



September 23-26, 2019
Santa Clara, CA

Using SmartNICs as New Platform for Storage Services

Fazil Osman
Broadcom Corporation



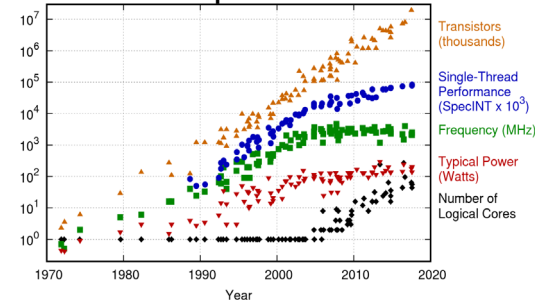
Why a SmartNIC

April 26, 2019
Santa Clara, CA

SDC¹⁹

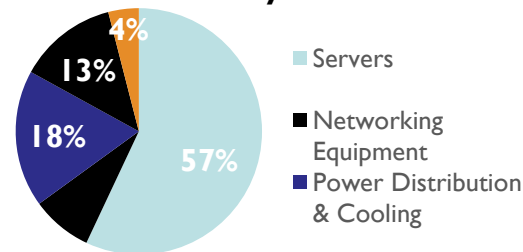
- **Moore's Law diminishing returns**
 - Vertical scaling power & cost model no longer viable
- **CPU costs increasing**
 - Economic benefits to limiting core count
- **Multi-socket interconnect bottleneck**
 - I/O, memory transactions across interfaces add latency
 - 2nd socket often used to get more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

42 Years of Microprocessor Trend Data



Original data up to 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

Monthly Costs



3yr server & 10yr infrastructure amortization

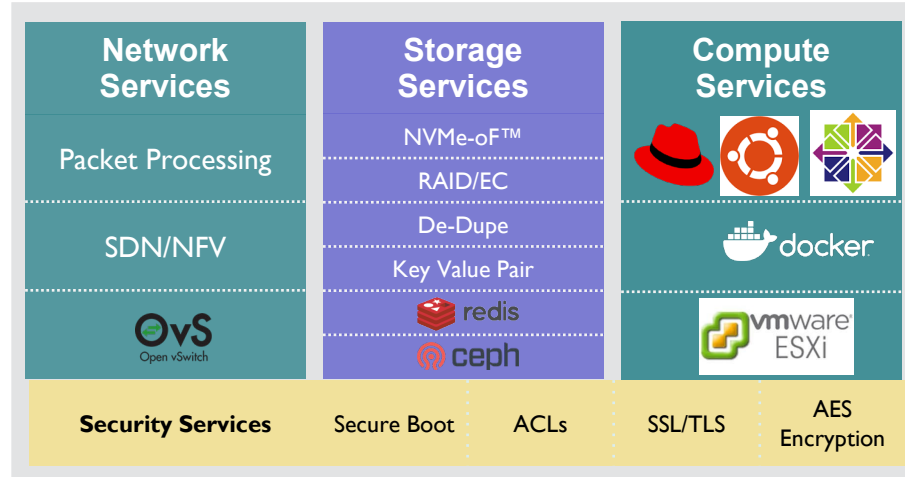
Source: James Hamilton, AWS

What is SmartNIC

September 23-26, 2019
Santa Clara, CA

SDC¹⁹

Architectural flexibility to quickly offload multiple overhead IaaS services

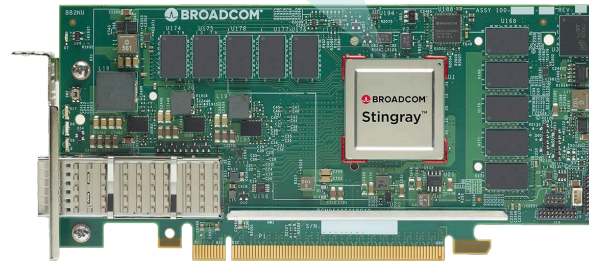


Onload 
Hardware Appliances...

- 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...

September 23-26, 2019

San Jose, CA



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



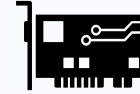
Network Function Processor

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

September 23-26, 2019
Santa Clara, CA

SDC¹⁹

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

Example
(165W, 18C)

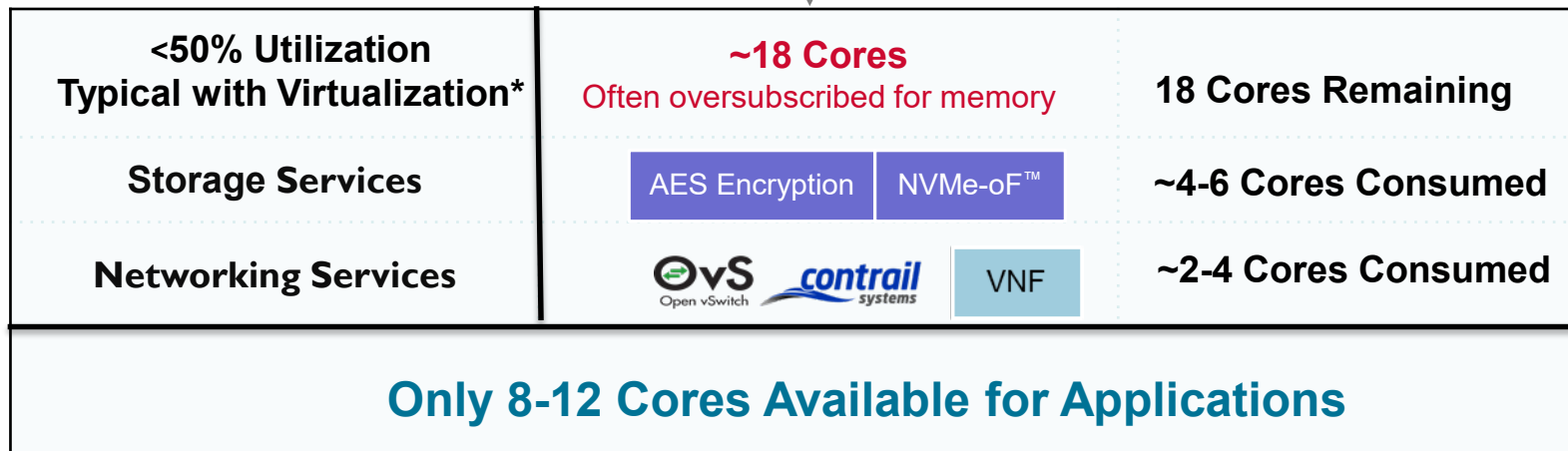


**36
Cores**



Example
(165W, 18C)

~ 380W Platform
Including Southbridge
and High-Performance NIC



Services can consume most of the remaining cores

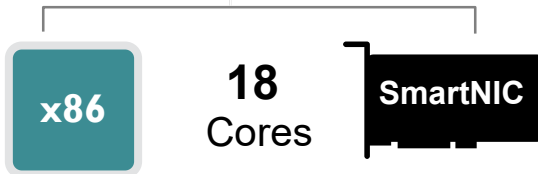
Platform Economics: SmartNIC Workload Partitioning

SDC¹⁹

September 23-26, 2019
Santa Clara, CA

~ \$4,000 Platform

Including Southbridge
and High-Performance NIC
Built Into SmartNIC



~ 200W Platform

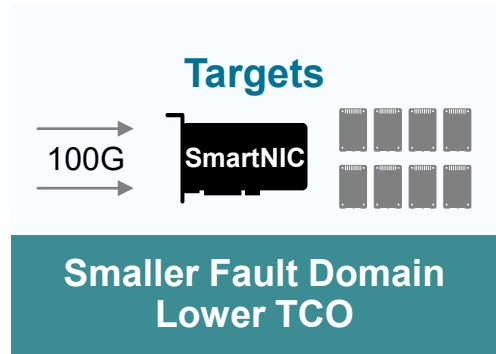
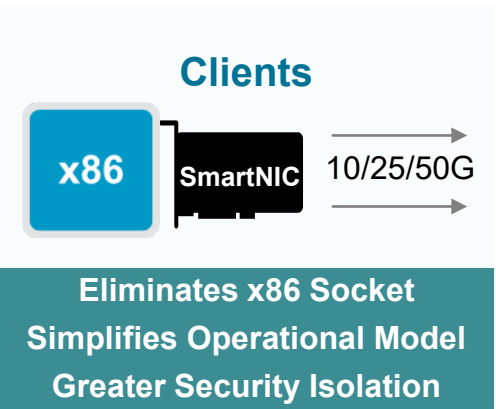
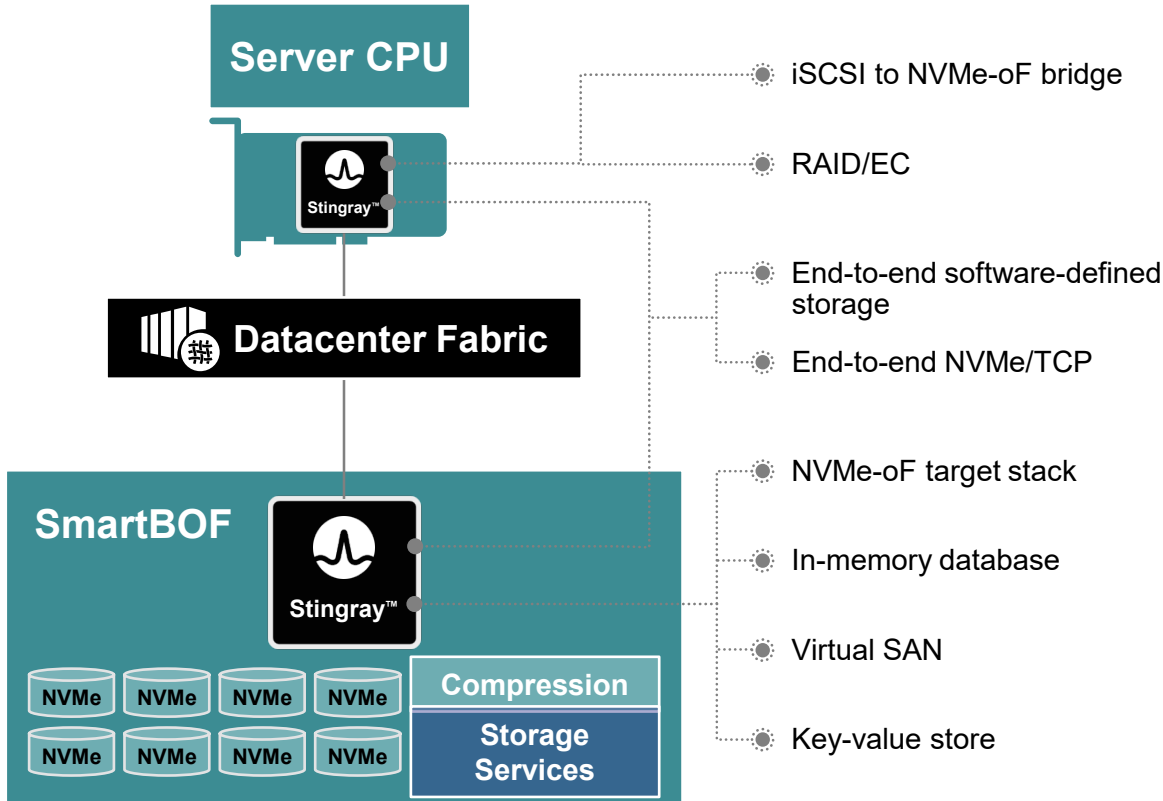
Including Southbridge
and High-Performance NIC
Built into SmartNIC

<50% Utilization Typical with Virtualization*	Minimal Virtualization Overhead	16-18 Cores Remaining
Storage Services	AES Encryption NVMe-oF™	Run on SmartNIC
Networking Services	OvS Open vSwitch contrail systems VNF	Run on SmartNIC
16-18 Cores Available for Applications		

Offloading services to SmartNICs frees up cores for applications

SmartNIC Storage Use Cases

September 23-26, 2019
Santa Clara, CA

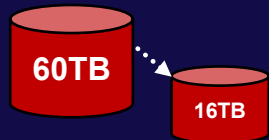


Example: Small vs. Large Fault Domains

NVMe-oF™

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)

Stingray-Based Storage Platforms

September 23-26, 2019
Santa Clara, CA

SDC¹⁹



Celestica Euclid



AIC *Manta*



wistron
Lymma



WDC F3100 and Stingray
Low power
NVMe-oF Enabled
Software-programmable
Low latency
Up to 10 Modules and >20M IOPS

Scale Out

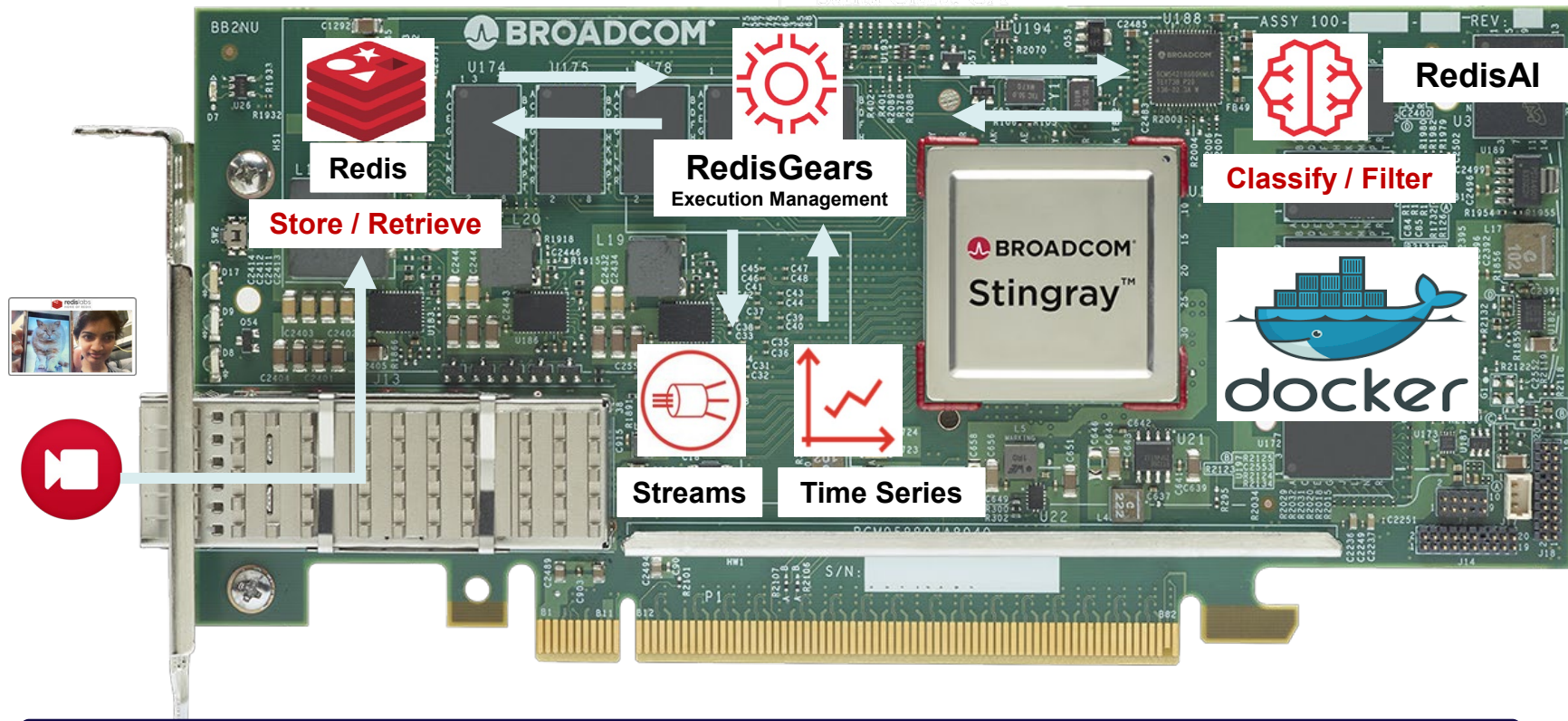
2U-24 Drive Systems
Full HA Support
NVMe-oF Enabled
>10M IOPS

Enterprise Class

RedisEdge on SmartNIC Use Case

September 23-26, 2019
Santa Clara, CA

SDC¹⁹



Deployed within hours using standard container

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

September 23-26, 2019
Santa Clara, CA

SDC¹⁹

Ecosystem

AIC



BROADCOM®



CNEXLABS

Excelero

EXTEN™



英業達
Inventec

kaminario.



SAMSUNG



wistron

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

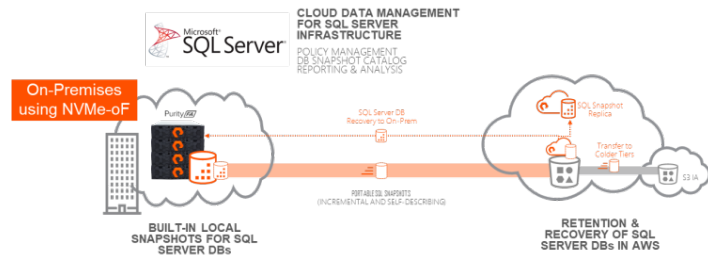
Ecosystem is maturing but broad adoption requires solution to OS support problem

Pure Storage Use Case

Bringing NVMe over Fabrics to Windows client solutions

SDC¹⁹

SQL SERVER DB SNAPSHOT PORTABILITY WITH CLOUDSNAP



- ✓ Simple, built-in local DB snapshot technology
- ✓ Offload local SQL Server DB snapshots to the cloud
- ✓ Run your Dev/Test environments in the cloud
- ✓ Business Continuity for your SQL Server infrastructure across clouds

Windows SQL server is largest application with Pure customers

STINGRAY ENABLING NVME-OF™ FOR PURE FLASHARRAY//X

iSCSI to NVMe-oF Mapping



x86

iSCSI



iSCSI to NVMe Bridging

NVMe-oF™

Glass Creek Virtualization

x86

Any OS



Looks Like NVMe Drive

NVMe-oF™

Bare Metal or Virtualized Environments

Problem:

Windows does not support NVMe-oF natively limiting FlashArray scalability

Solution:

Co-developed Windows iSCSI to NVMe-oF on Stingray SmartNIC

Broadcom Glass Creek Adapter

September 25-27, 2019
Santa Clara, CA

Introducing

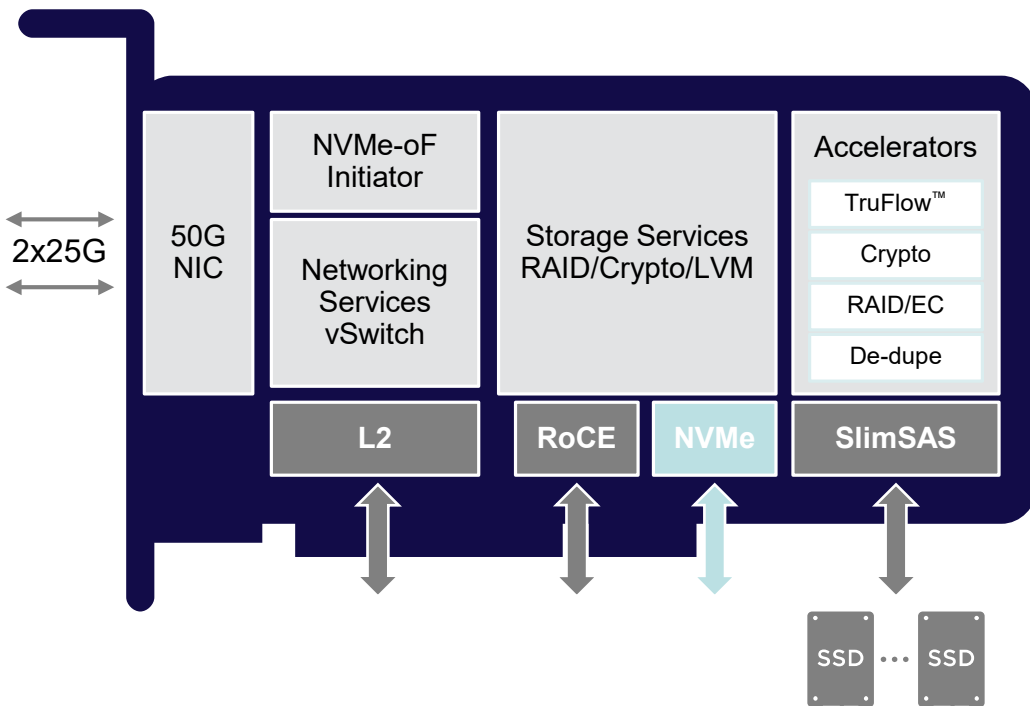
Glass Creek

19

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

13



Stingray Solutions

SDC¹⁹

September 25-26, 2019
Santa Clara, CA

- **NVMe-oF SmartBOFs**
 - High performance fabric storage appliances
 - Enterprise-capable high-availability solutions
 - Cloud scale with low blast radius
- **Expanding NVMe-oF™ ecosystem**
 - Multiple vendors
 - NVMe-oF and **now** NVMe virtualization
- **Highest performance SmartNIC**
 - System architecture, cost and performance
 - End of Moore's Law
 - Dual socket architectures are inefficient
- **SmartNIC Use cases**
 - General-purpose programmability
 - Offloading storage and networking services
 - Bare metal and virtualization servers
 - Security