



Benefits and Use Cases for NVMe-oF SmartNICs

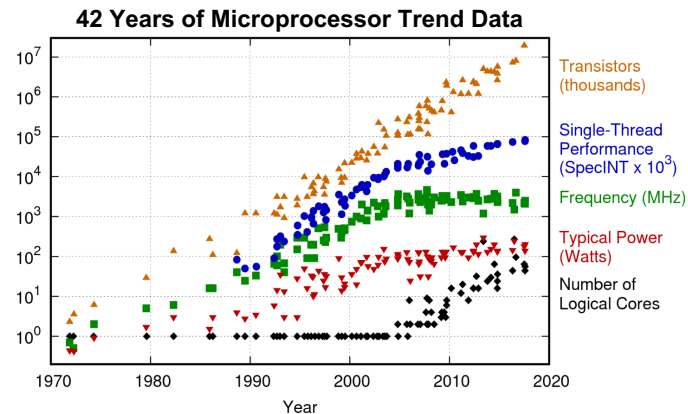
Fazil Osman

Distinguished Engineer

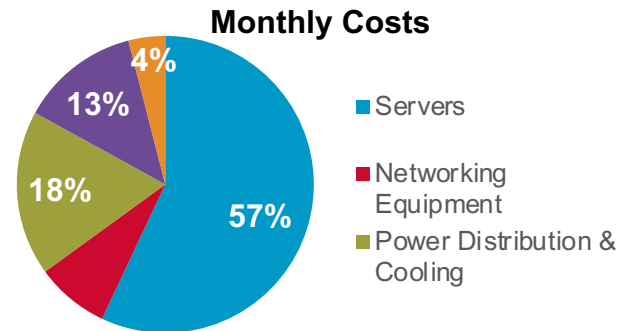
Broadcom

Why a SmartNIC

- **Moore's Law diminishing returns**
 - Vertical scaling power & cost model no longer viable
- **CPU costs increasing**
 - Economic benefits to limiting core count
- **Multi-socket interconnect bottleneck**
 - I/O, memory transactions across interfaces add latency
 - 2nd socket often used to get more memory and I/O
 - TCO penalty for 2nd socket
- **Distributed cloud architecture**
 - Smaller fault domains



Original data up tot the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010 – 2017 by K. Rupp



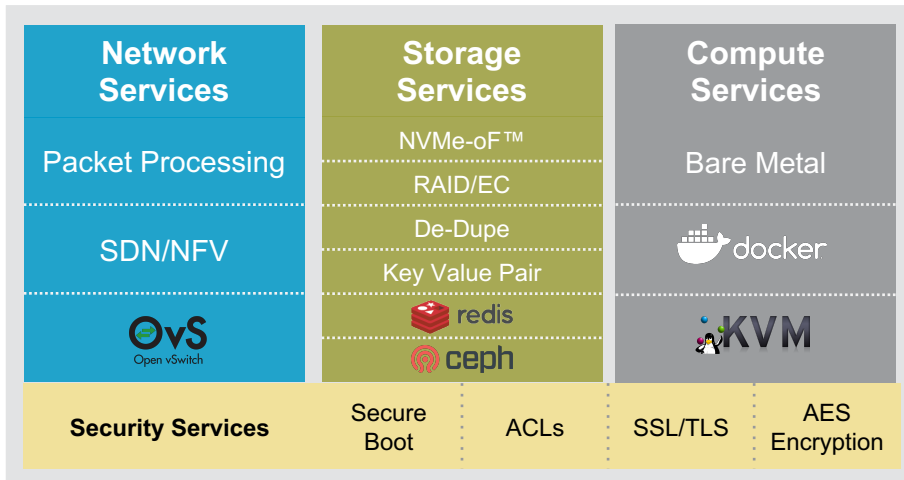
3yr server & 10yr infrastructure amortization

Source: James Hamilton, AWS



What is SmartNIC

Architectural flexibility to quickly offload multiple overhead IaaS services

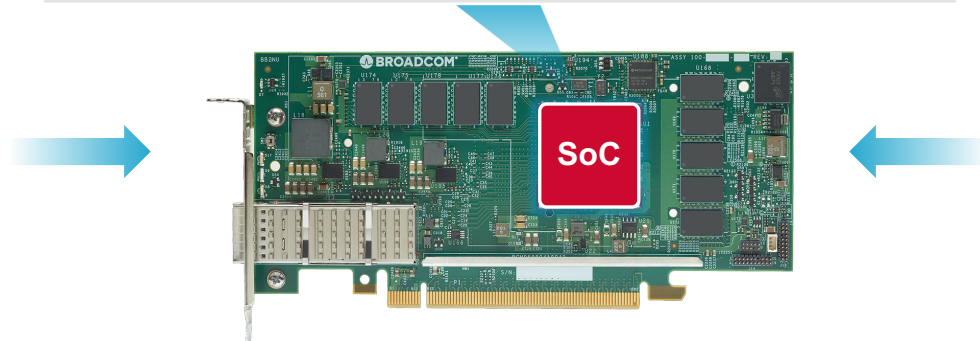


Onload 
Hardware Appliances...

- Firewall
- IDS/IPS
- SD-WAN
- Router
- ADC
- vTAP
- Packet Broker

...Offload 
SDS, SDN, NFV Services

- NVMe-oF
- RAID/EC
- KV Store
- IPSec/SSL/TLS
- vSwitch
- vRouter
- NFV VNFs





Evolution of SmartNIC...



FPGA + NIC

Pros

- Typical single function offload
- Good performance

Cons

- Hard to design for performance
- Slow feature velocity (RTL)
- High power
- Large devices are expensive



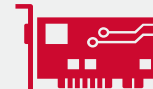
Network Function Processor

Pros

- More than single function

Cons

- Non-standard programming
- Can be expensive
- High power



SmartNIC

Pros

- Performance/Watt
- General-purpose with standard programming
- Great feature velocity

Cons

- Performance varies based on CPUs, DDR, and availability of integrated accelerators

HFT, HPC, Telco I/O

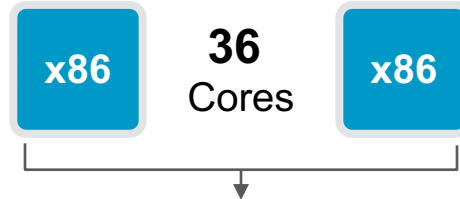
Telco I/O

Cloud DC & Telco

Platform Economics: CPU Workload Partitioning

~ \$8,000 Platform
Including Southbridge
and High-Performance NIC

Example
(165W, 18C)



Example
(165W, 18C)

~ 380W Platform
Including Southbridge
and High-Performance NIC

<p><50% Utilization Typical with Virtualization*</p>	<p>~18 Cores Often oversubscribed for memory</p>	<p>18 Cores Remaining</p>
<p>Storage Services</p>	<p>AES Encryption NVMe-oF™</p>	<p>~4-6 Cores Consumed</p>
<p>Networking Services</p>	<p>Open vSwitch contrail systems VNF</p>	<p>~2-4 Cores Consumed</p>
<p>Only 8-12 Cores Available for Applications</p>		

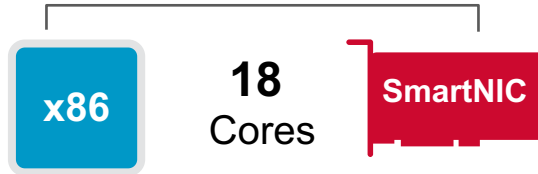
Services can consume most the remaining cores



Platform Economics: SmartNIC Workload Partitioning



~ \$4,000 Platform

Including Southbridge and High-Performance NIC Built Into SmartNIC



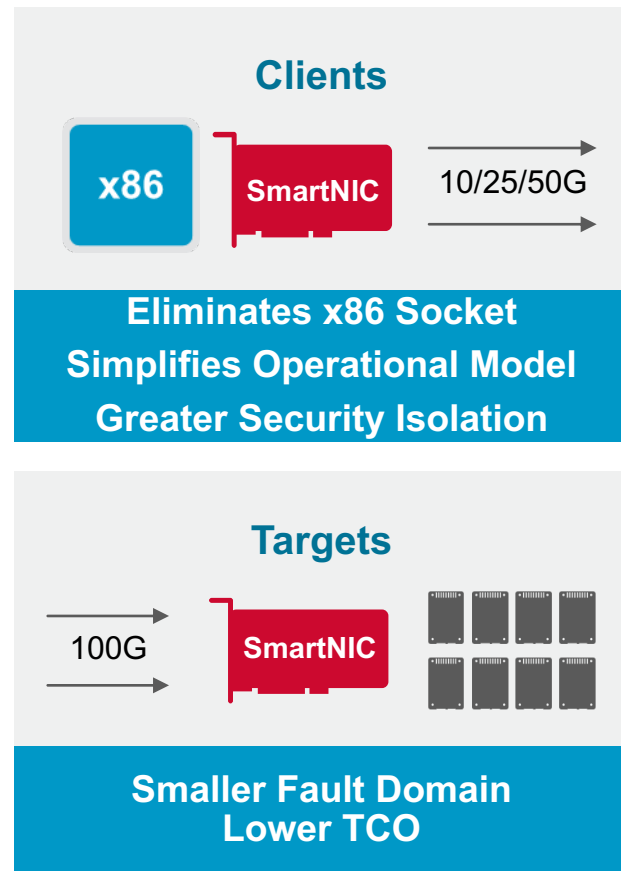
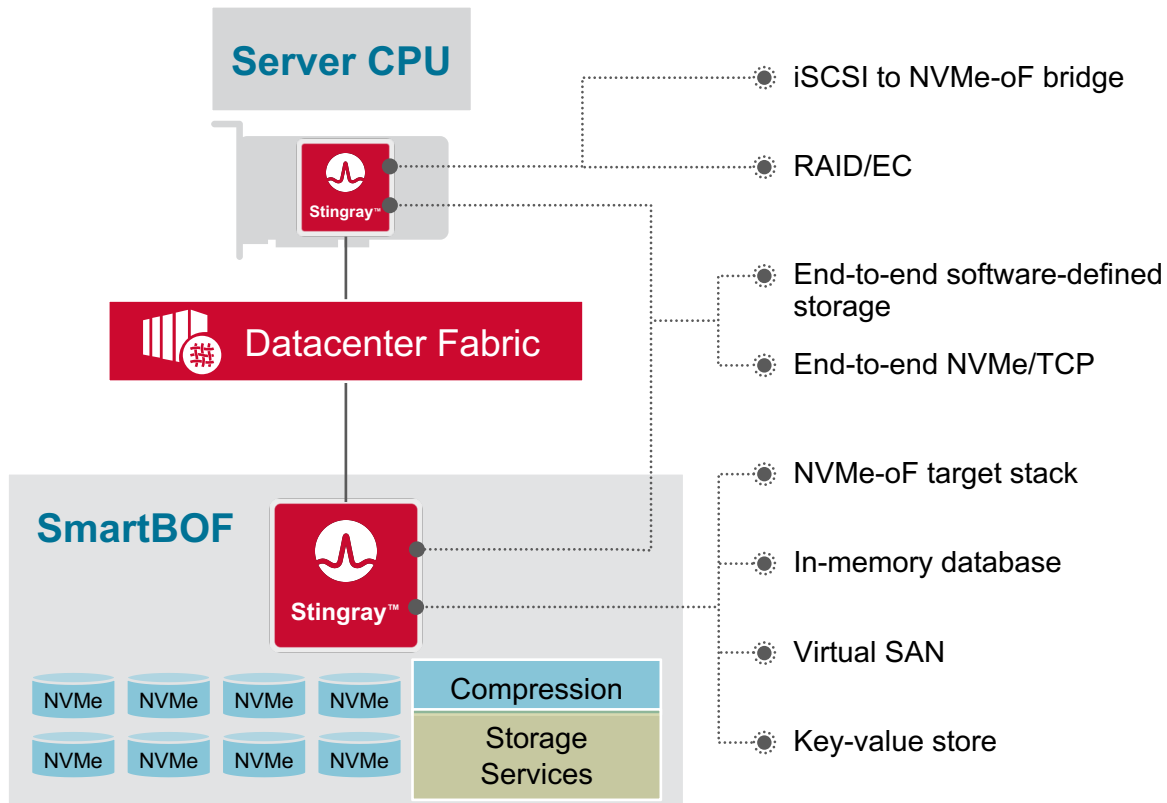
~ 200W Platform

Including Southbridge and High-Performance NIC Built into SmartNIC

<50% Utilization Typical with Virtualization*	Minimal virtualization overhead	16-18 Cores Remaining
Storage Services	<div style="display: flex; justify-content: space-around;"> <div style="background-color: #d9ead3; padding: 5px;">AES Encryption</div> <div style="background-color: #d9ead3; padding: 5px;">NVMe-oF™</div> </div>	Run on SmartNIC
Networking Services	<div style="display: flex; justify-content: space-around; align-items: center;">   <div style="background-color: #d9ead3; padding: 5px; border: 1px solid #ccc;">VNF</div> </div>	Run on SmartNIC
16-18 Cores Available for Applications		

Offloading services to SmartNICs frees up cores for applications

SmartNIC Storage Use Cases





Example: Small vs. Large Fault Domains

Test Summary

Parameter	4x Stingray Targets	2 Socket-x86 Target
Network Link	4x 25G	1x 100G
NVMe SSDs (x2 Gen3)	32	30
4K Random Read	2.0M IOPS 🏆	1.8M IOPS
512K Sequential Write	37K IOPS 🏆	18K IOPS
Tail Latency (mean – P90% – P99.9%)	2 ms – 6.2 ms – 11 ms 🏆	2.3 ms – 12.9 ms – 23.5 ms
CPU+DRAM Power (estimated)	160W 🏆	300W



SmartNIC Disaggregated Storage Advantages

- Better performance
- Lower power
- Smaller fault domain reduces blast radius exposure (16TB vs 60TB)





SmartNIC in NVMe-oF™ – We Have Come A Long Way but...

Ecosystem

OS



OS Support for NVMe-oF

- Limited to recent versions of Linux
- No announced support for other operating systems



Industry



NVMe-oF™ July 2016



University of New Hampshire
InterOperability
Laboratory

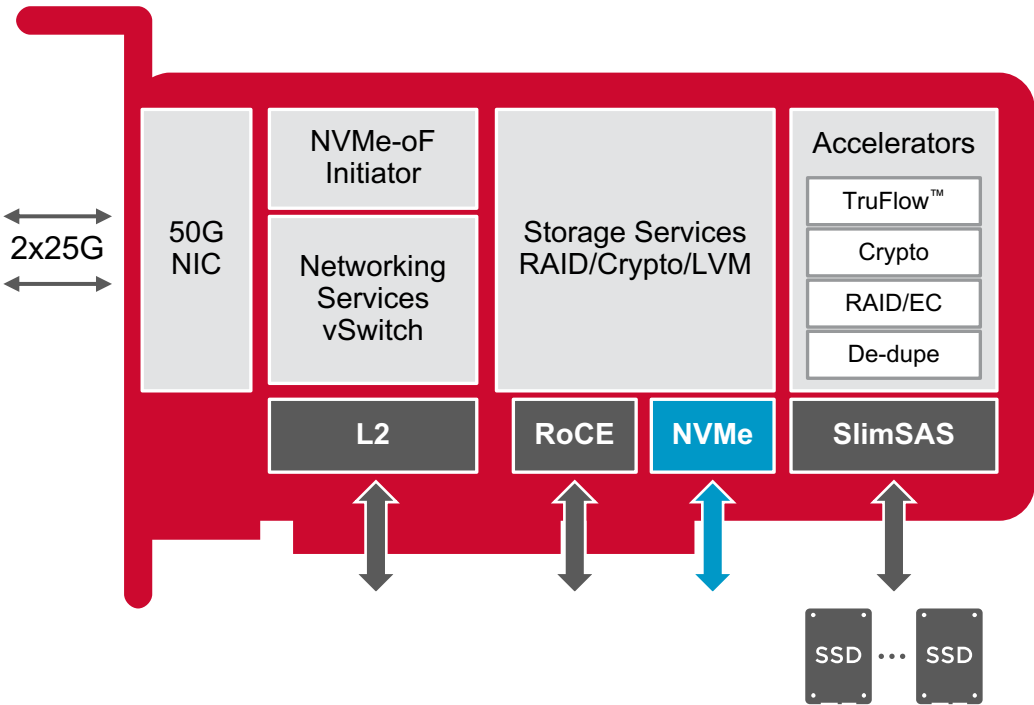


NVMe/TCP™ Ratified Nov 2018

Ecosystem is maturing but broad adoption requires solution to OS support problem

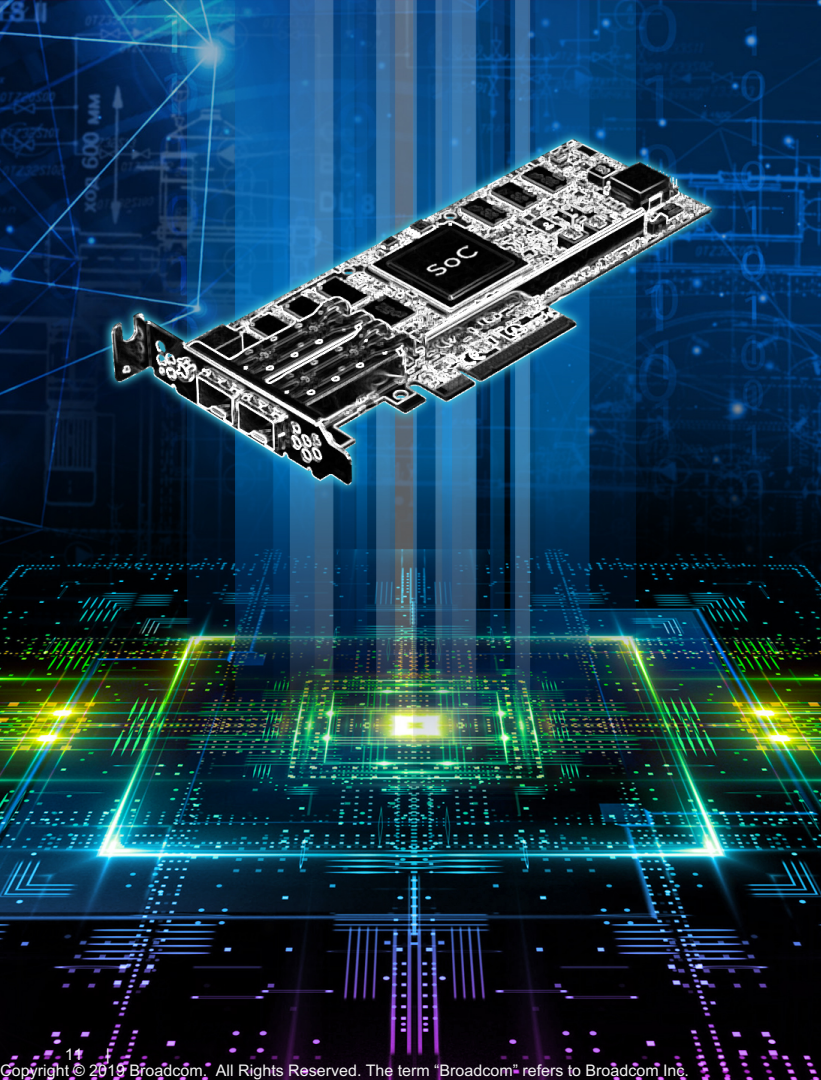


Broadcom Glass Creek Adapter



Applications

- Storage disaggregation for any OS
- Works with standard NVMe drivers
- Storage virtualization
 - Bare metal and virtualized servers
- Storage services offload
 - Logical Volume Management
 - RAID/EC, De-dupe, Crypto



Summary

- **Why SmartNIC**

- System architecture, cost and performance
- Dual socket architectures are inefficient
- End of Moore's Law

- **Market adoption of SmartNIC**

- Highly programmable
- CPU-based
- More flexible architecture

- **Use cases**

- Offloading storage and networking services
- Software-defined storage
- Security

- **Expanding ecosystem**

- Multiple vendor support
- NVMe-oF and **new** NVMe virtualization