INDONESIA
OpenInfra Days

02.11.2019 | Surabaya, Indonesia

Cloud Done Right!

Bringing Web-Scale Innovations to Every Data Center

David Iles – Senior Director of Ethernet Switching
Mellanox Technologies

davidi@mellanox.com

Biznet GioCloud Biznet Mellanox TECHNOLOGIES BANK BRI OSF OpenStack Foundation

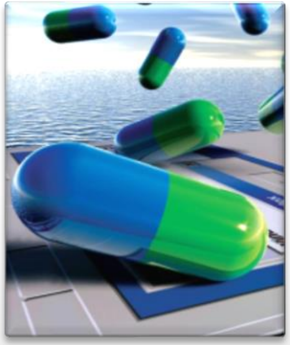
Mellanox Leadership Across Industries

5 of Top 6

Global Banks

10 of Top 10

Automotive Manufacturers



3 of Top 5

Pharmaceutical Companies

9 of Top 10

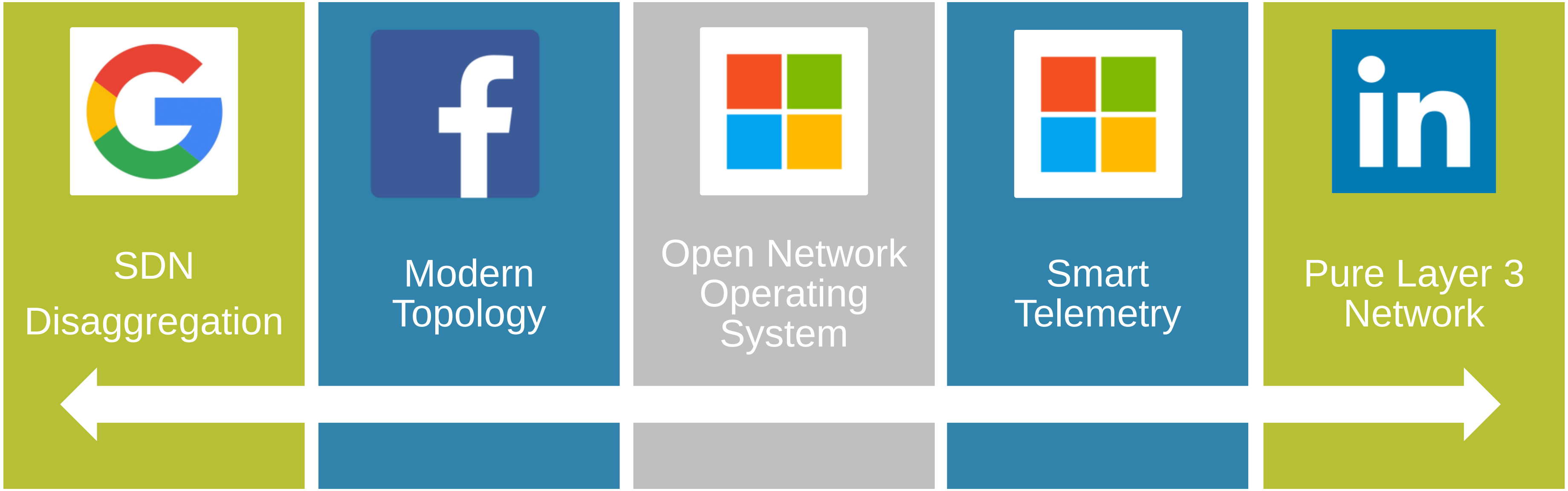
Oil and Gas Companies

9 of Top 10

Hyperscale Companies

Mellanox Interconnect Solutions Deliver Highest Return on Investment

What have Cloud Titans taught the Industry?



SDN Disaggregation

Modern Topology

Open Network Operating System

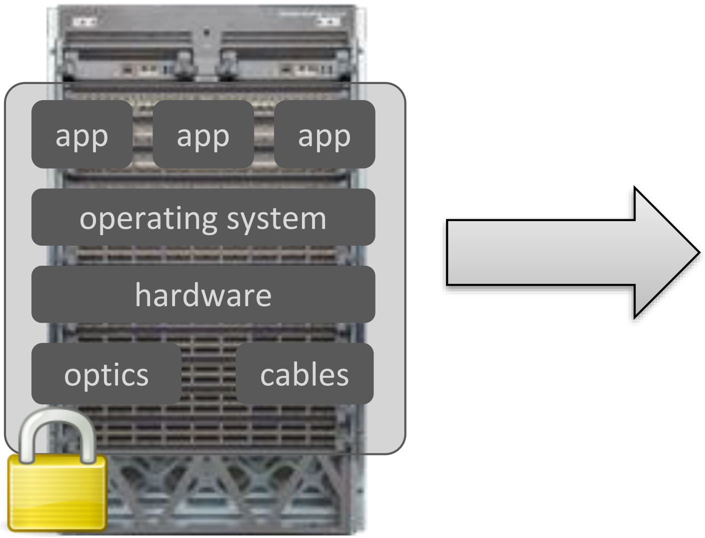
Smart Telemetry

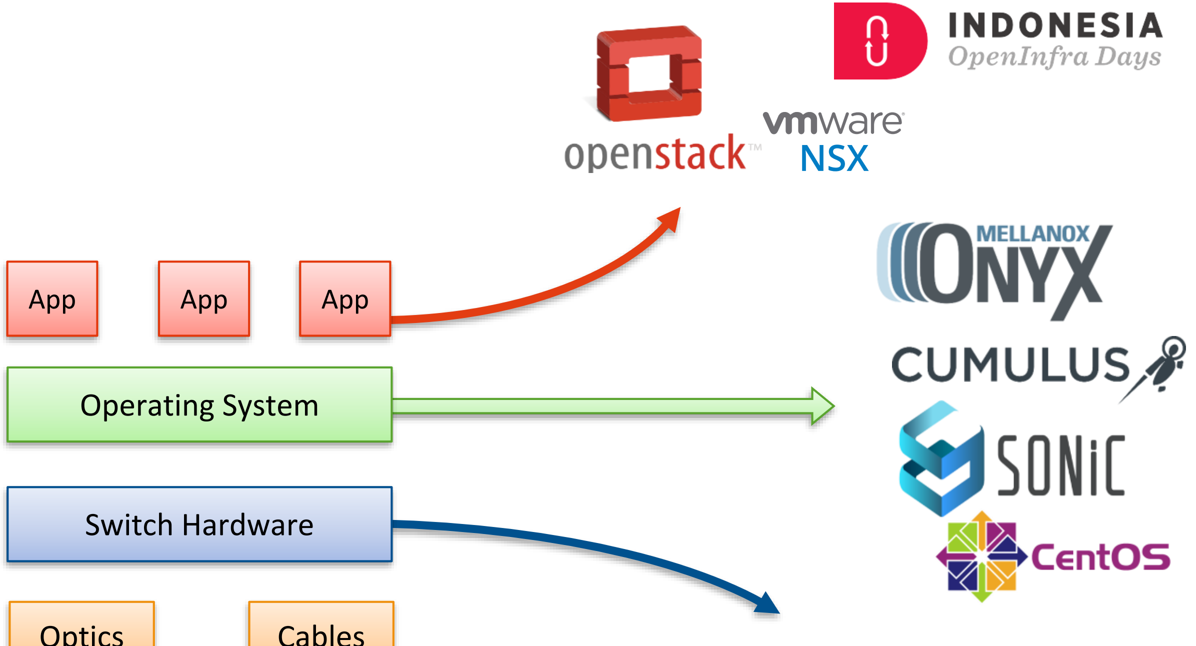
Pure Layer 3 Network

We bring Cloud Titan innovations to you!

Web-Scale Innovation:

Leverage Open Platforms





Mainframe-like Networks:

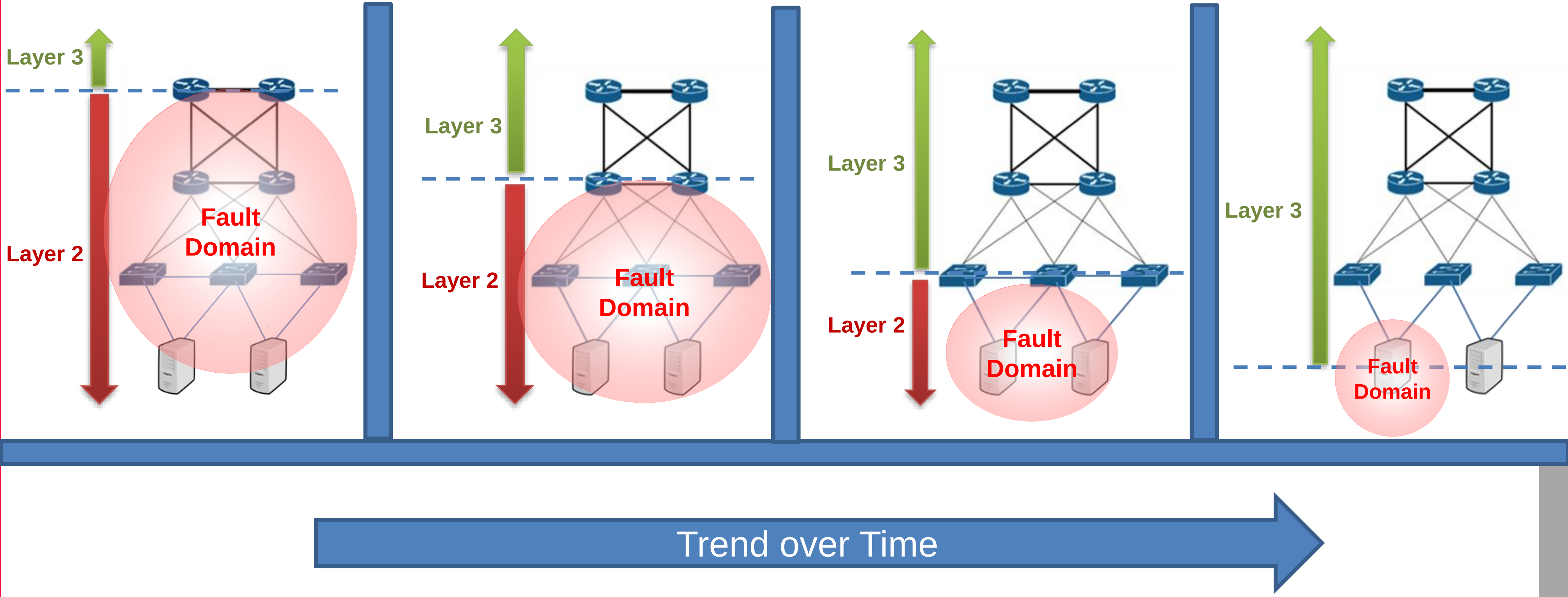
- Vendor lock-in
- Higher switch prices
- Higher support prices

Open Networking:

- Best of breed hardware
- Best of breed software
- No hidden licenses/royalties

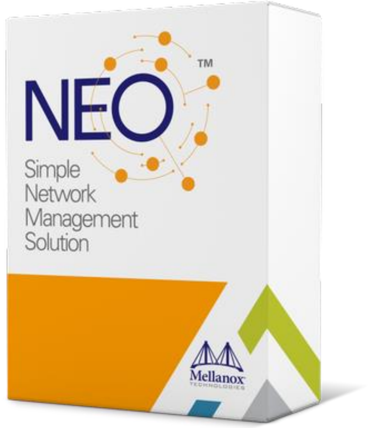
Web-Scale Innovation:

From Layer 2 to Layer 3



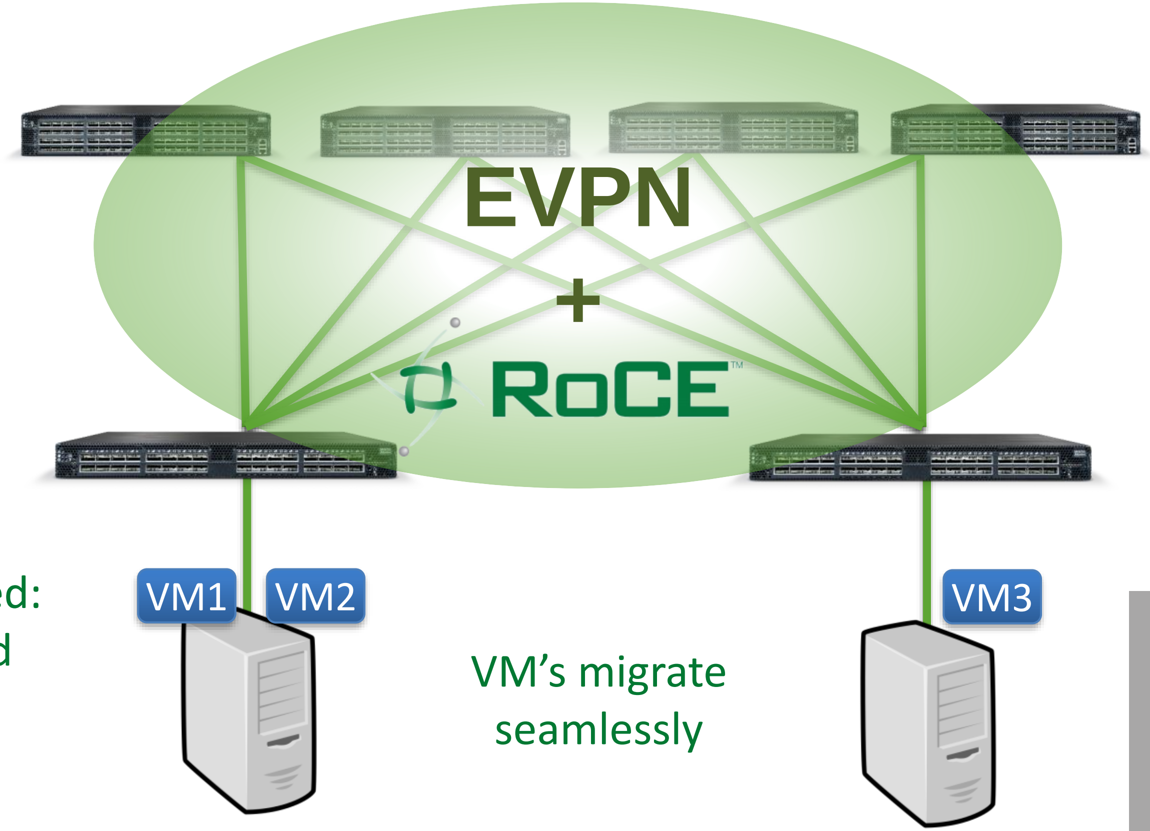
Web-Scale Innovation:

VXLAN without compromise



openstack.

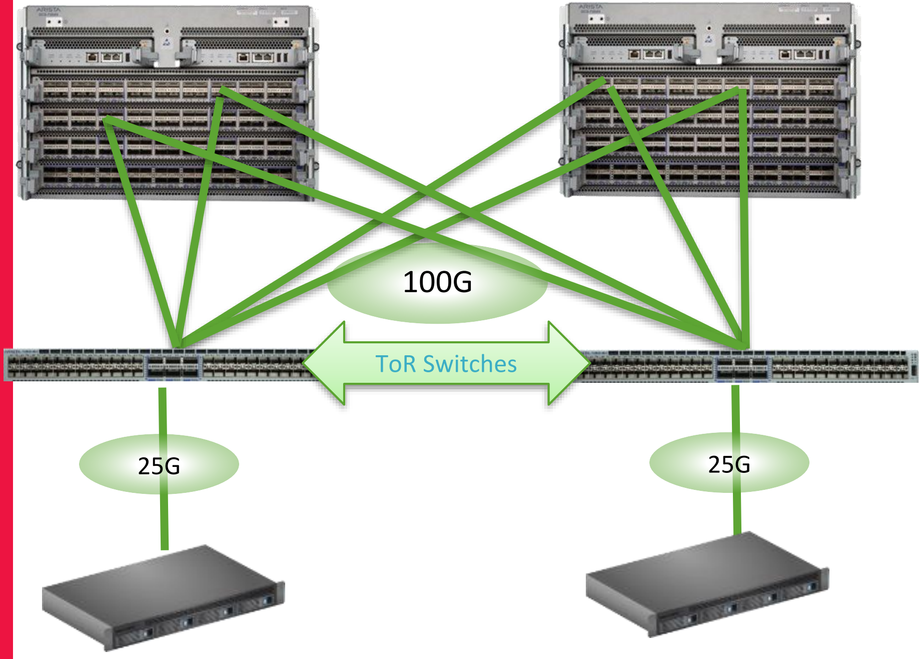
When VMs are deployed:
VLAN Auto-configured
&
mapped to VXLAN

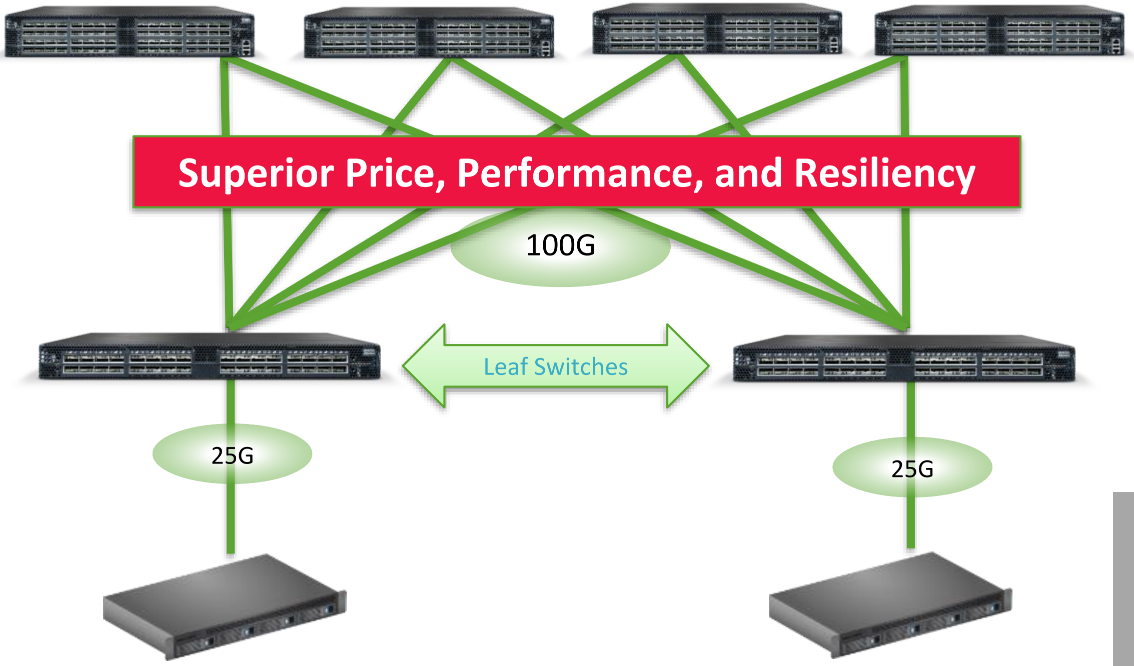


VM's migrate seamlessly

Web-Scale Innovation:

Leaf/Spine Networks

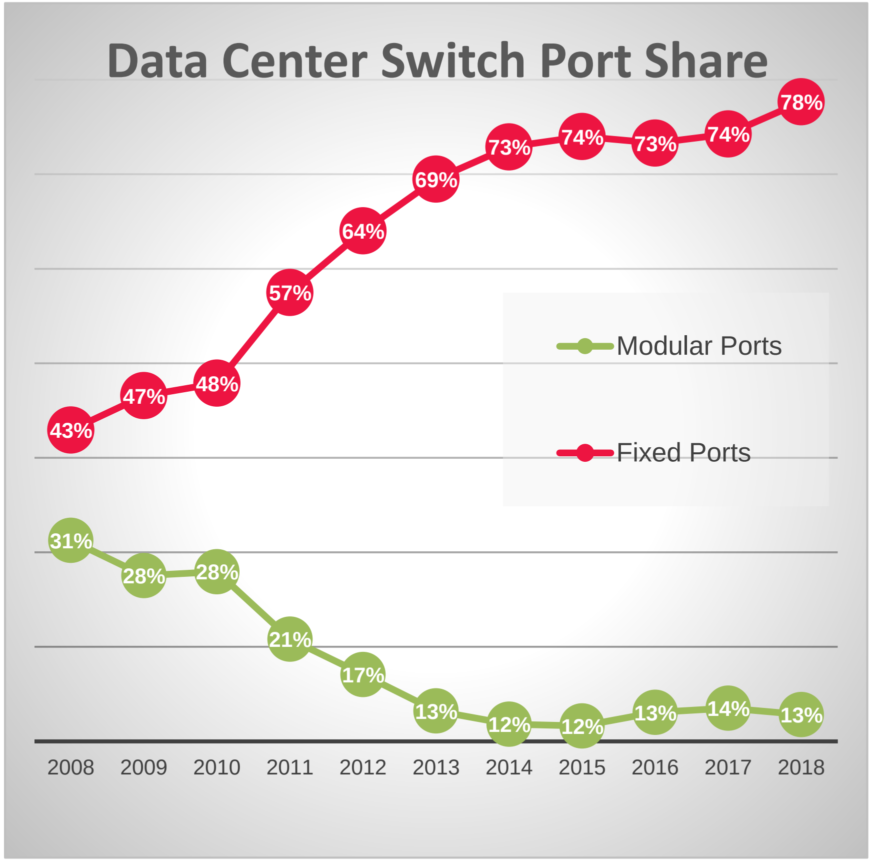




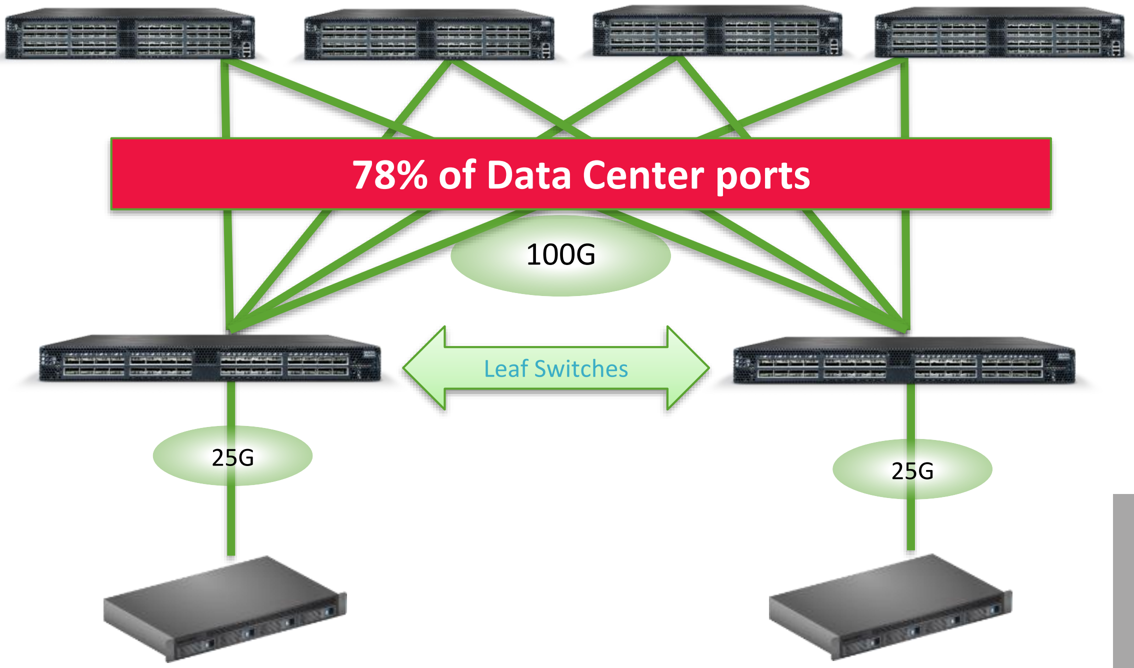
Scale Up Network → **Scale Out Network**

Web-Scale Innovation:

Trend from Modular to Fixed Port Switches



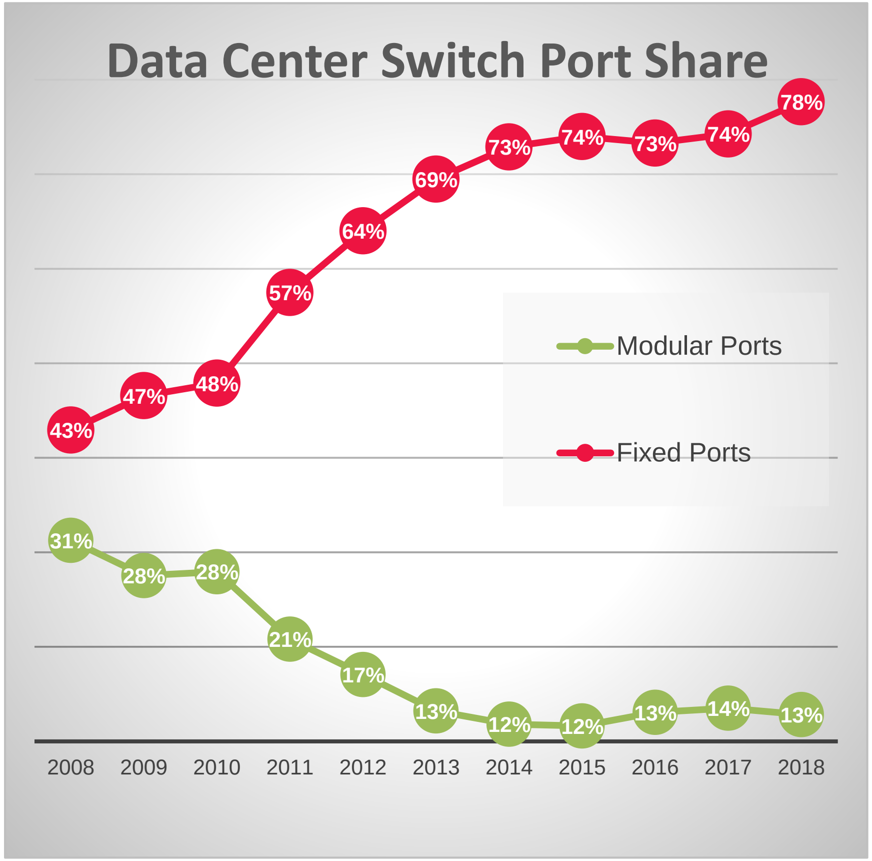
Crehan Quarterly Market Shares - Data Center Ethernet, July 2019



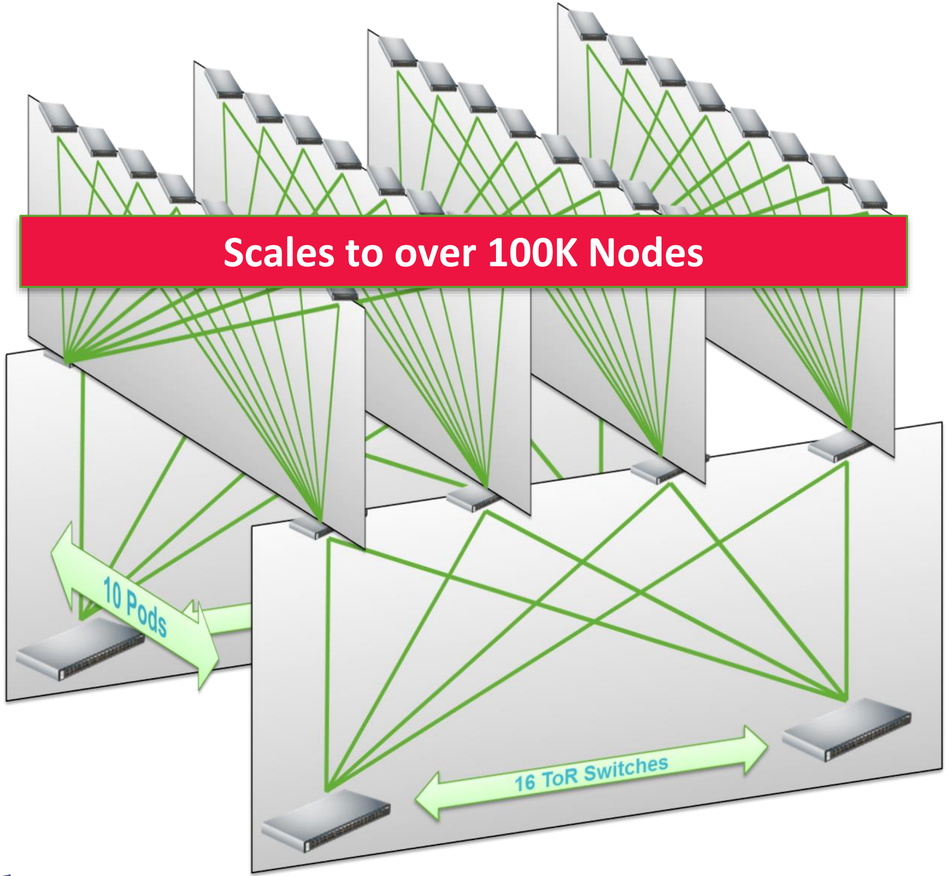
INDONESIA
OpenInfra Days

Web-Scale Innovation:

Trend from Modular to Fixed Port Switches



Crehan Quarterly Market Shares - Data Center Ethernet, July 2019



Biznet GioCloud Biznet Mellanox TECHNOLOGIES BANK BRI OSF OpenStack Foundation

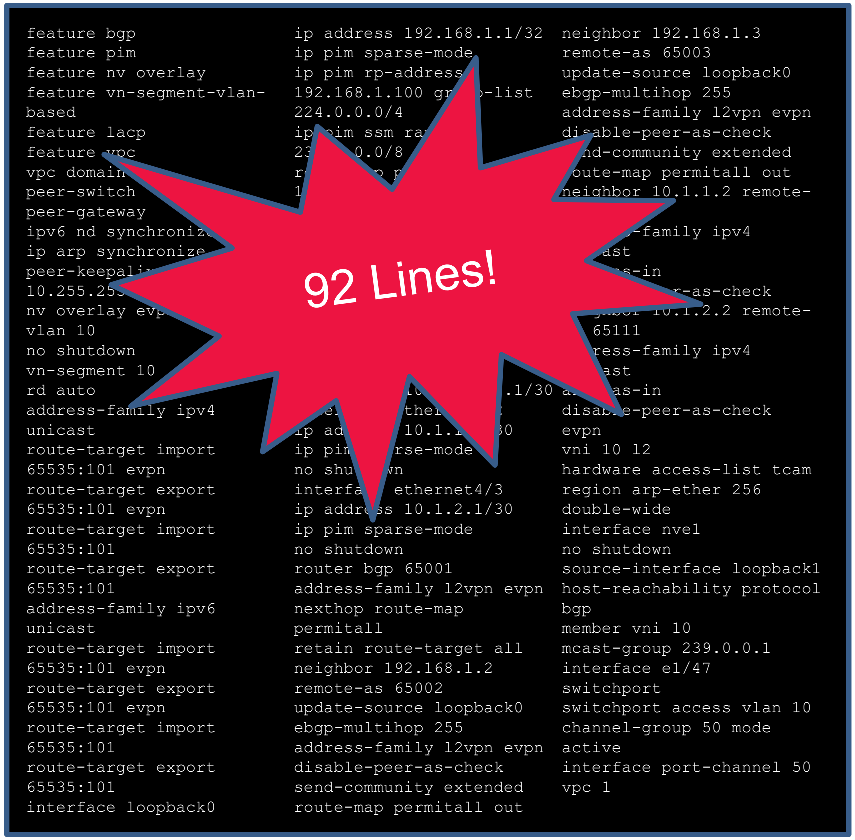
Web-Scale Innovation:

Simplified Configs

Cumulus EVPN Config - 11 Lines

Server bond	1.	`net add bond bond01 bond slaves swp1`
	2.	`net add bond bond01 bridge access 10`
	3.	`net add bonding bond01 clag id 1`
MLAG (vPC)	4.	`net add clag peer sys-mac 44:38:39:FF:00:01 interface swp49`
Loopback IP	5.	`net add loopback lo ip address 192.168.1.1/32`
VxLAN	6.	`net add vxlan vni10 vxlan id 10`
	7.	`net add vxlan vni10 vxlan local-tunnelip 192.168.1.1`
	8.	`net add interface vni10 bridge access 10`
BGP	9.	`net add bgp autonomous-system 65000`
	10.	`net add bgp neighbor swp51,swp52 interface remote-as external`
	11.	`net add bgp neighbor l2vpn evpn neighbor swp51,swp52 activate`

Other's EVPN Config



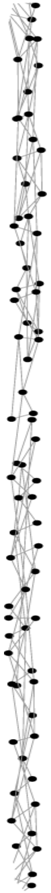
```
feature bgp                  ip address 192.168.1.1/32  neighbor 192.168.1.3
feature pim                  ip pim sparse-mode         remote-as 65003
feature nv overlay           ip pim rp-address          update-source loopback0
feature vn-segment-vlan-     192.168.1.100 gr  -list    ebgp-multihop 255
based                        224.0.0.0/4                address-family l2vpn evpn
feature lacp                 ip  im ssm ra              di able-peer-as-check
feature vpc                  23   0.0/8                  nd-community extended
vpc domain                   r     p p                   oute-map permitall out
peer-switch                  1                          neighbor 10.1.1.2 remote-
peer-gateway
ipv6 nd synchroniz                                          -family ipv4
ip arp synchronize                                          ast
peer-keepali                                                s-in
10.255.25                                                  r-as-check
nv overlay evp                                              10.1.2.2 remote-
vlan 10                                                    65111
no shutdown                                                ress-family ipv4
vn-segment 10                                              ast
rd auto                          10        .1/30        a   as-in
address-family ipv4           e      ther               disab e-peer-as-check
unicast                      ip ad     10.1.1   30      evpn
route-target import          ip pin    rse-mode         vni 10 l2
65535:101 evpn               no shu    wn               hardware access-list tcam
route-target export          interfa   ethernet4/3      region arp-ether 256
65535:101 evpn               ip addr ss 10.1.2.1/30     double-wide
route-target import          ip pim sparse-mode         interface nve1
65535:101                    no shutdown                no shutdown
route-target export          router bgp 65001           source-interface loopback1
65535:101                    address-family l2vpn evpn  host-reachability protocol
address-family ipv6          nexthop route-map          bgp
unicast                      permitall                  member vni 10
route-target import          retain route-target all    mcast-group 239.0.0.1
65535:101 evpn               neighbor 192.168.1.2       interface e1/47
route-target export          remote-as 65002            switchport
65535:101 evpn               update-source loopback0    switchport access vlan 10
route-target import          ebgp-multihop 255          channel-group 50 mode
65535:101                    address-family l2vpn evpn  active
route-target export          disable-peer-as-check      interface port-channel 50
65535:101                    send-community extended    vpc 1
interface loopback0          route-map permitall out
```

Web-Scale Innovation:

Simplified Configs

Mellanox “Do RoCE”

```
switch (config) # roce
```



Other’s RoCE Config

Step 1 – Ingress Traffic Classification

```
class-map type qos match-all CNP
match dscp 48
class-map type qos match-all
RDMA
match dscp 26
policy-map type qos QOS_MARKING
class RDMA
set qos-group 3
class CNP
set qos-group
```

Step 2 – QoS Policies

```
policy-map type n
QOS_NETWORK
class type netwo
pause pfc-cos 3
mtu 2240
policy-map type     uing
QOS_QUEUEING
class type queuing c-out 8q-q3
random-detect minimum-threshold
150 kbytes maximum-threshold
1500 kbytes drop-probability 100
weight 0 ecn
bandwidth remaining percent 20
class type queuing c-out-8q-q6
priority level 1
policy-map type queuing
INPUT_QOS_QUEUEING
class type queuing c-in-q3
queue-limit dynamic 3
system qos
```

```
service-policy type queuing
input INPUT_QOS_QUEUEING
service-policy type queuing
output QOS_QUEUEING
service-policy type network-qos
OS_NETWORK
```

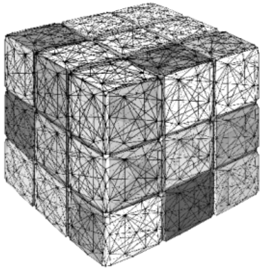
ep 3 –Res urce Allocation

```
 rd   access-list tcam region

          -list tcam region
           ce 0
          ss-list tcam region

         ccess-list tcam region
          -lite 0
   a      ss-list tcam region
   56
   are access-list tcam region
e-q  256
```

24 Lines!



Web-Scale Innovation:

Automate everything

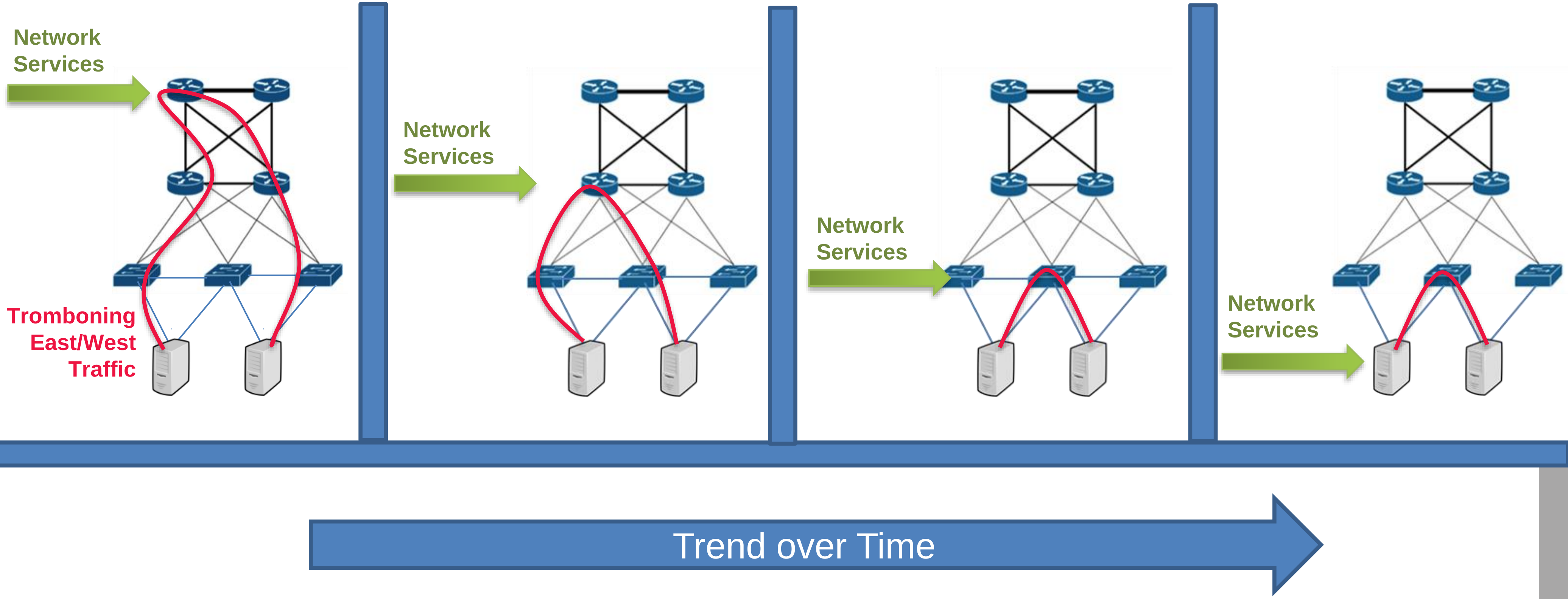
INDONESIA
OpenInfra Days

	TRADITIONAL NETWORKING	**Web-Scale NETWORKING**
Operational Leverage	1 admins : 4 Switches	**1 admin : 500 Switches**
Provisioning	Weeks	**Minutes**
Supply Chain	Locked-in	**Open Supply Chain**
3rd Party Integration	Vendor Determines	**Customer Determines**
Management Tools	Vendor Driven	**Customer Choice**
Robustness / Reliability	Manual & Highly Error Prone	**Automated & Reduced Network Downtime**

$ Up to **75%** Reduction in OpEx with Web-Scale Networking

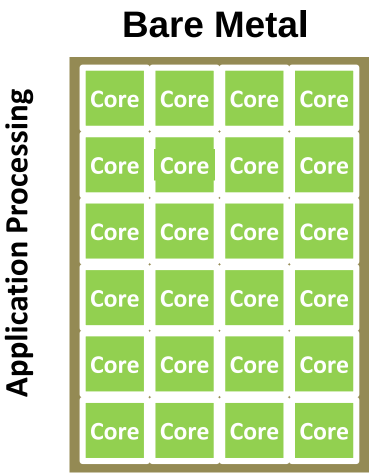
Web-Scale Innovation:

Moving Intelligence to the Edge



Software Defined Everything

Creates Bottlenecks



Bare Metal

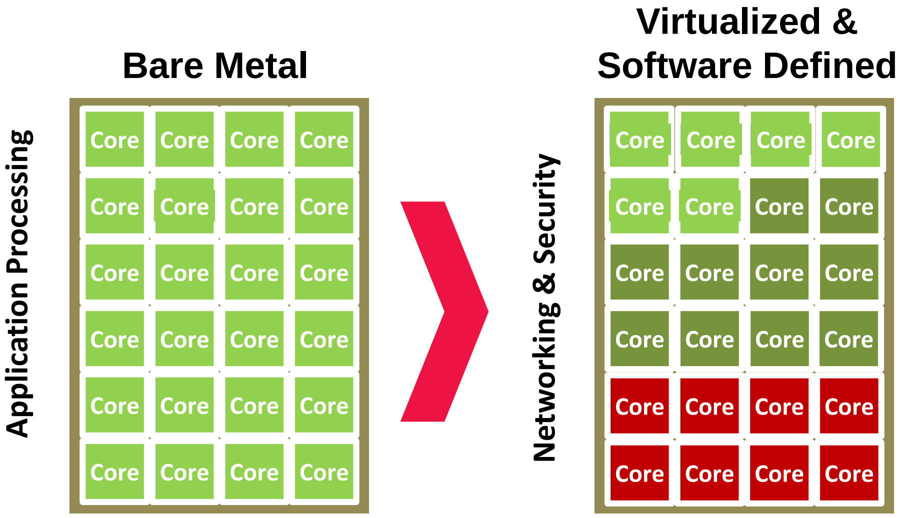
Application Processing

Core	Core	Core	Core
Core	Core	Core	Core
Core	Core	Core	Core
Core	Core	Core	Core
Core	Core	Core	Core
Core	Core	Core	Core

Core Available for Application Processing

Software Defined Everything

Creates Bottlenecks



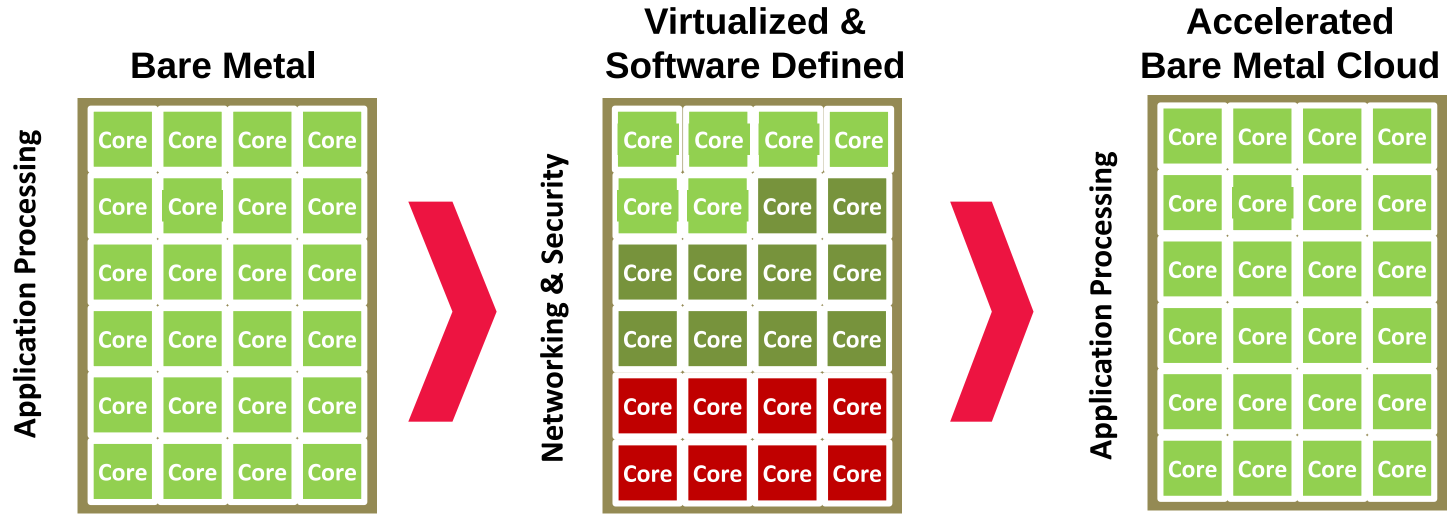
Core Available for Application Processing

Core Software Defined Everything (SDX) Consumes CPU cores for Packet Processing
- Virtualization, Storage, Switching, Routing, Load Balancing

Core Security: Consumes CPU cores for Security Processing
- Layer 4 Firewall, encryption, host introspection
- Intrusion detection & prevention

Software Defined Everything

Creates Bottlenecks



Core Available for Application Processing

Core Software Defined Everything (SDX) Consumes CPU cores for Packet Processing

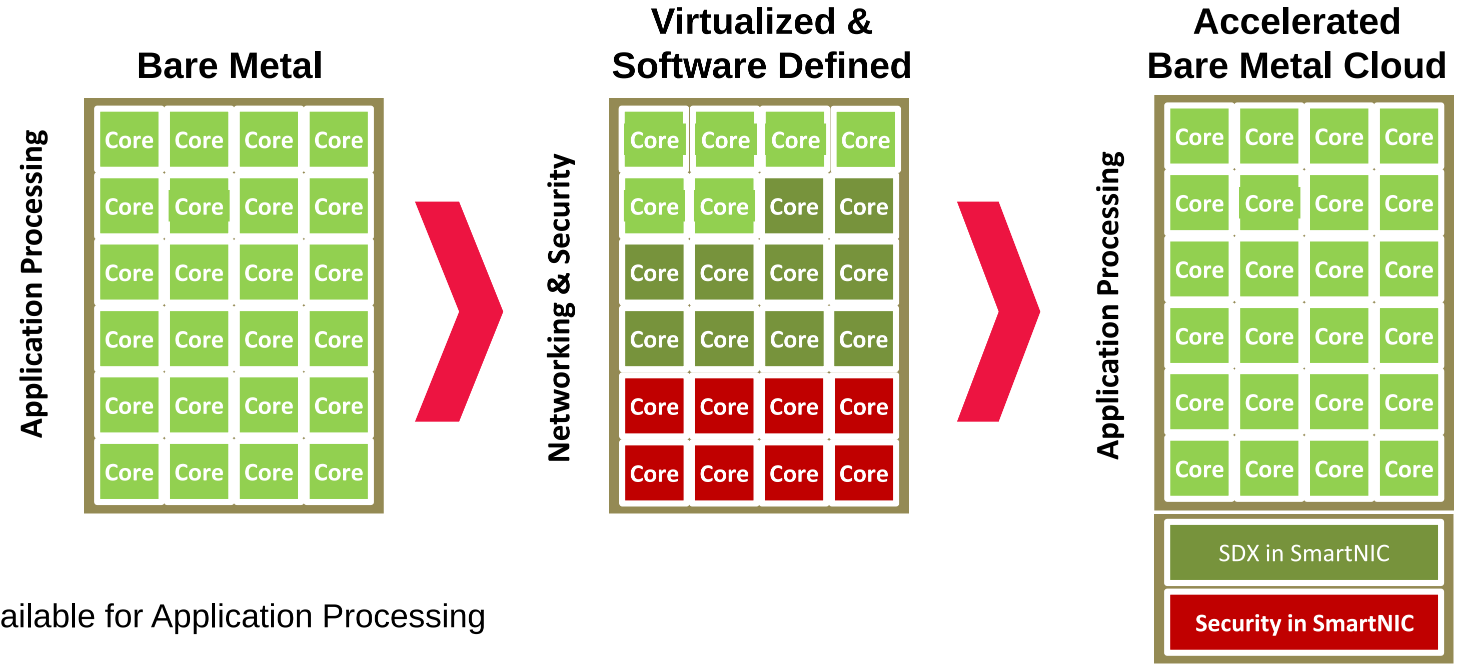
- Virtualization, Storage, Switching, Routing, Load Balancing

Core Security: Consumes CPU cores for Security Processing

- Layer 4 Firewall, encryption, host introspection
- Intrusion detection & prevention

Software Defined Everything

Creates Bottlenecks



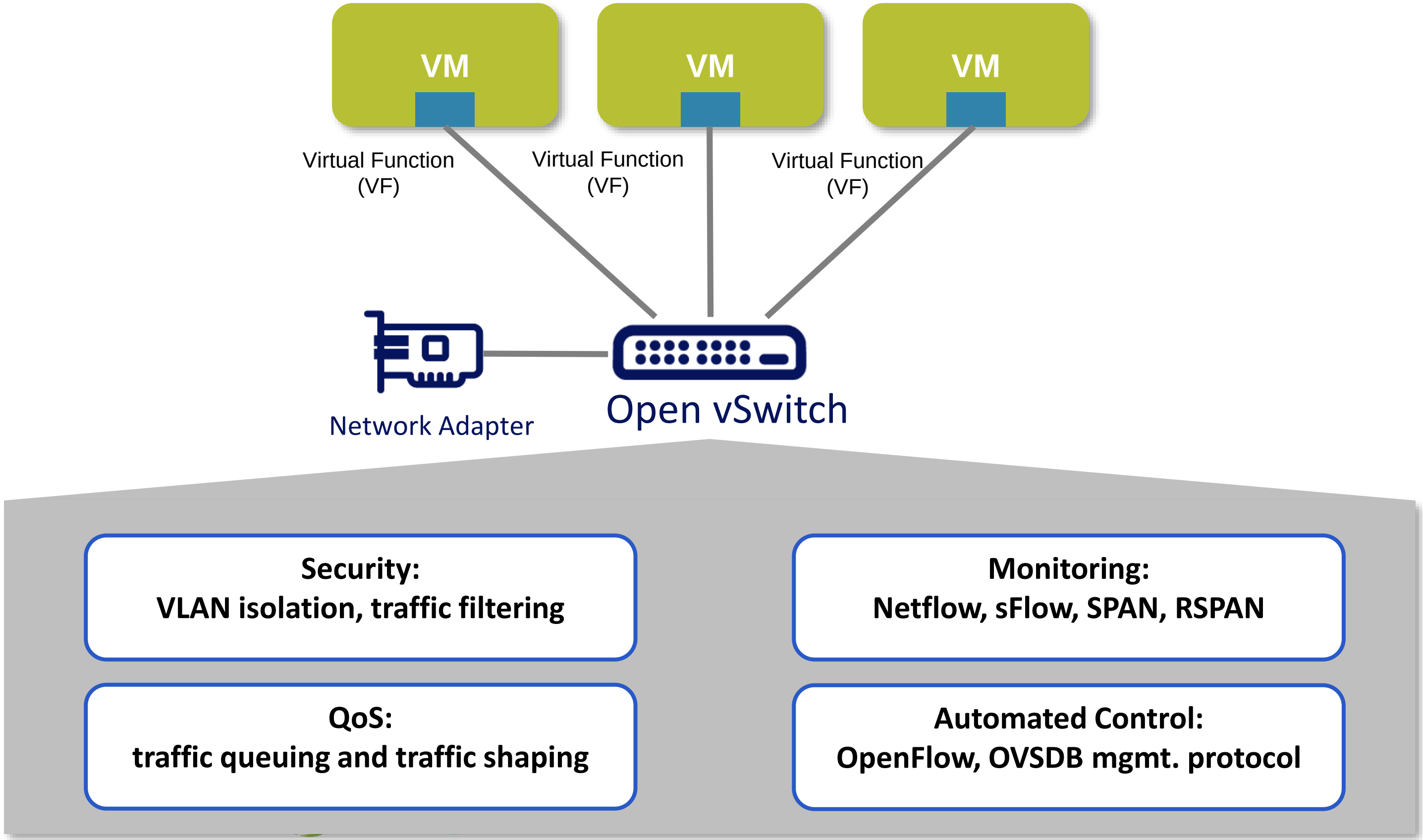
🟩 Core Available for Application Processing

🟩 Core Software Defined Everything (SDX) Consumes CPU cores for Packet Processing
- Virtualization, Storage, Switching, Routing, Load Balancing

🟥 Core Security: Consumes CPU cores for Security Processing
- Layer 4 Firewall, encryption, host introspection
- Intrusion detection & prevention

Open vSwitch (OVS)

Cloud ready virtual switch



Security:
VLAN isolation, traffic filtering

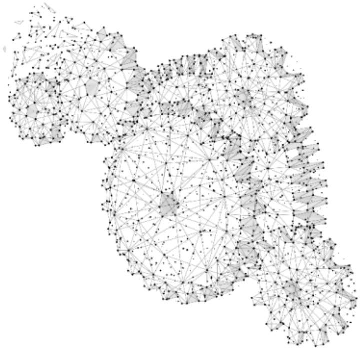
Monitoring:
Netflow, sFlow, SPAN, RSPAN

QoS:
traffic queuing and traffic shaping

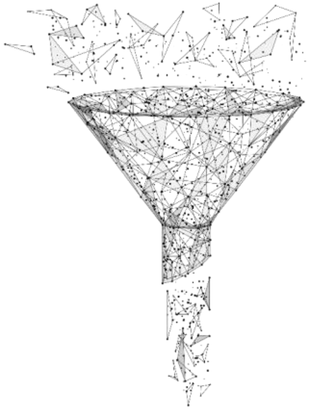
Automated Control:
OpenFlow, OVSDB mgmt. protocol

OVS Performance Challenges

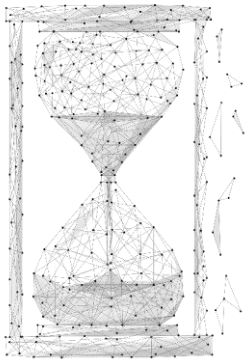
OVS performance burdens:



High CPU Utilization

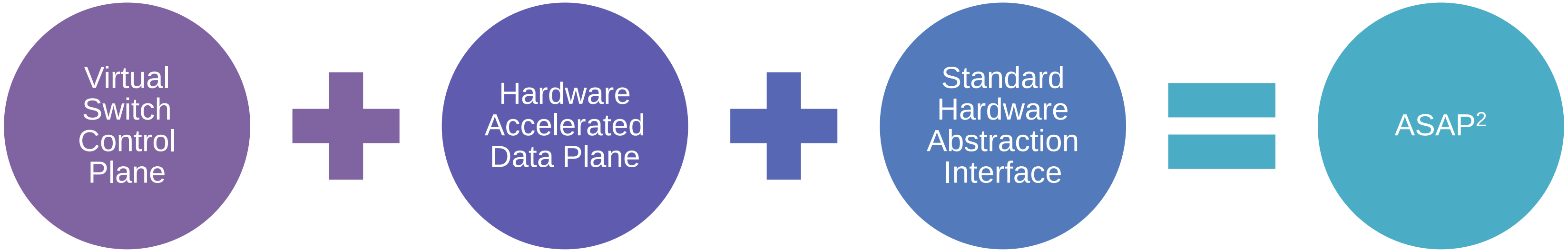


Limited Throughput



Higher Latency

Accelerated Switching and Packet Processing ASAP^2


Virtual Switch Control Plane
+
Hardware Accelerated Data Plane
+
Standard Hardware Abstraction Interface
=
ASAP2

Best of both worlds:

Hardware Accelerated Data Plane

+

Software Define Control Plane

Software vs. Hardware OVS

Legacy

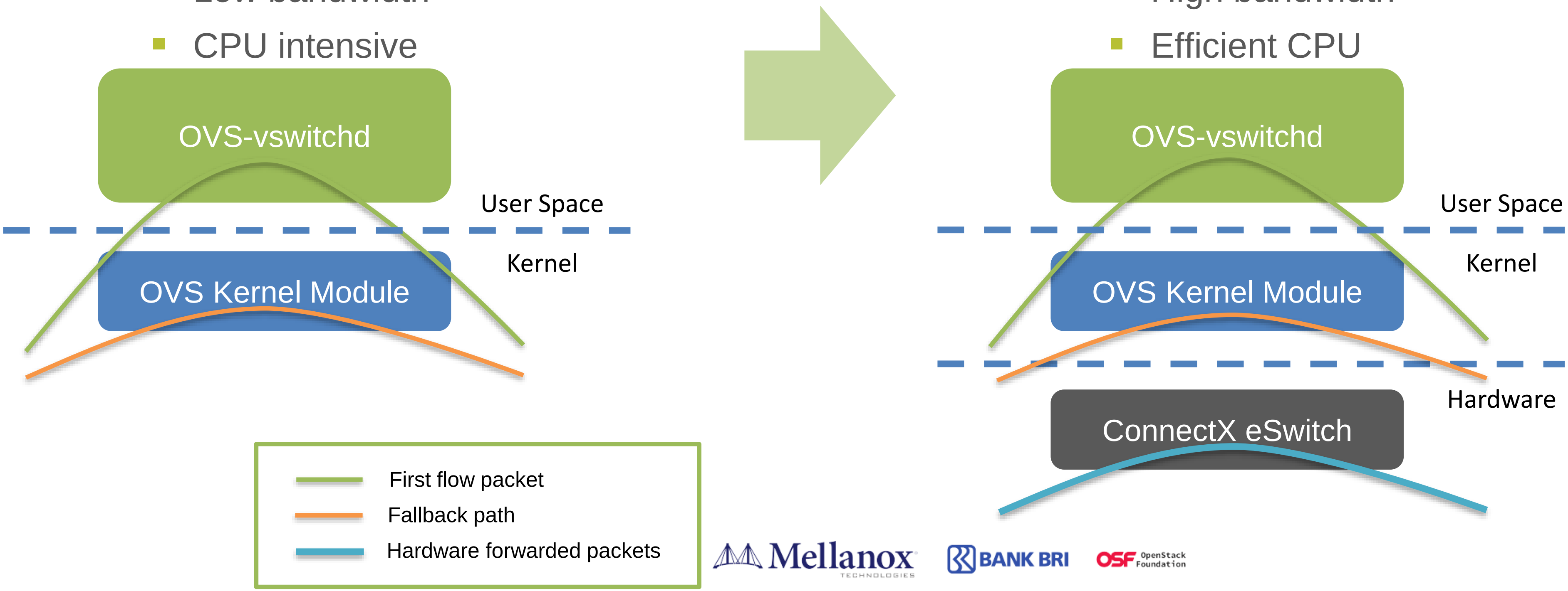
OVS Software Implementation

- High latency
- Low bandwidth
- CPU intensive

Mellanox TECHNOLOGIES

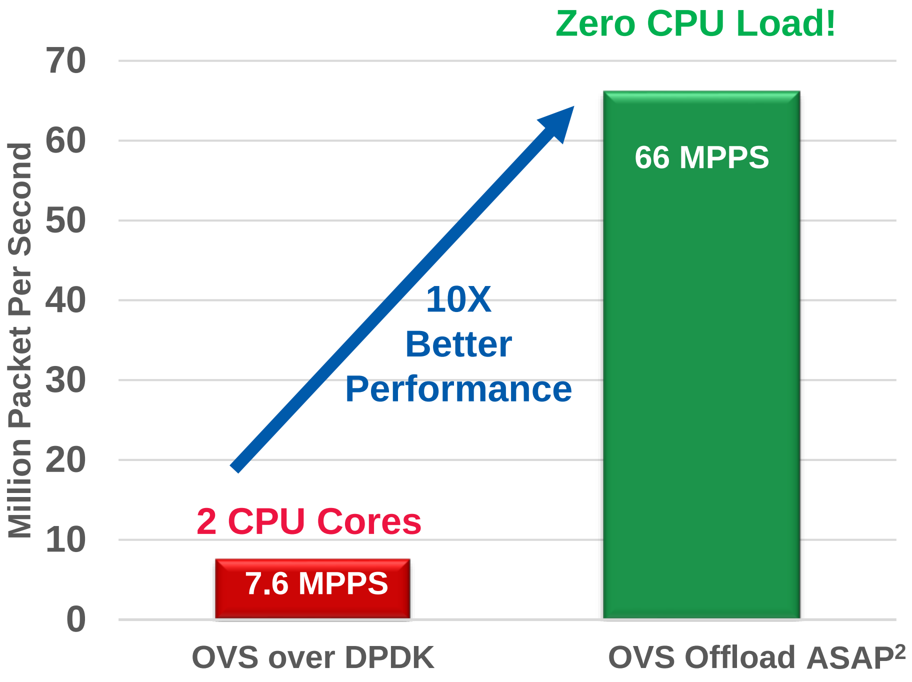
ASAP^2 on ConnectX Hardware

- Low latency
- High bandwidth
- Efficient CPU



Improving OVS Performance

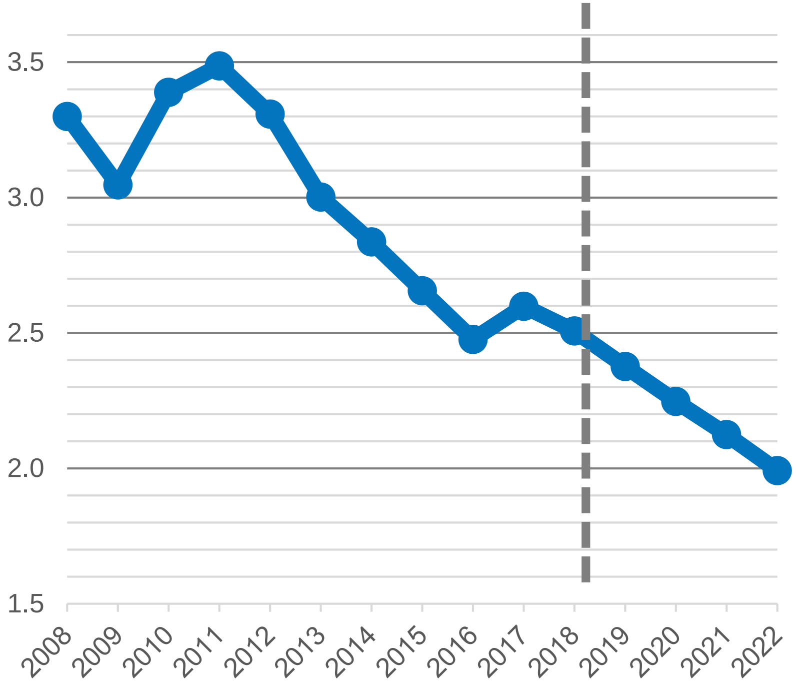
- **Mellanox OVS Offload - ASAP^2**
  - **20X** higher performance than vanilla OVS
  - **10X** better performance than OVS-DPDK
  - **Line rate performance** at 25/40/50/100Gbps



ASAP^2: 10X packet rate with Zero CPU Load

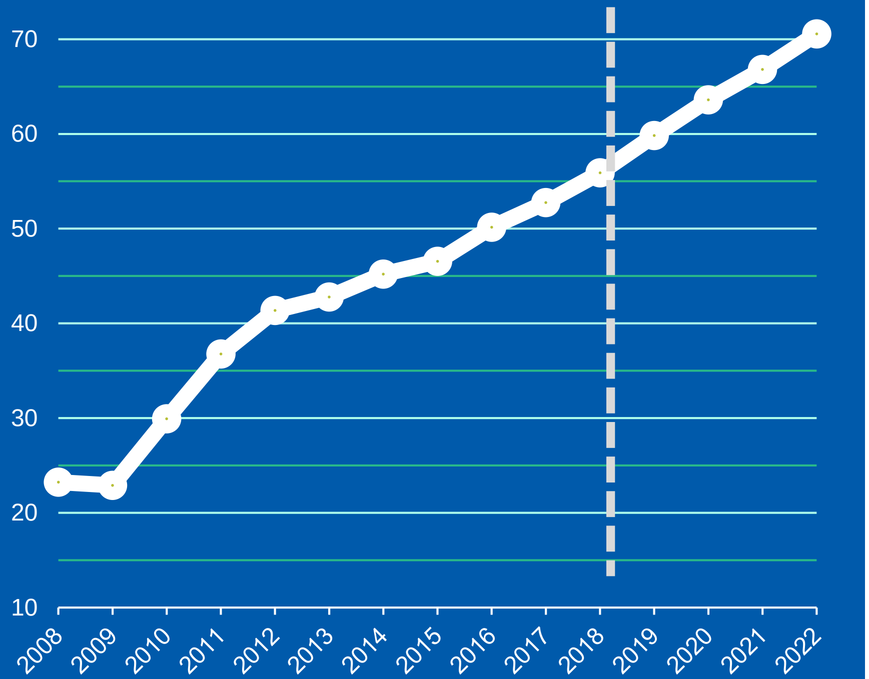
Storage Networking Trend

Fibre Channel Port Shipments (in Millions)



Source: Crehan Research, Host Adapter Port Shipments, January 2018

Ethernet Port Shipments (in Millions)



Storage Networking Trend

1997

Feature	Fibre Channel	Ethernet
Bandwidth	1 G	100 M
Supports	Block	Block, file
Lossless	Yes	No
Cost	High $$$$	Medium $$
Cloud / HCI	No / No	No / No
Vendors	Several	Many
SDS / Scale-out	No / No	No / No

2019

Feature	Fibre Channel	Ethernet
Bandwidth	8/16/32 G	10/25/40/100 G
Supports	Block	Block, file, object
Lossless	Yes	Yes
Cost	Medium $$	Low $
Cloud / HCI	No / No	Yes / Yes
Vendors	2 / 2	Many / Many
SDS / Scale-out	Rare / No	Yes / Yes

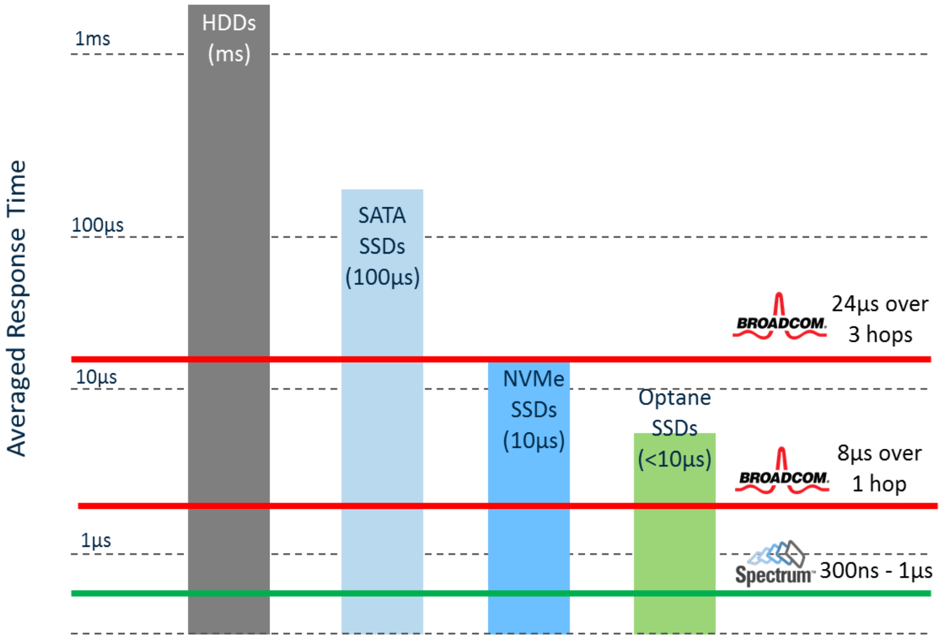
Yesterday: Storage Network = FC

- Fibre Channel offered best performance
- All interesting storage was tier-1 block
- No cloud or hyperconverged

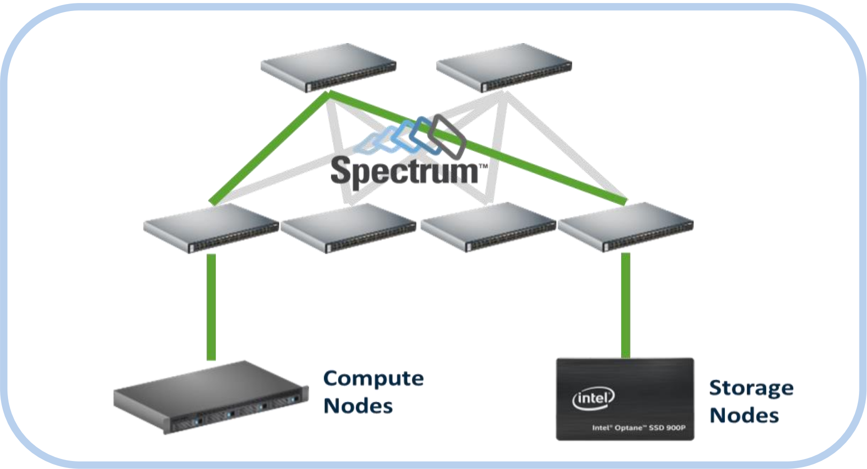
Today: Ethernet for storage networks

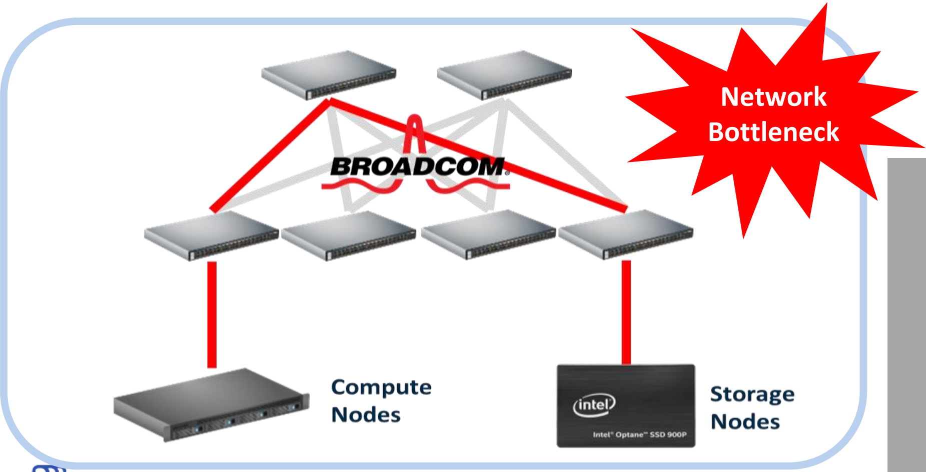
Web-Scale Innovation:

Matching Network & Storage Latency



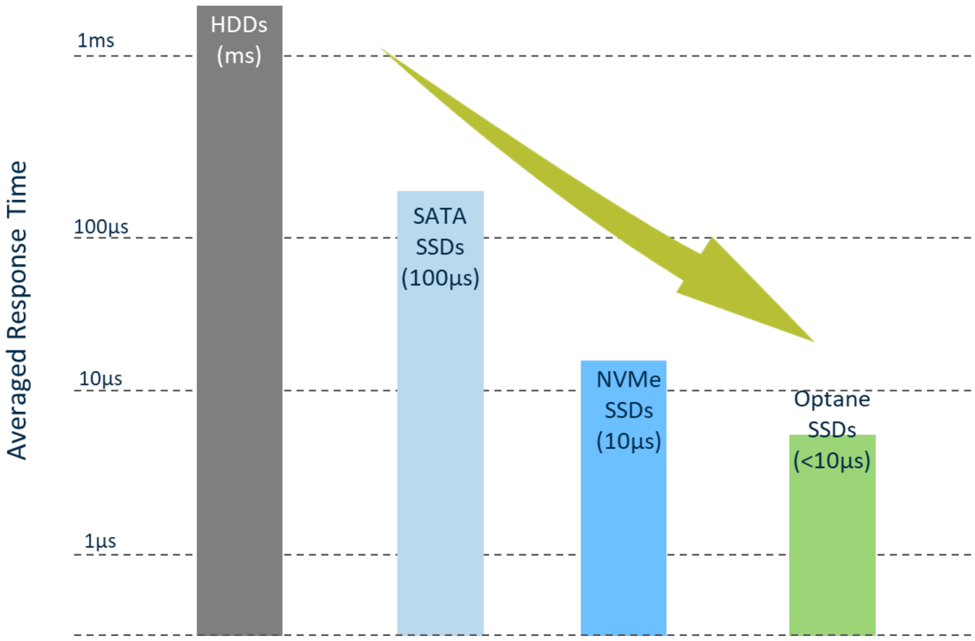
Flash Storage is Getting a Lot Faster!



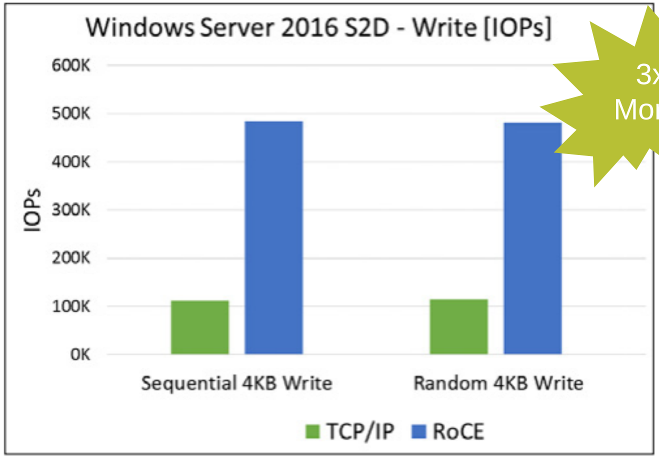


Web-Scale Innovation:

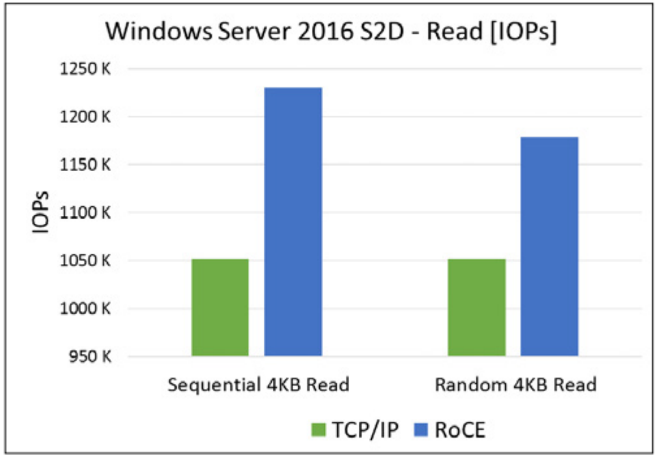
Accelerate Scale-out Storage with ROCE



Storage is Getting Faster!



3x More!



Source: Outstanding S2D Efficiency

Web-Scale Innovation:

Storage Optimized Switch Form Factors

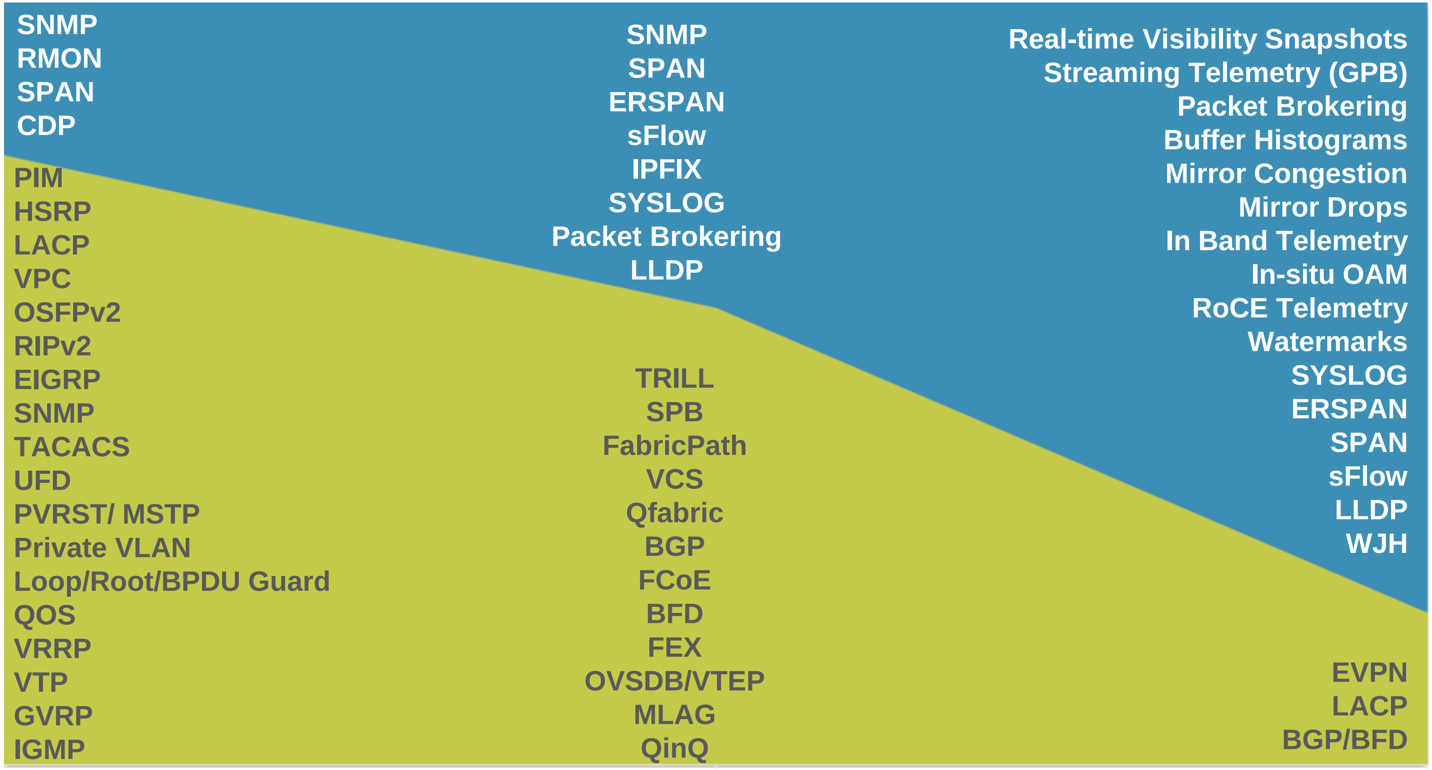
- Performance
- High Availability
- Simplicity
- Automated
- Scalability
- Cost Efficient

Spectrum®

- ✓ 2 Switches in 1RU
- ✓ Ideal Port Count for Storage /HCI / ML
- ✓ Zero Packet Loss
- ✓ Low Latency
- ✓ RoCE optimized
- ✓ Network automation/visibility
- ✓ Cost optimized

Web-Scale Innovation:

Measure Everything!



SNMP
RMON
SPAN
CDP

PIM
HSRP
LACP
VPC
OSFPv2
RIPv2
EIGRP
SNMP
TACACS
UFD
PVRST/ MSTP
Private VLAN
Loop/Root/BPDU Guard
QOS
VRRP
VTP
GVRP
IGMP

SNMP
SPAN
ERSPAN
sFlow
IPFIX
SYSLOG
Packet Brokering
LLDP

TRILL
SPB
FabricPath
VCS
Qfabric
BGP
FCoE
BFD
FEX
OVSDB/VTEP
MLAG
QinQ

Real-time Visibility Snapshots
Streaming Telemetry (GPB)
Packet Brokering
Buffer Histograms
Mirror Congestion
Mirror Drops
In Band Telemetry
In-situ OAM
RoCE Telemetry
Watermarks
SYSLOG
ERSPAN
SPAN
sFlow
LLDP
WJH

EVPN
LACP
BGP/BFD

Legacy Mindset **Webscale Mindset**

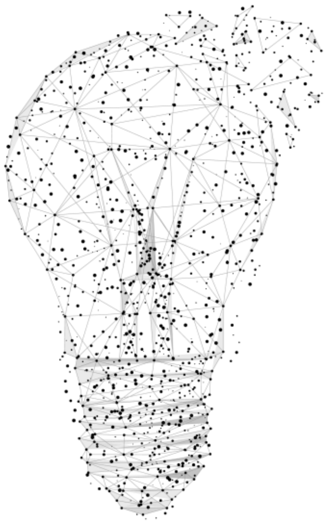
🟩 Protocols 🟦 Telemetry Features

Biznet GioCloud Biznet Mellanox TECHNOLOGIES BANK BRI OSF OpenStack Foundation

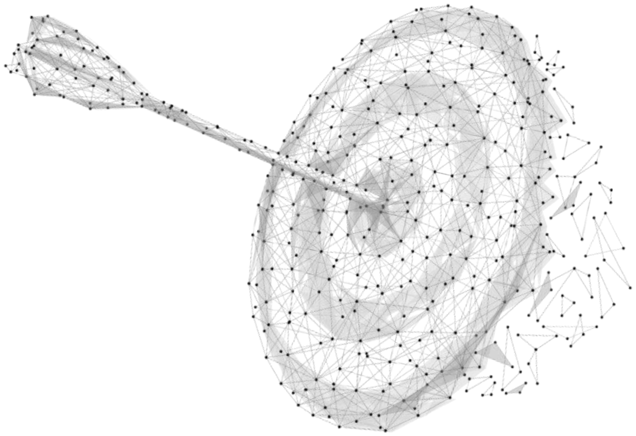
Why Do We Need Telemetry?

Faster Time to Innocence

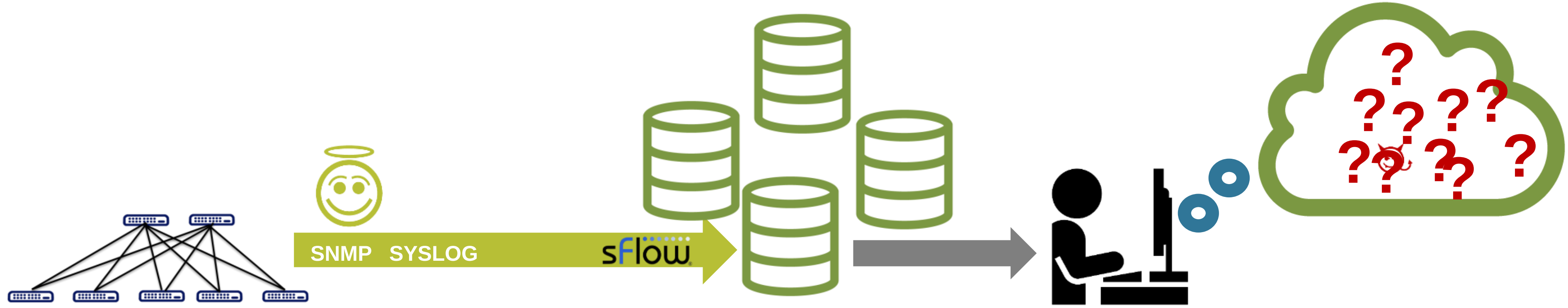
Faster Time To Resolution

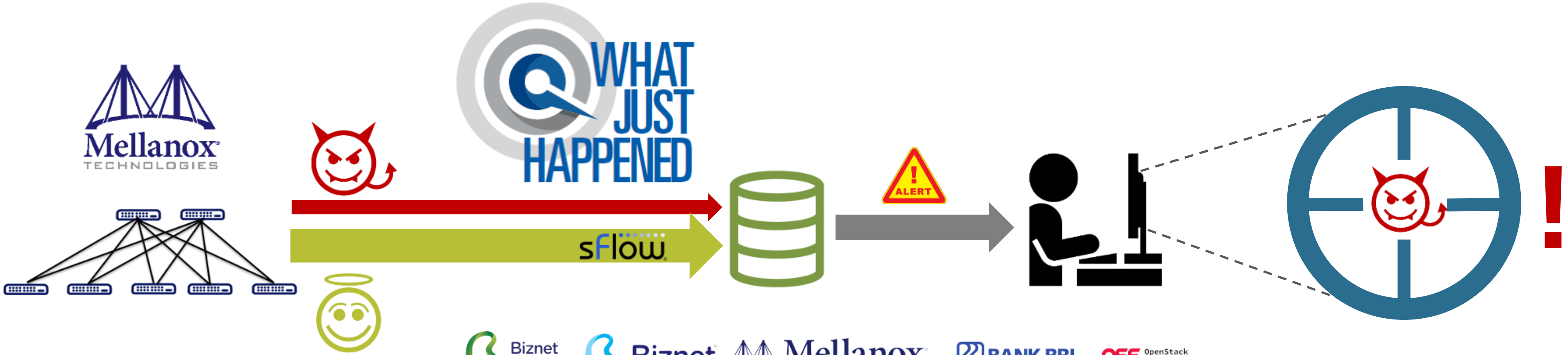


Get more out of the Network



WJH™ Accelerates the Time to Root-Cause



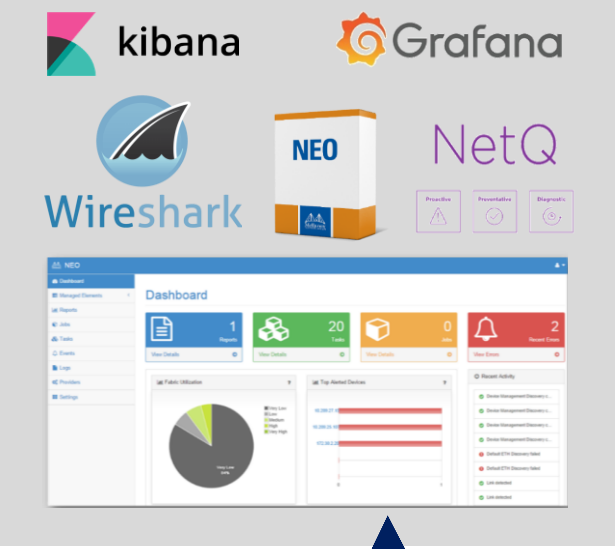


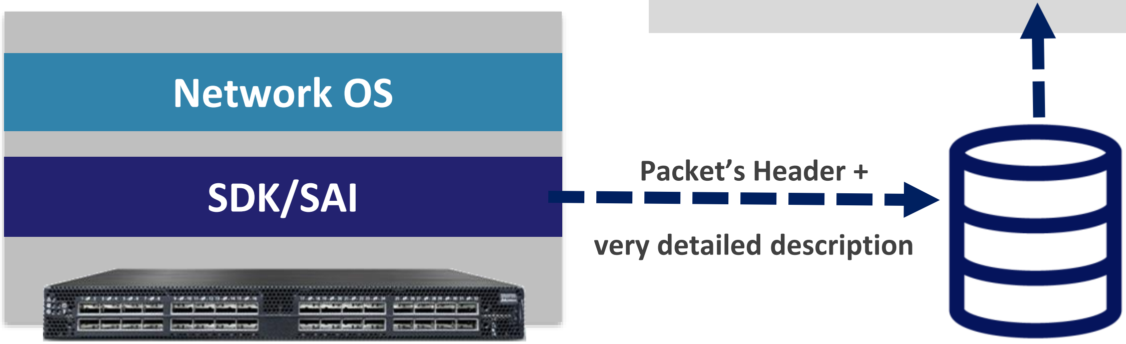
WJH™ – How Does It Work?

1. SDK generates:
WJH messages

2. WJH Agent:
Streams to a Database

3. Presentation layer shows:
What Just Happened

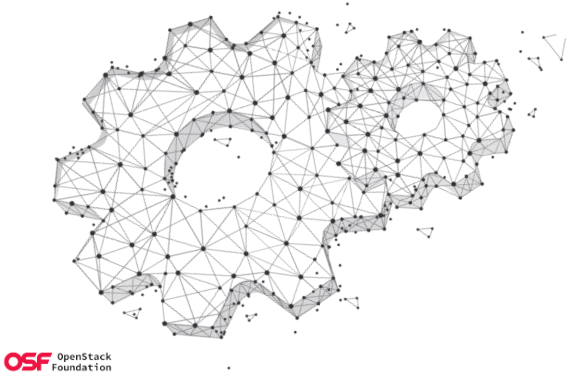




Packet's Header +

very detailed description

The Important Questions	
WHO	is being impacted
WHAT	is causing the problem
WHEN	it happened
WHERE	is the problem
WHY	it is happening



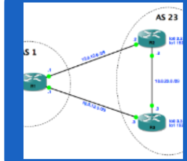
Web-Scale Innovation:

Extreme Visibility

Packet Drop

L1

- Bad CRC
- Flaky cable



L2/L3/Overlay

- BGP
- VLAN

Buffer

- Incast
- Rate Limit

ACLs

- Deny based on IP
- Deny based on VLAN

No Packet Drop

Congestion

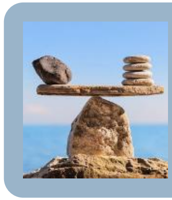
- Incast
- Busy storage device

Latency

- Pause frames
- Congestion→ latency

Suboptimal Route

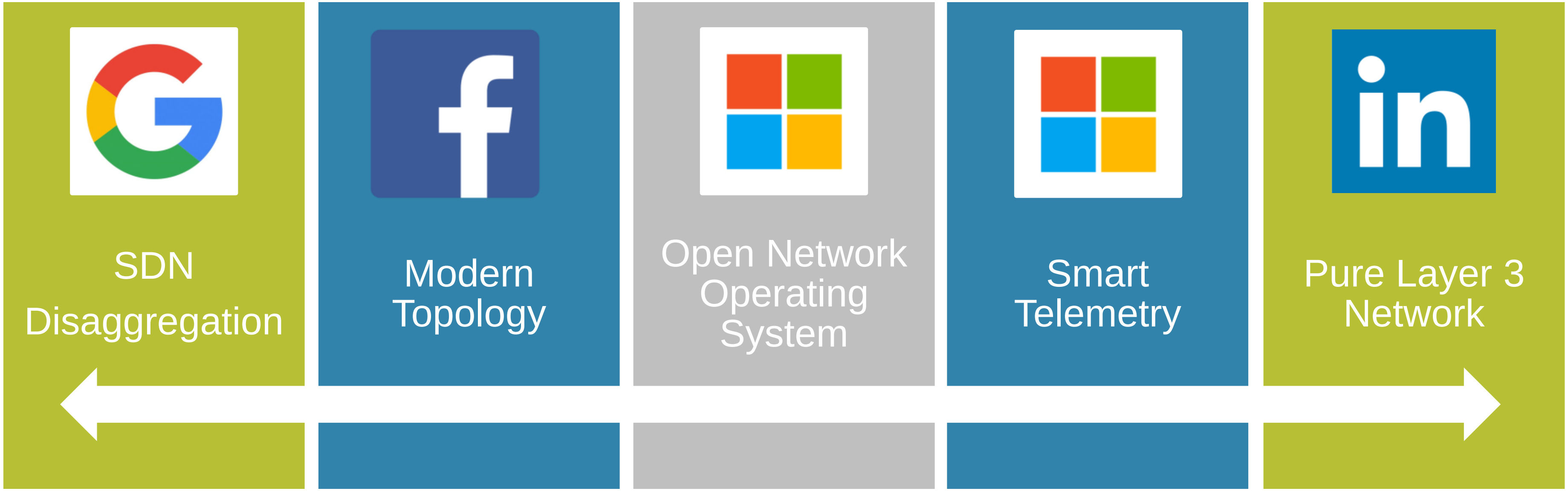
- Packet doesn't reach the firewall
- Packet go through a sub-optimized path



Suboptimal Load Balance

- Suboptimal ECMP
- Suboptimal LAG

What have Cloud Titans taught the Industry?



SDN Disaggregation

Modern Topology

Open Network Operating System

Smart Telemetry

Pure Layer 3 Network

We bring Cloud Titan innovations to you!

Biznet GioCloud Biznet Mellanox TECHNOLOGIES BANK BRI OSF OpenStack Foundation

INDONESIA
OpenInfra Days

Thank you!

Biznet GioCloud Biznet Mellanox TECHNOLOGIES BANK BRI OSF OpenStack Foundation