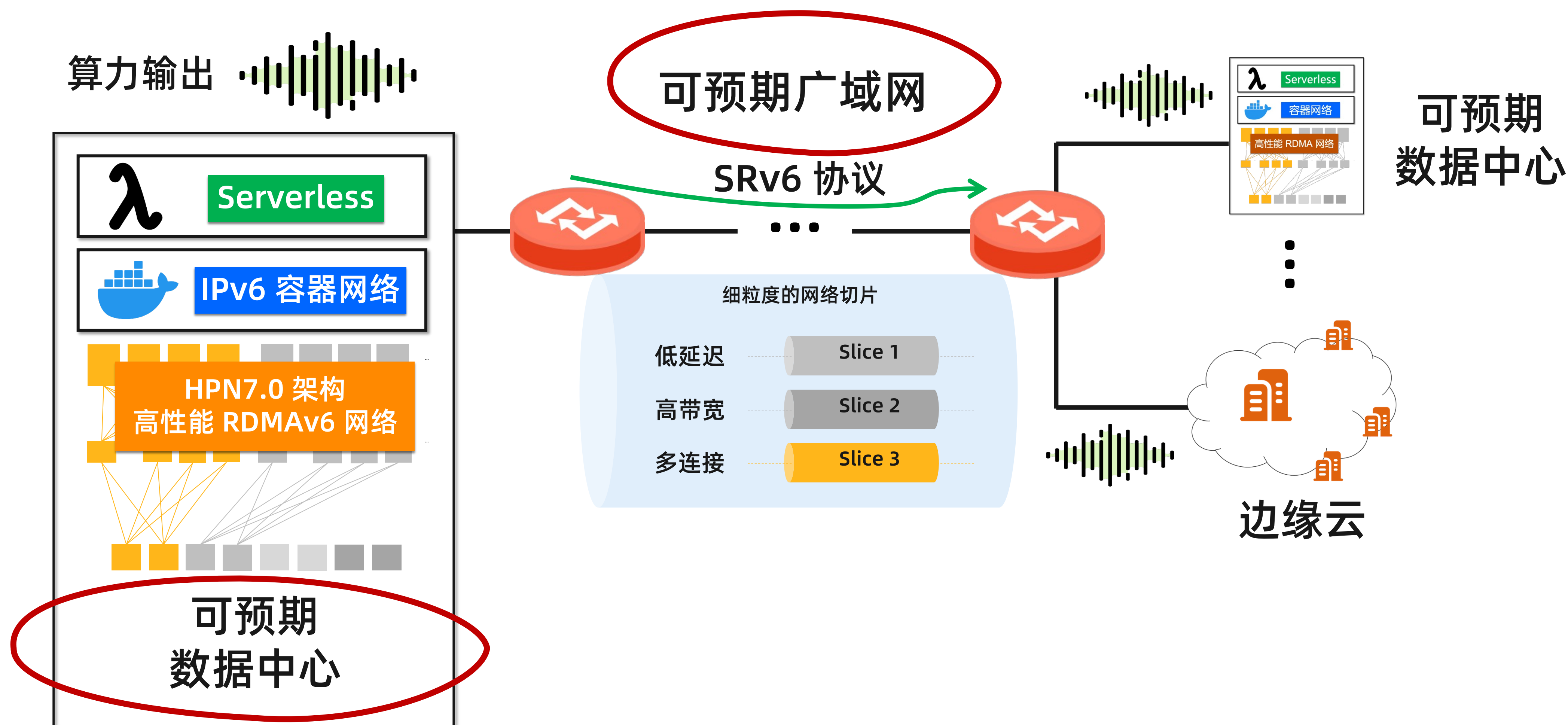
可编程交换网 [SIGCOMM'20, '21]

in-network拥塞、路径精确反馈, 自适应路由调节

端网融合架构创新: 10 篇顶会论文 SIGCOMM/NSDI/etc, 30多专利

规模部署: 2017年开始高性能存储, 2019 年开始高性能AI集群

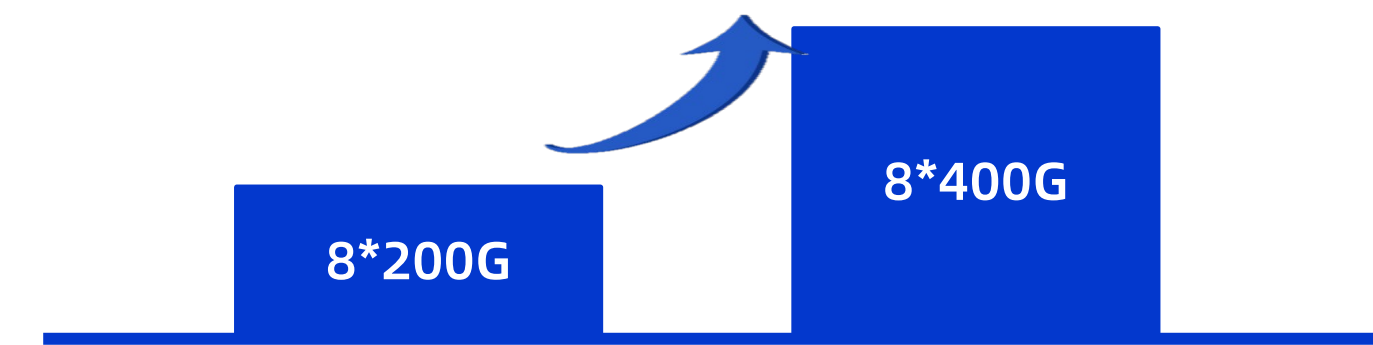
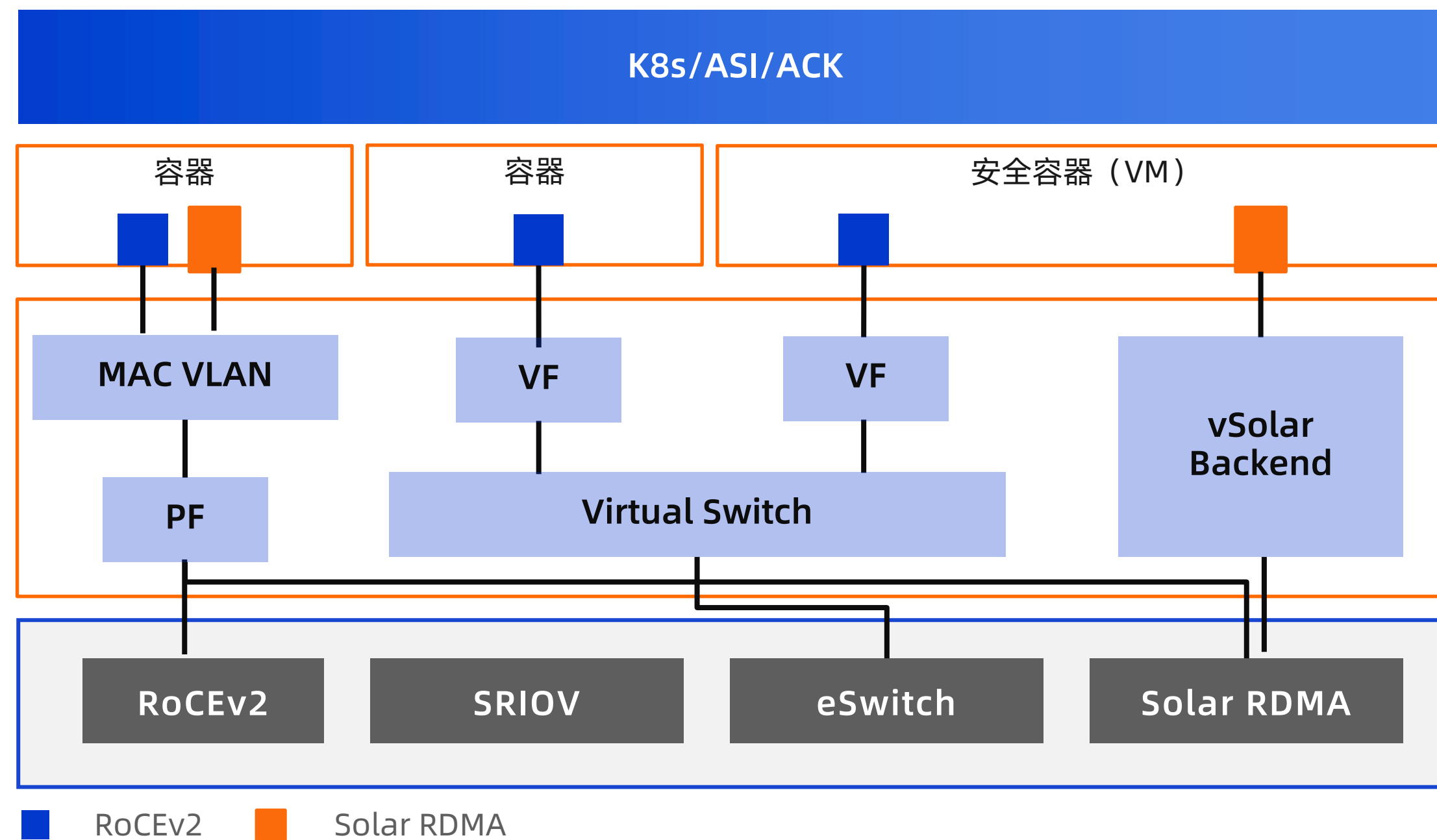
IPv6 端网融合创新：构建新型智算网络



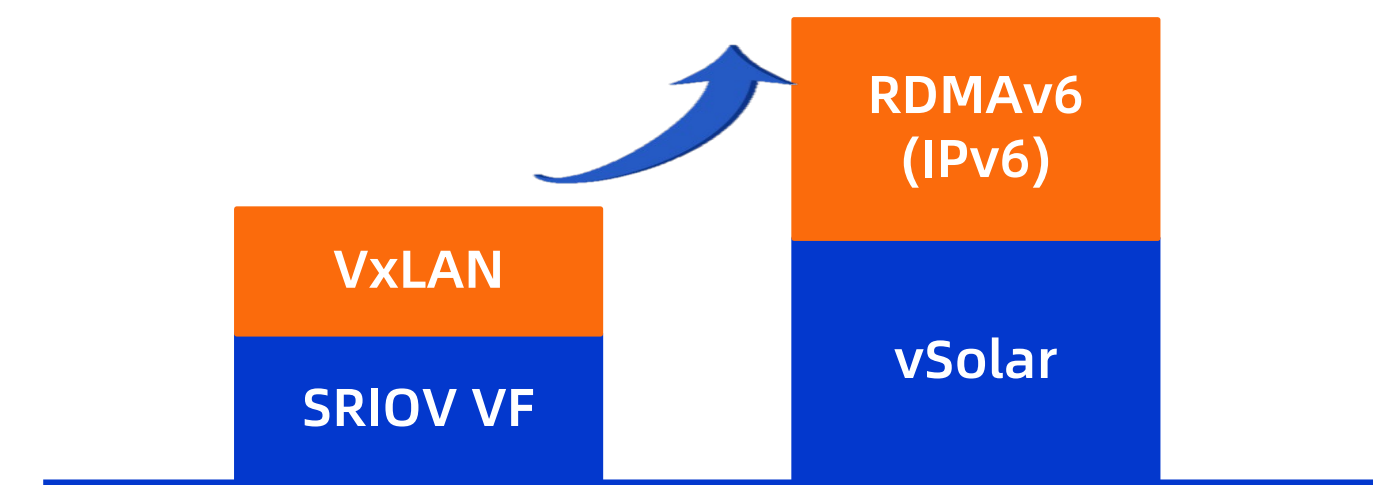
在端到端IPv6可达性的基础上，实现端网融合的IPv6可预期网络
支撑下一代面向 AI 的新型 Infrastructure - 从资源售卖到算力输出

HPN7.0 架构: IPv6高性能网络 RDMAv6

基于IPv6的, 容器化的 **云原生智算集群**网络, 同时获得高性能、敏捷和高弹性, 释放无限可能



带宽升级, 使能大模型训练极致性能



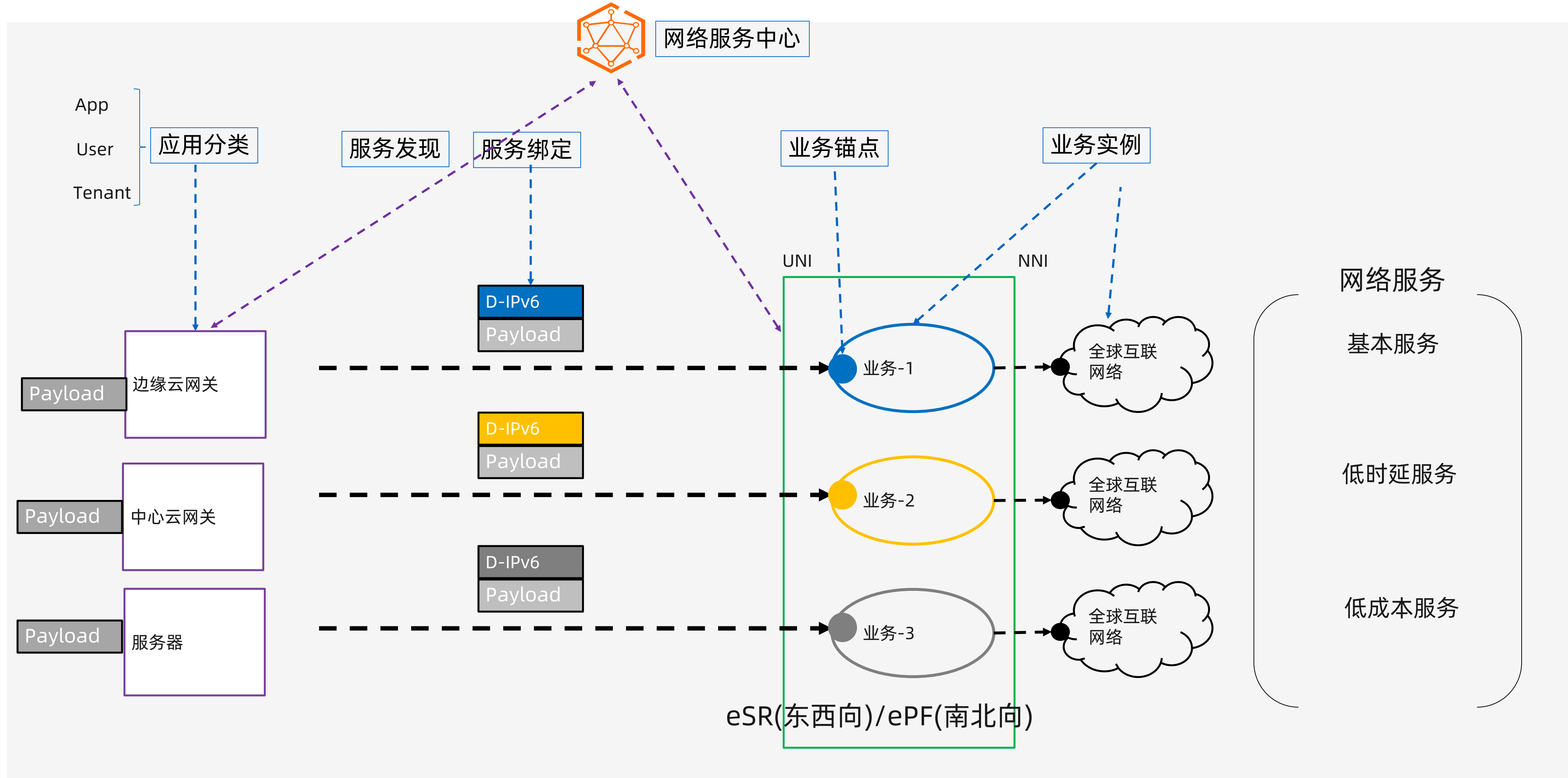
虚拟化技术升级

vSolar: 基于IPv6的高性能AI集群 RDMA 网络方案

- 支持RDMA操作, 支持多租隔离, 支持GPU Direct
- 消除了VXLAN封装, 协议头开销减少30B, 网卡流水线减少1次封装和解封装操作: 小包时延开销小于100ns, 大包带宽几乎无损

可预期广域网：基于 IPv6/SRv6 的部署实践

基于SRv6扩展实现阿里云数据中心网络的**网络服务化能力**，及**精细化算力调度能力**和**水平扩展能力**



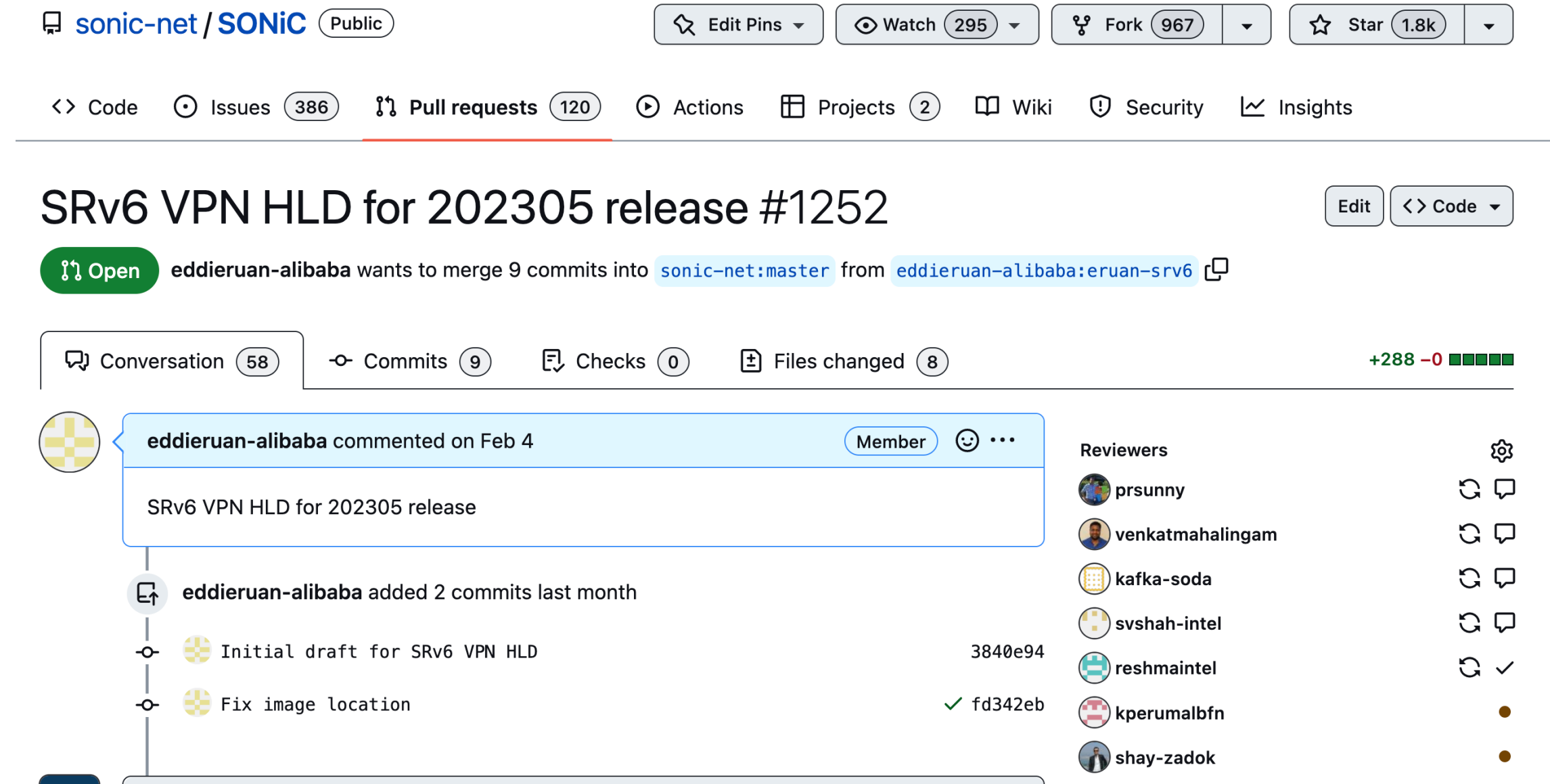
IPv6 国际生态影响

依托全球开源社区推动SRv6技术的普惠化

SONiC (Software for Open Networking in the Cloud) 是一个开源项目，旨在通过提供一个可扩展、灵活的网络操作系统，为云服务提供商和数据中心网络管理员提供更好的工具来管理他们的网络设备。

Premier Members

The SONiC Fund supports the activities of the SONiC Foundation. For information or to join the SONiC Fund, please click [here](#).



创立基于 SRv6 的SONiC 路由工作组：

- 挑战1：如何满足路由器相关特性以及 SRv6VPN 部署需求
- 挑战2：控制平面的高效性，提高 FRR、Linux 内核和 SONiC 之间通信的灵活性
- 挑战3：如何处理大规模 SRv6 VPN 路由

解决方案：与社区合作，对BGP协议进行了补充，并引入了BFD硬件卸载等优化措施，以提高网络性能和可靠性

解决方案：与FRR和SONiC社区合作，重新设计了控制平面模块之间的通信管道，优化了转发对象的下发方式

解决方案：优化内存使用，优化数据结构，引入新的高效且快捷的故障处理方式

技术影响力和生态领导力

阿里云被权威机构评选为世界十大网络研究机构（中国唯一）

SONiC/ODCC 开源社区

SONiC 社区创始成员（中国唯一）

创立 S3IP 白盒硬件开源生态

创立白盒路由器和 OTN 工作组

网络顶会

网络顶级会议论文40+篇
SIGCOMM, NSDI, SOSP,
OFC/ECOC

SIGCOMM
过去5年19篇论文

IETF 协议标准化

协议标准化
HPCC+, IFA, Multi-paths QUIC
SRv6

国内生态 国家奖项, IPv6 生态

中国通信学会
技术发明一等奖
科技进步二等奖（2次）
北京市科技进步二等奖

大规模 IPv6 部署实践反馈生态

浙江省IPv6规模化部署和下一代互联网创新实验室

