云计算开源产业联盟

全球云计算开源峰会 2017

聚合云计算新势力，拥抱全世界新开源

GLOBLE CLOUD COMPUTING OPEN SOURCE SUMMIT

IT大咖说 知识分享平台

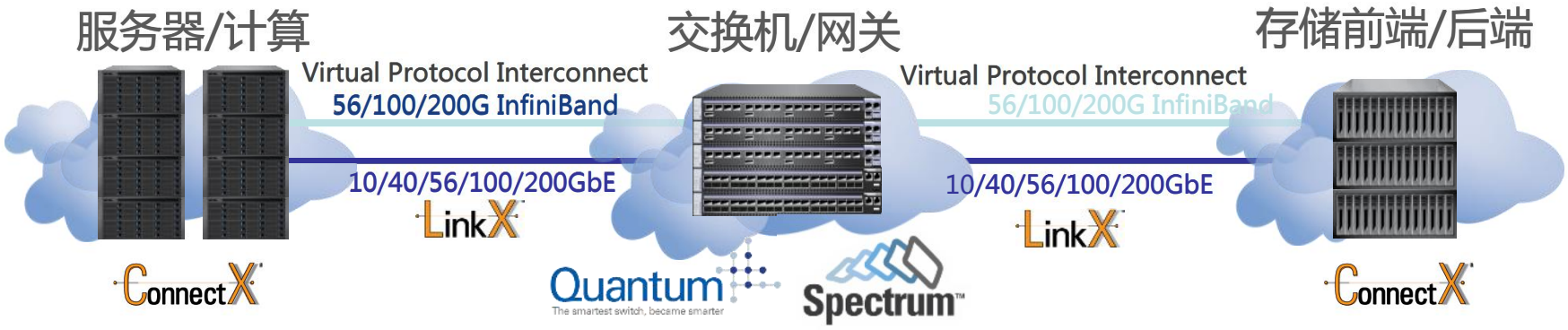
中国 北京

智能网络构建高效云计算平台

张辉， 亚太及中国区解决方案营销总监

April 19th . 2017

领先的端到端互联解决方案提供商



完整的端到端互联产品家族

虚拟化

不牺牲性能的高效、弹性

实现终极云性能的三个障碍

云计算开源产业联盟
IT大咖说 知识分享平台
聚合云计算新势力，拥抱全世界新开源
GLOBLE CLOUD COMPUTING OPEN SOURCE SUMMIT

低效的网络协议

有状态的传输协议(如TCP使卸载更复杂,造成额外的CPU开销）

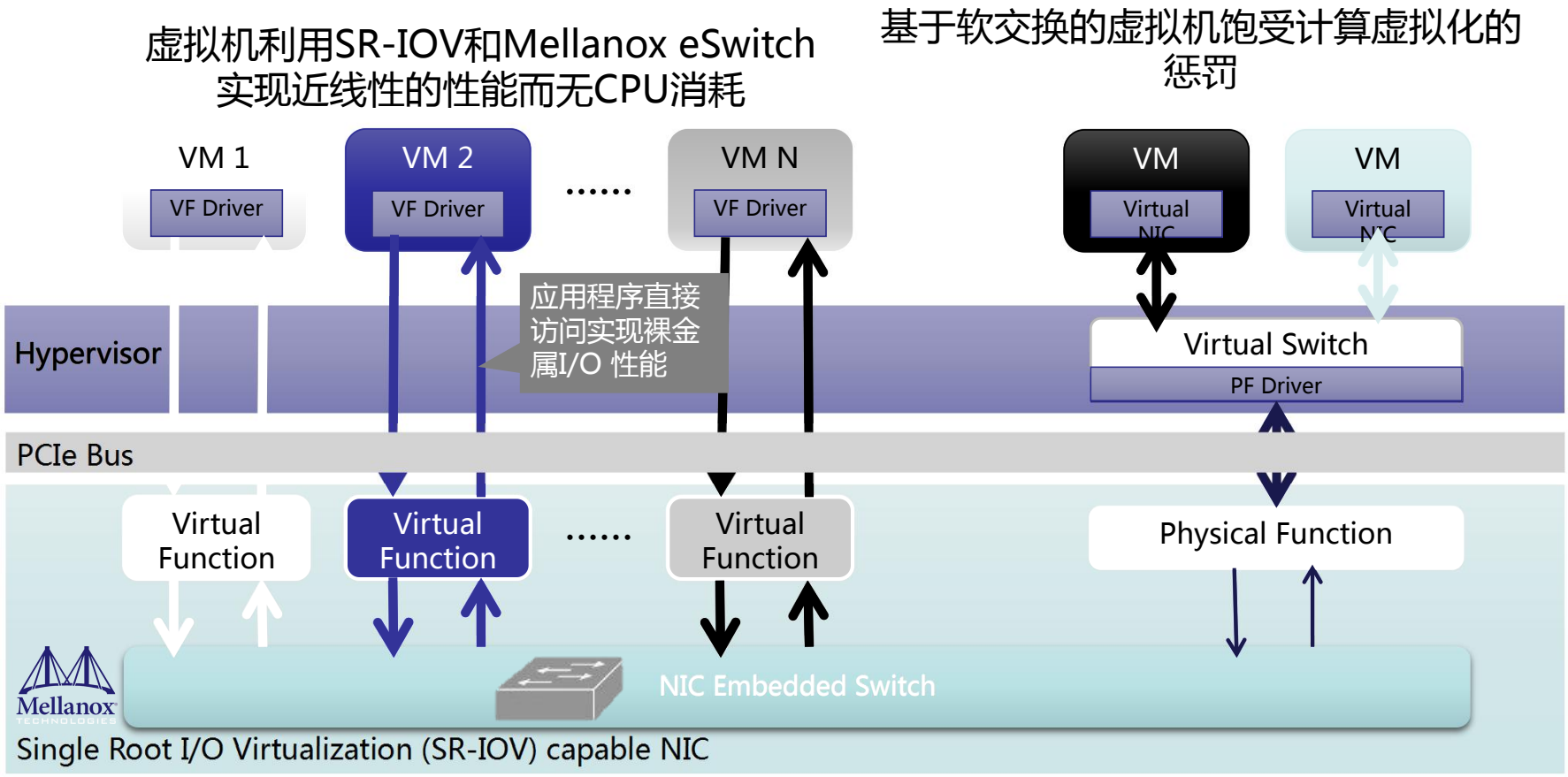
计算虚拟化的惩罚

由于传输额外的软件程序开销引起的I/O性能下降

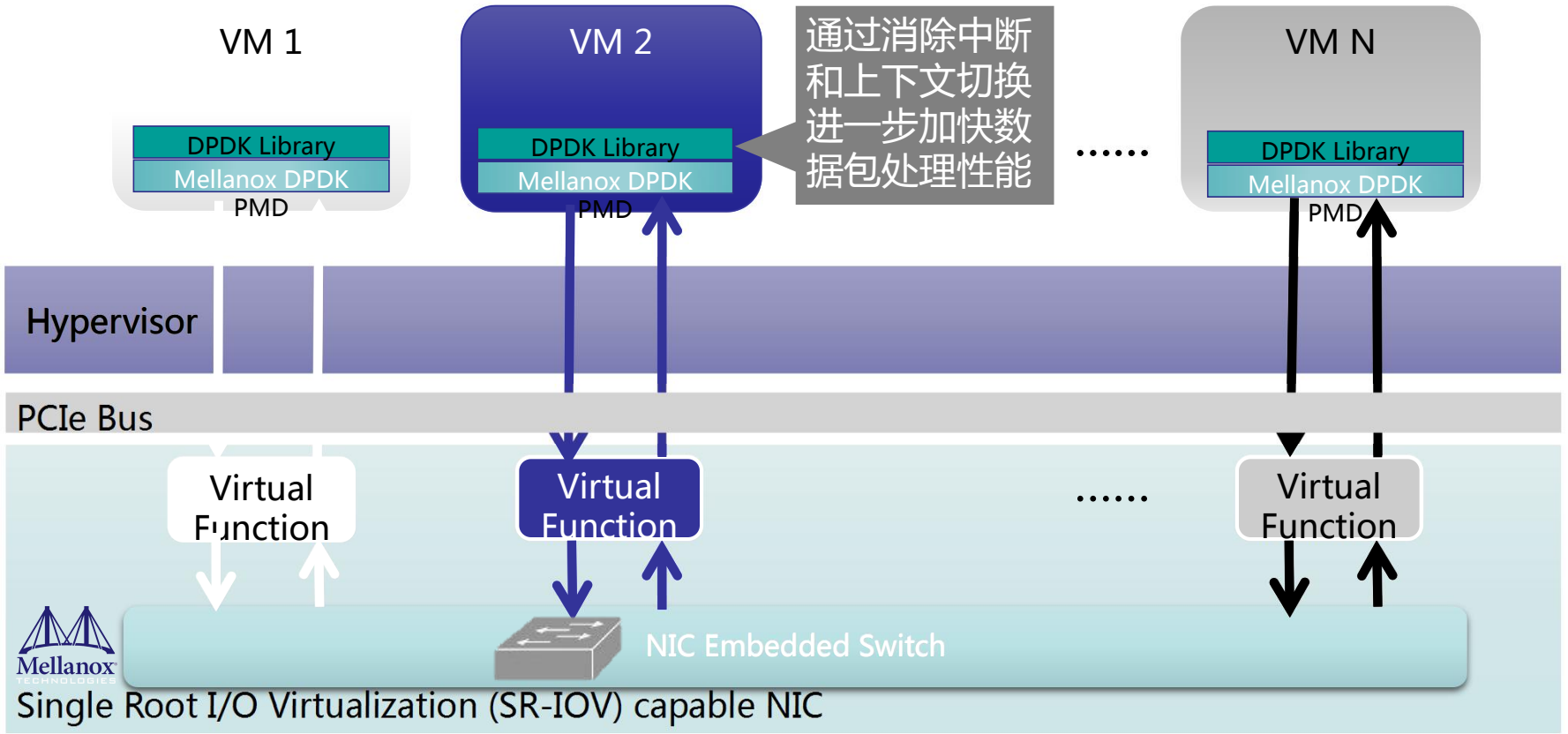
网络虚拟化的惩罚

Overlay 网络虚拟化导致新的数据包报头格式的I/O，系统可能无法处理

SR-IOV – 克服计算虚拟化的惩罚

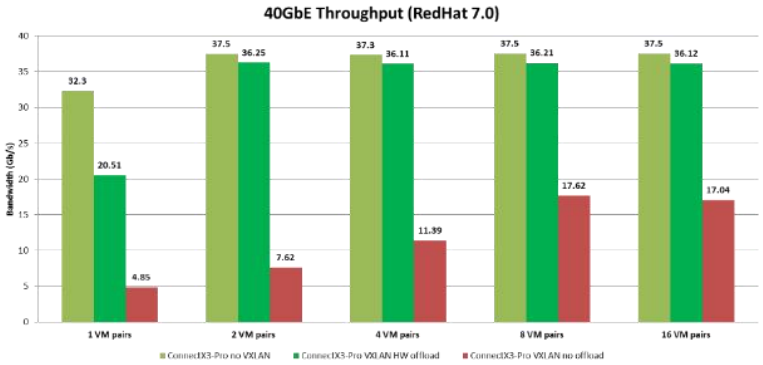


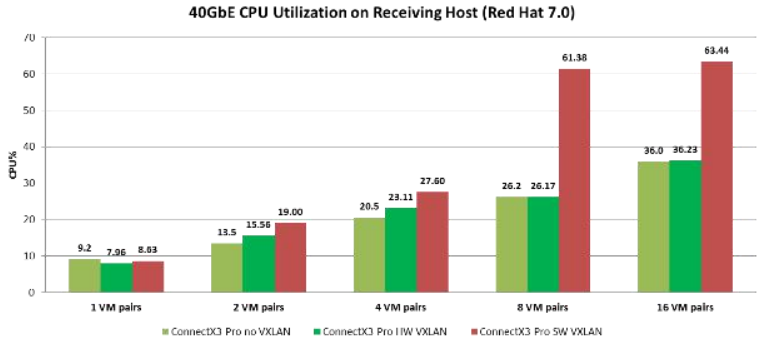
SR-IOV + DPDK: 与Mellanox PMD协同效果更佳



VXLAN卸载 – 克服网络虚拟化的惩罚

- 解决方案：
  - 网卡中的Overlay网络加速器
  - 以裸金属速度运行无性能牺牲
  - 主流的SDN厂商已验证和集成
- 收益：
  - 40Gb链路上达到37.5Gb/s，2倍的性能提升（相对没有使用VXLAN卸载）
  - 在一个20个Core的系统中，释放出7个Core 用于运行虚拟机，节省35%的内核，同时实现吞吐率两倍提升！



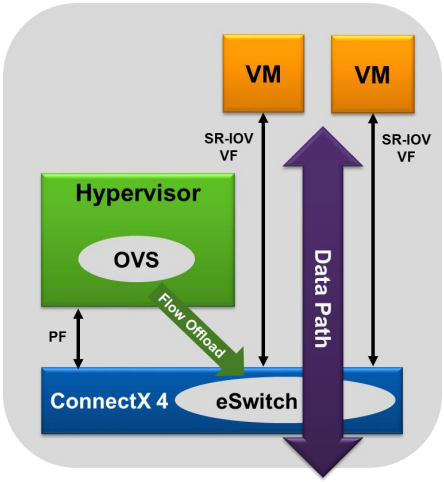


PLUMgrid nuagenetworks Hewlett Packard Enterprise midokura vmware NSX

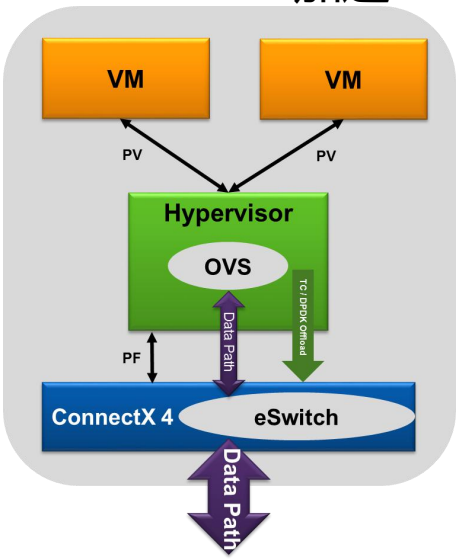
Accelerated Switch And Packet Processing (ASAP²)

- ASAP² 利用ConnectX-4能力加速或卸载“主机”网络堆栈
- 三种主要使用方式

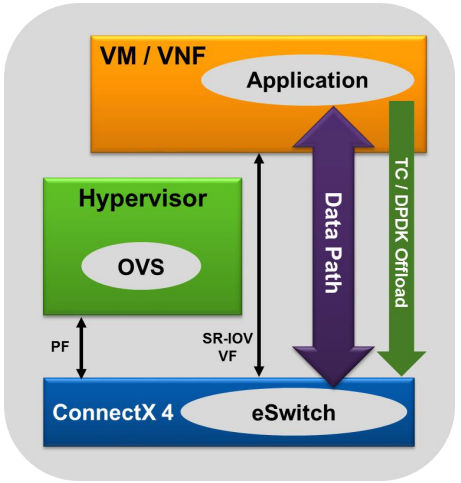
ASAP² Direct
全vSwitch卸载



ASAP² Flex
vSwitch 加速

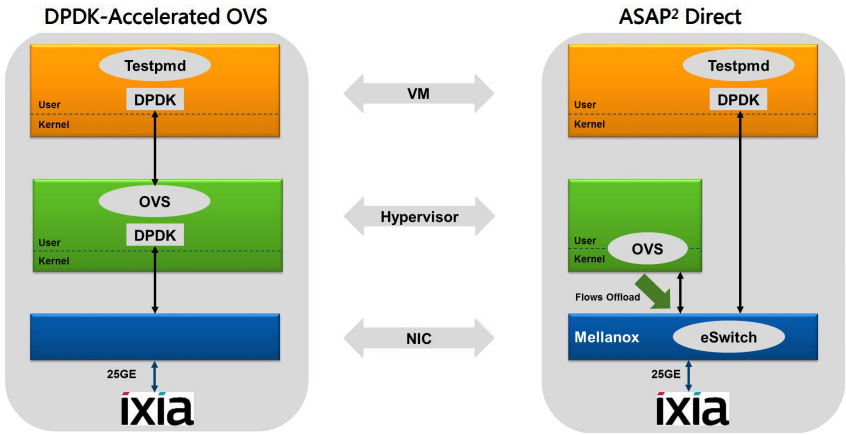


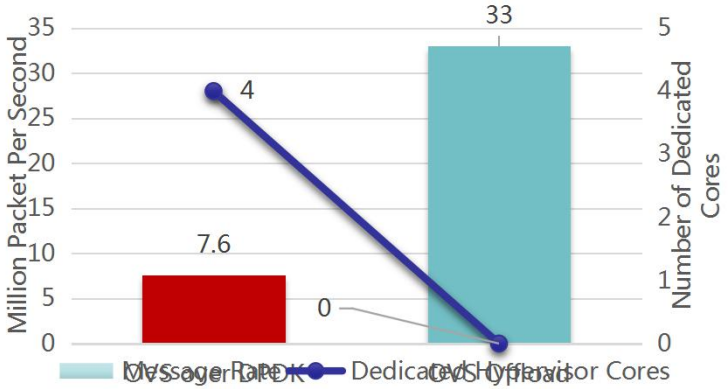
ASAP² Flex
VNF/VM 加速



DPDK加速的OVS vs Mellanox ASAP² Direct

- 1 flow, 没有VXLAN
  - 消息传输率是OVS over DPDK的3.3倍
  - CPU 消耗率为零！对比OVS over DPDK却使用4个CPU内核
    - 释放出来的CPU可以用于运行虚拟机
- 2000 flows, VXLAN硬件封装/解封装
  - OVS offload 实现 ~25MPPS
  - CPU消耗依然为零，DPDK却使用4个CPU内核





加速

为横向扩展的应用程序启用快速网络和存储访问

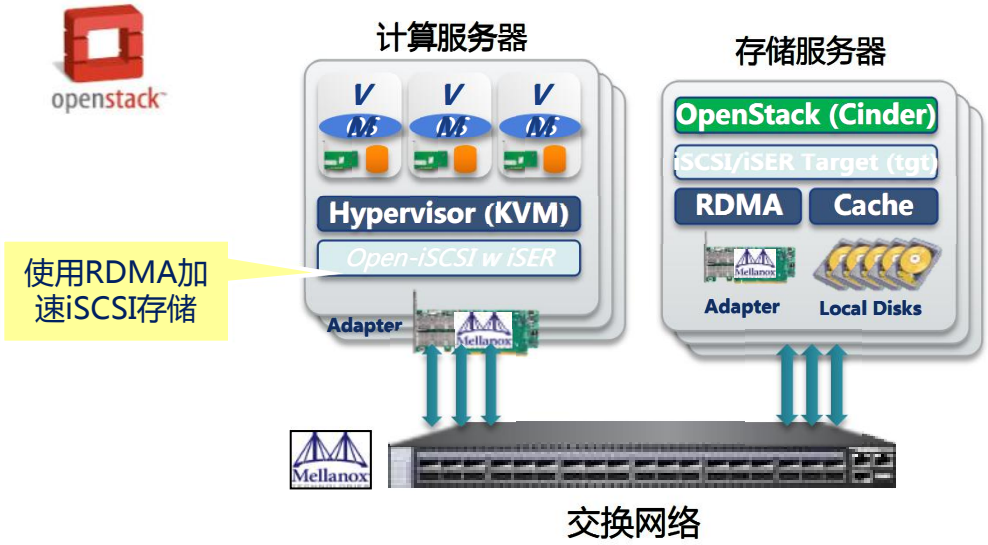
Spectrum: 25/100GbE交换机中的旗舰

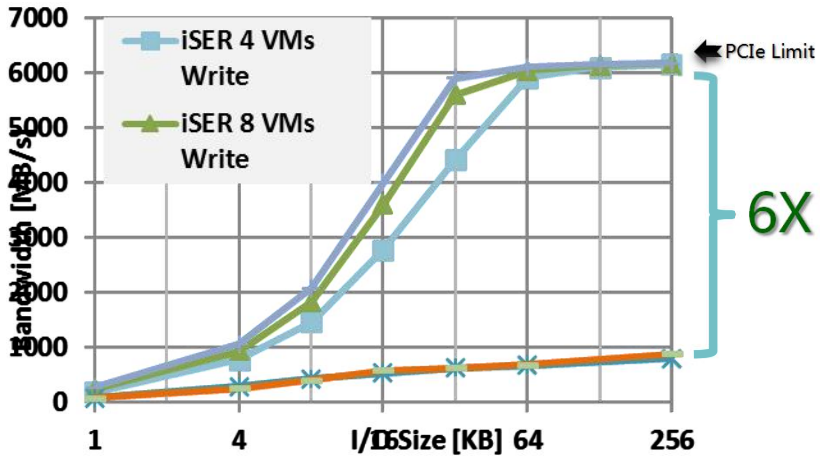
Spectrum™

简单易用	性能	扩展性
• 智能可视 • 云敏捷及QoS • 自动化	• 10倍延迟 • 包零丢失 • 智能缓冲	• 更低的功耗 • 2倍的转发表 • 更低的成本$$/Gbps

- 最高的包转发率
- 更好的缓冲：10-15倍的 microburst性能提升
- 可预测的运行状态
- 更低的延迟: 降低50%
- 更低的能耗

RDMA加速访问OpenStack存储



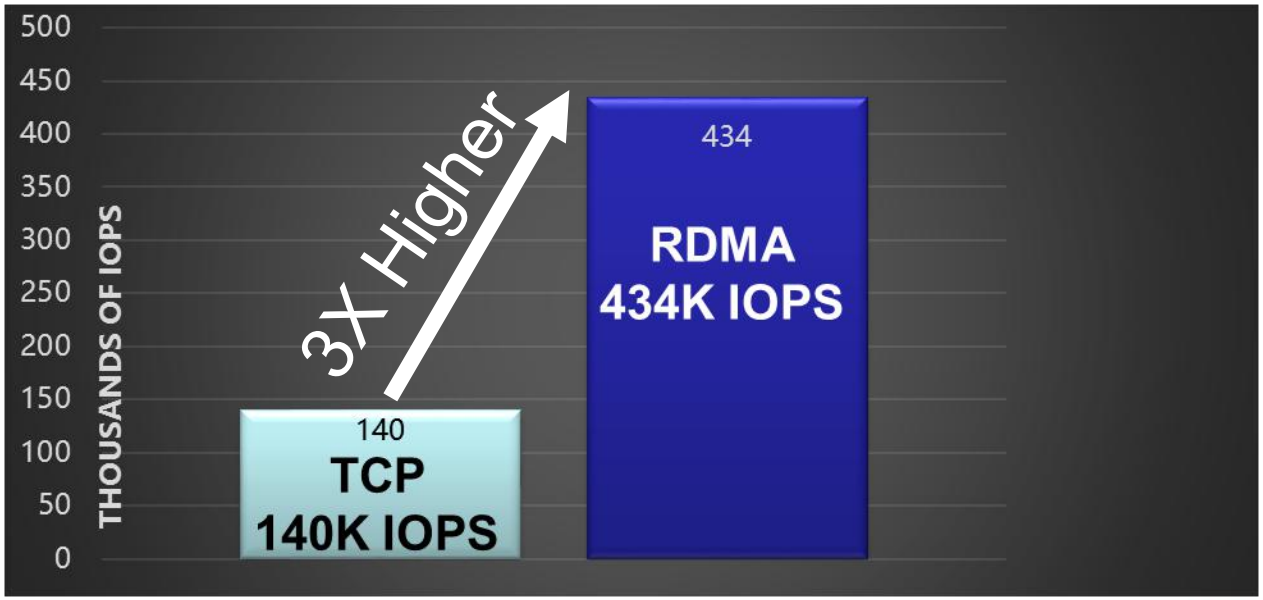


- 内置的OpenStack组件和管理
  - 不需要额外的软件
  - RDMA已经内置到OpenStack并被用户广泛使用

RDMA实现带宽提升6倍，延迟降低5倍，并且降低CPU利用率

使用RDMA加速Ceph

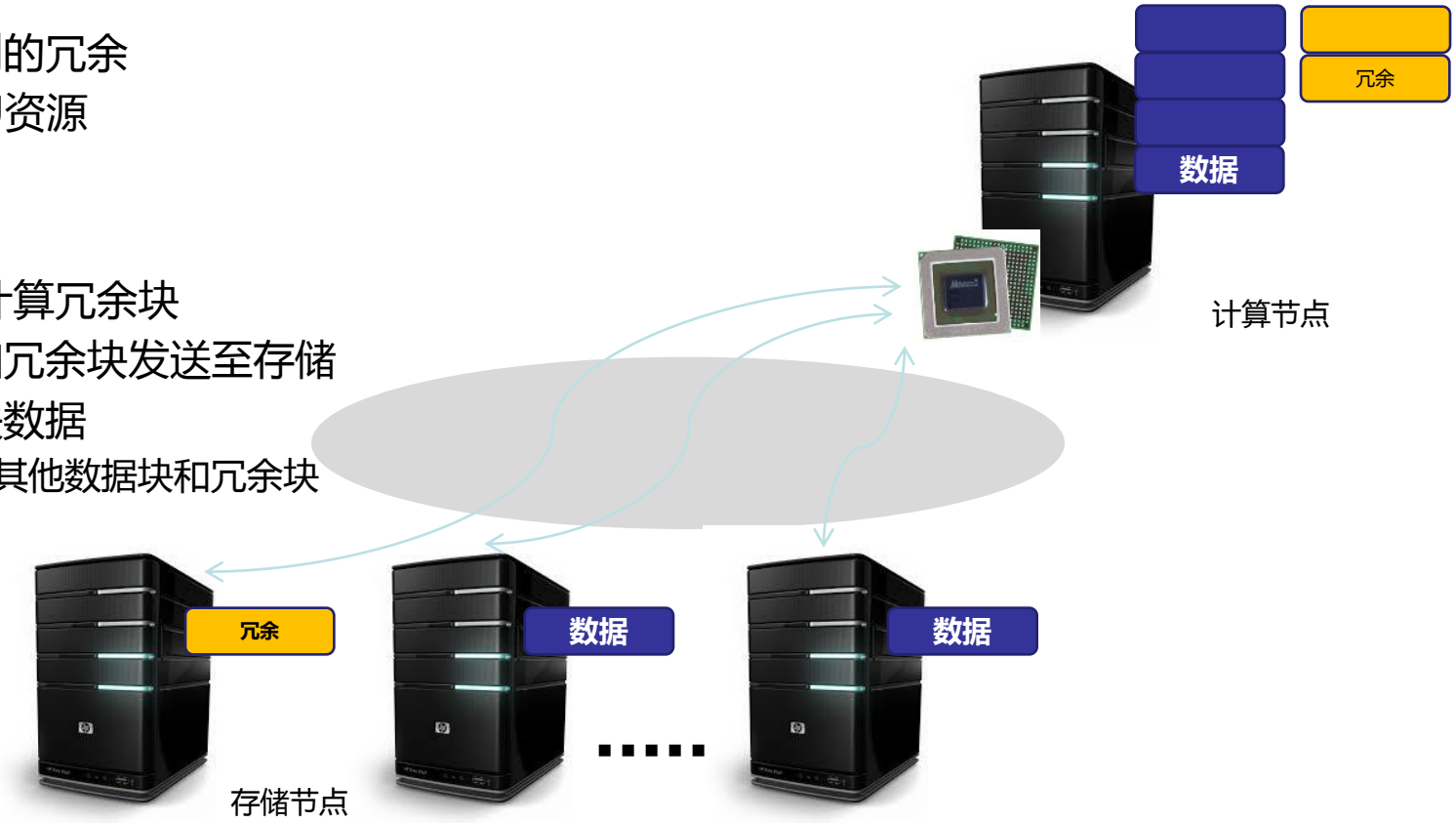
- 更快的RDMA集成至应用
- 最大化消息和CPU并发
- 单节点 > 10GB/s
- 延迟 < 10usec



Ceph Read IOPS: TCP vs. RDMA

纠删码卸载（Erasure Coding）

- 价值
  - 集群级别的冗余
  - 释放CPU资源
- 硬件卸载
  - RS编码计算冗余块
  - 数据块和冗余块发送至存储
  - 重算丢失数据
    - 给予其他数据块和冗余块



自动化

操作简单、高效

开放组合网络（OCN）

Mellanox TECHNOLOGIES

Microsoft SONiC　Metaswitch Networks　Open Switch　Cumulus® Linux® Network OS　Mellanox TECHNOLOGIES

精选的网络操作系统

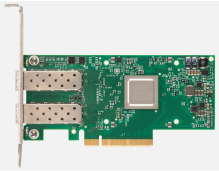
sai　switchdev

开放的API

Spectrum™

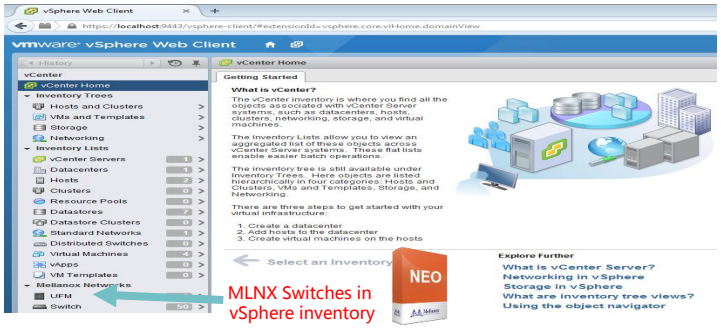
NEO

自动化

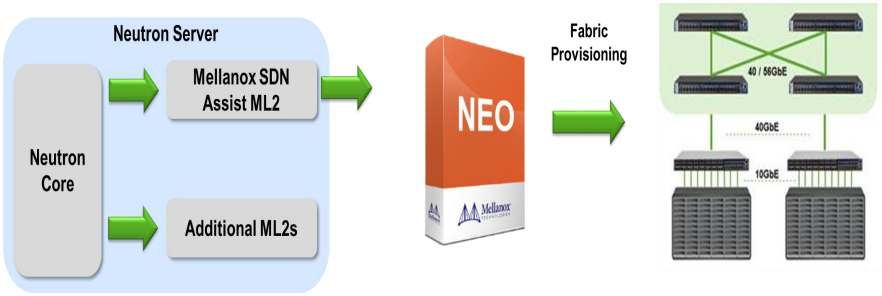


端到端互连

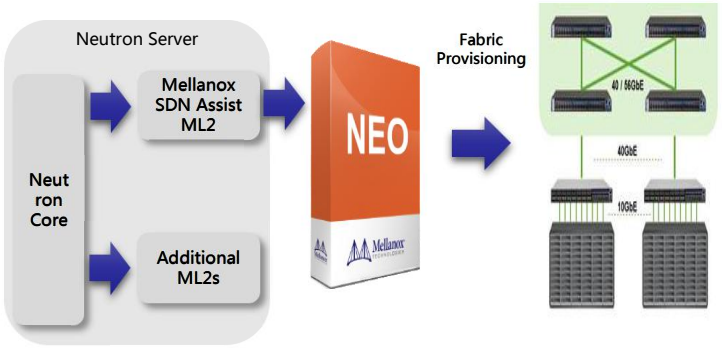
操作简单、高效



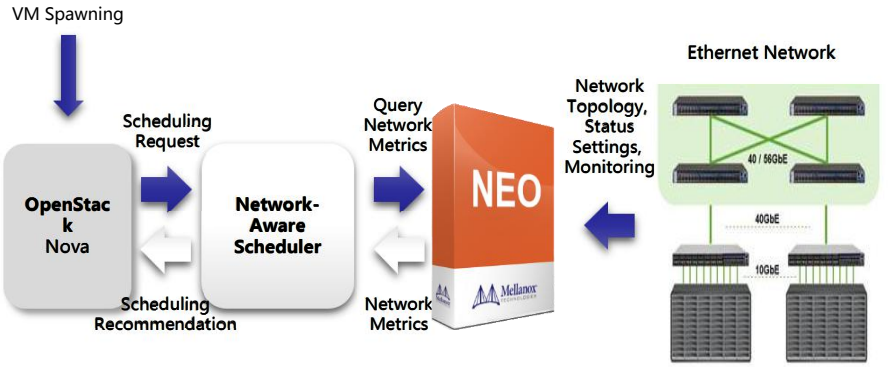
NEO与vSphere集成



全面支持以太网和InfiniBand



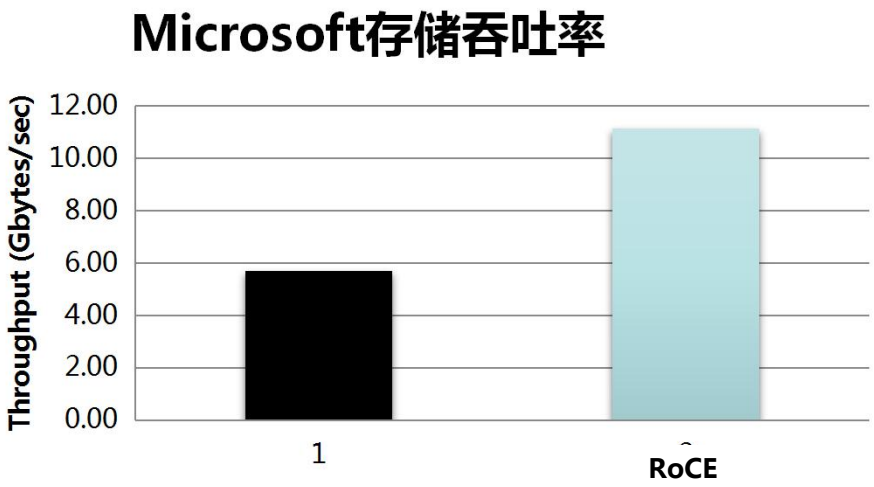
零接触网络配置

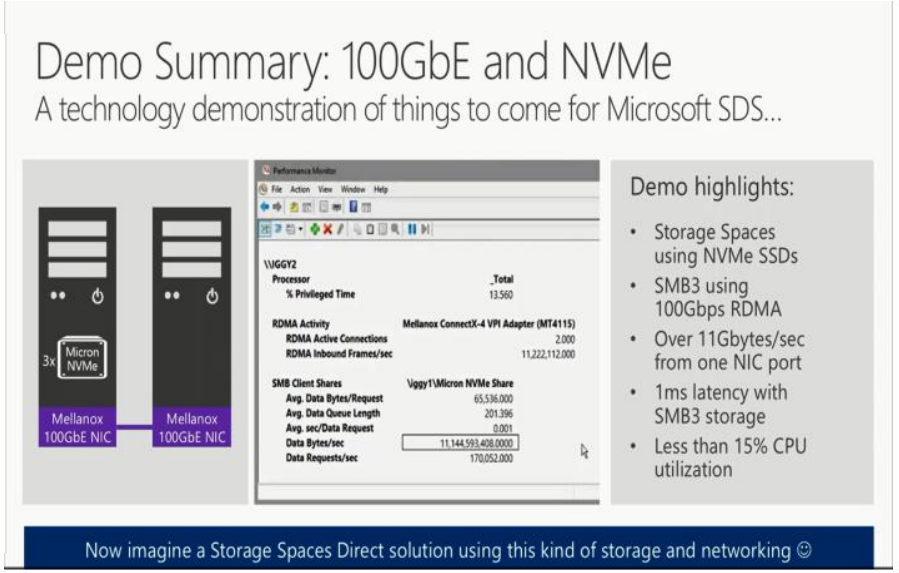


网络感知调度

案例分享

Mellanox 100G方案使云效率最大化





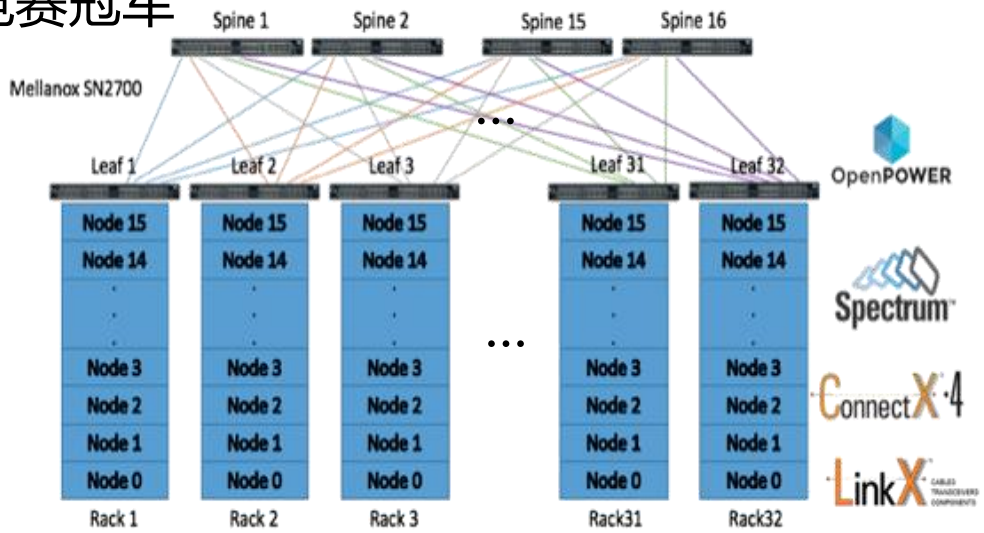
- RDMA在Windows 2016及 AzureStack被广泛支持
- 使用 RoCE 可以提升2倍带宽及2倍CPU利用率
- RoCE 充分利用的闪存存储带宽
  - 远程存储再不妥协

Microsoft

腾讯云创造 Terasort 世界纪录

TeraSort 年度排序基准全球计算竞赛冠军

- Mellanox 和 IBM 携手 Tencent 实现超过 1TB/秒的 TeraSor 性能新纪录
- 打破了 GraySort 和 MinuteSort 类别的纪录，将去年的成绩提高了5.8 倍
- 每节点的性能比上一届提升了33 倍



* http://sortbenchmark.org/

非常感谢