COMPUTERWORLD SNIA

SNW

# Transforming the Information Infrastructure: Build, Manage, Optimize.

FALL 2011

# Ethernet –
# The Basis for SAN / LAN Convergence in the Data Center

Greg Scherer

VP, Server and Storage Strategy

gscherer@broadcom.com

Broadcom Corporation

# Agenda

- Broadcom Introduction
- Reasons for Convergence in the Data Center
- Key Market Drivers for Convergence
- Technology Enablers for Convergence
- Adapter Capability and Architectural Requirements
- Conclusions and Call to Action

# About Broadcom

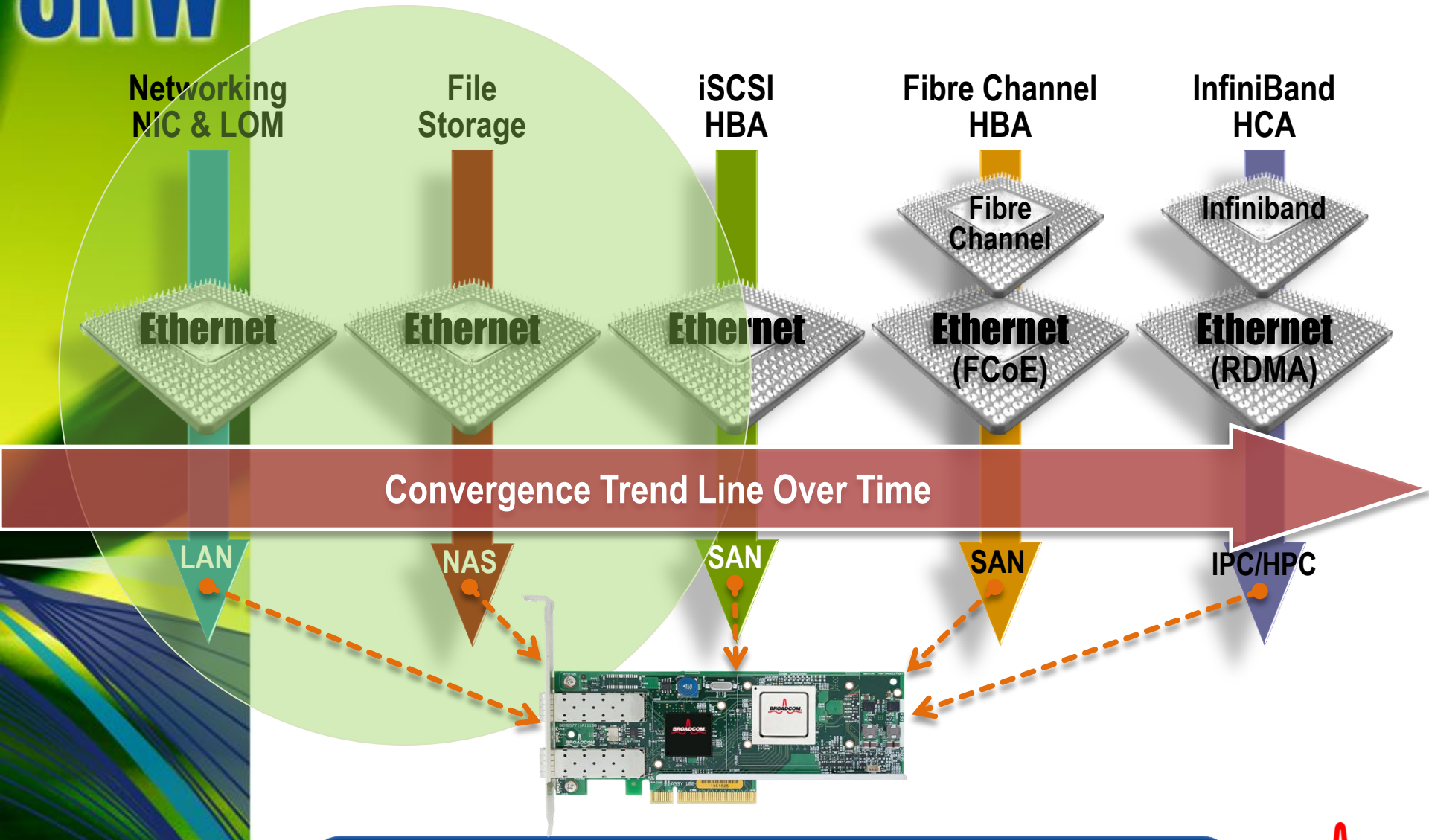
- Global leader in semiconductors for wired and wireless communications
- Founded 1991
- Fortune 500 company
  - Ranked most innovative semiconductor
- 2010 Gartner Top 10 Semiconductor Companies (Revenue)
  - 2010 net revenue of $6.82 billion

- One of the largest volume fabless semiconductor suppliers
- Broad IP portfolio with over 15,000 U.S. and foreign patents and applications
  - Strongest patent portfolio among fabless semiconductor companies (IEEE)
- Approximately 9,052 employees worldwide

Driving convergence through unique & innovative product offerings

# Ethernet: Consolidation Followed by Convergence

## (Multiple Networks vs. one Unified Network)




Consolidation over Ethernet is underway; Convergence is Next!

# Networking and Storage Convergence

## Why?

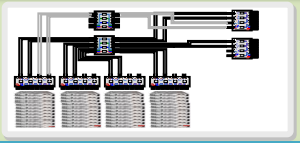
Ethernet: Leading Economics

Ethernet: Leading Technology

### 1. *Ethernet*

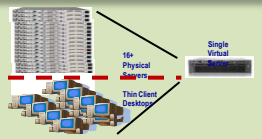
Open, Interoperable, Simple, Low Cost, and High Volume!

### 2. *Unified & Flat Networks*



Significant Opportunity for Lowering CAPEX and OPEX

### 3. *Virtualization*



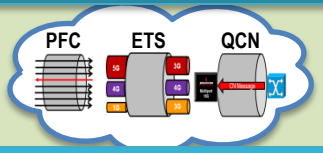

Network Storage is Required for Migration, DR, etc.

### 4. *Scalable Bandwidth*

2/4/8G FC → 10GbE
16/32G FC → 40GbE
64G FC → 100GbE

Clear Path Beyond 10Gb to 40Gb, 100Gb, etc.

### 5. *Continued Innovation*



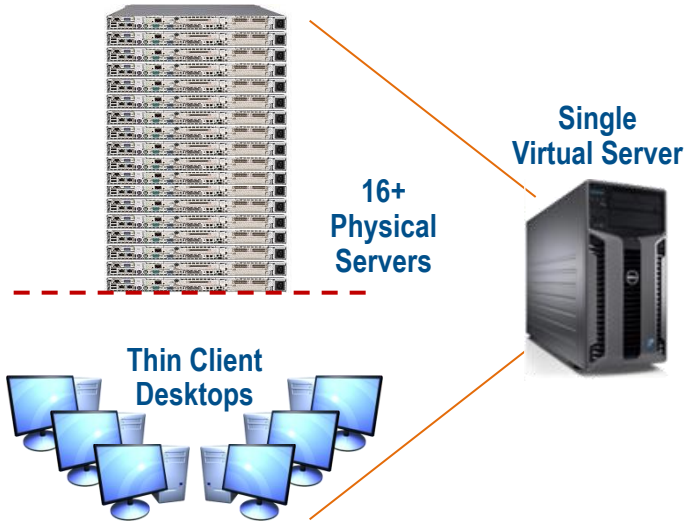

Loss-Less (DCB), Energy Efficient Ethernet (EEE),
Congestion Mgmt (802.1Qau) , Advanced Routing (TRILL), etc.

Ethernet : Technology leadership fueled by economics

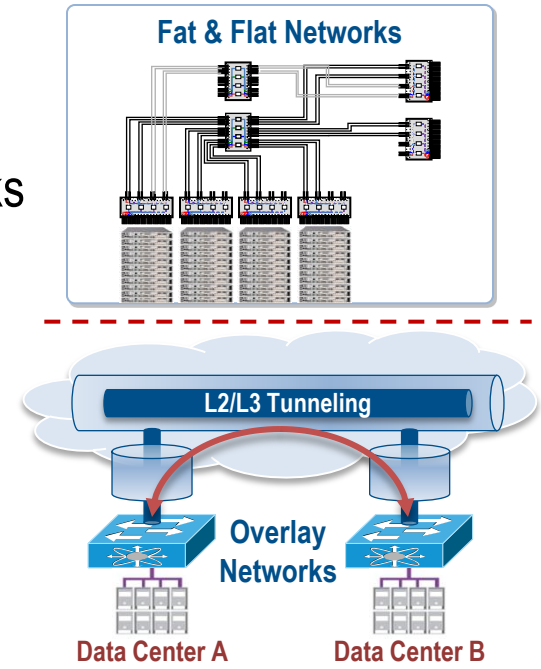
BROADCOM
Connecting everything

# Key Trends Driving Convergence

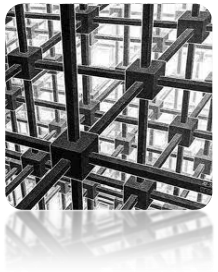
- **Server Virtualization**
  - Server consolidation
  - Desktop virtualization




- **Network Virtualization**
  - Fat & Flat L2/L3 networks
    - Driven by public Cloud
  - Overlay networks



# Cloud Computing Characterization



"Cloud" – a metaphor for on-demand IT "In the Network"

Not a technology but an *architectural practice*

Implemented with a suite of technologies

*Distributed computing environment of arbitrary scale*

Where hardware and software are abstracted from services

## Essential Cloud Attributes

### *Leading Economics:*

- Extreme efficiencies – do more with less
- Utility metering – pay as you go
- Utilization – capacity/demand impendence
- Business service entry cost

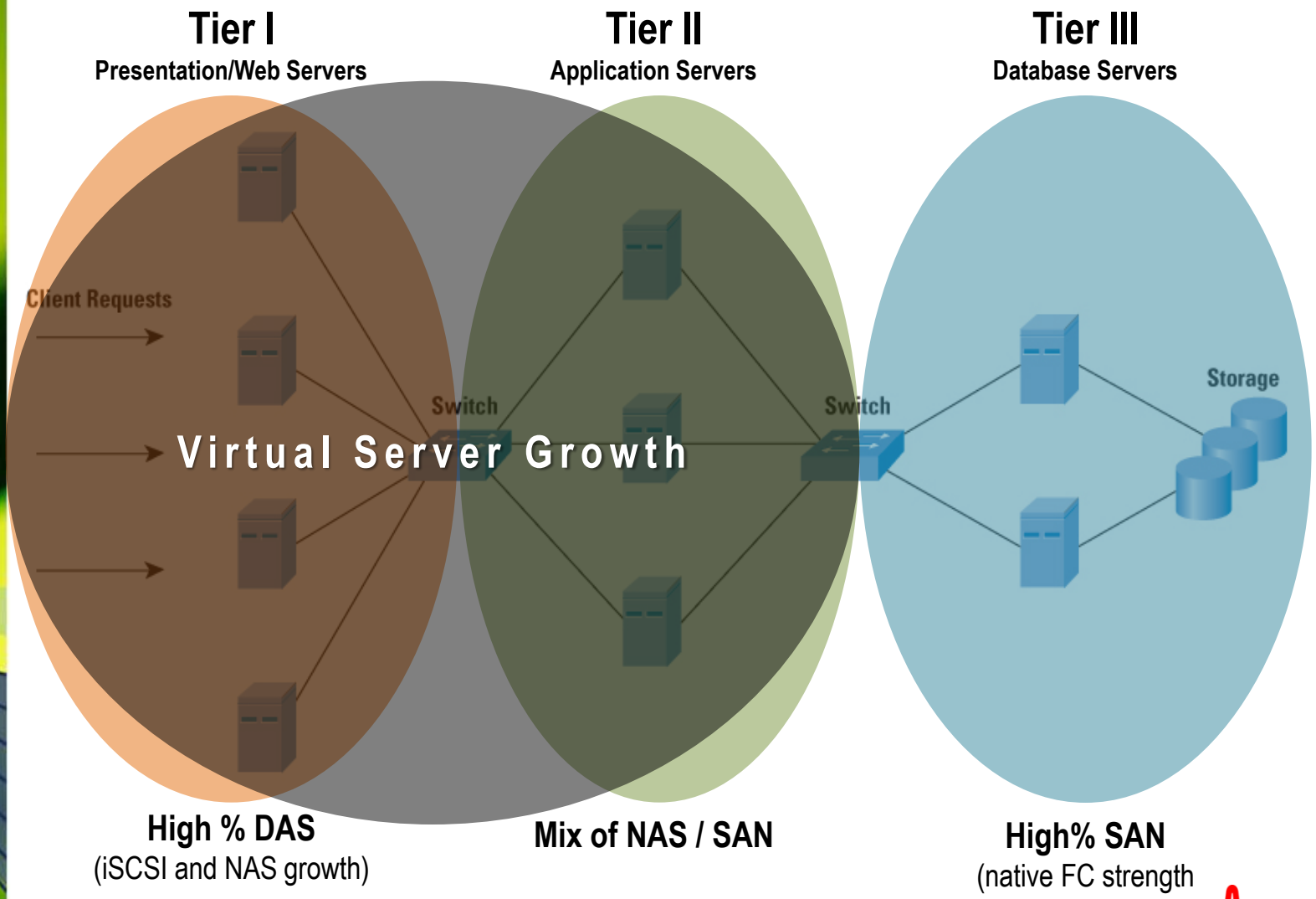
### *Flexibility:*

- Any server can be used for any application
- Application time-to-market
- No lock in – Standard interfaces
- Programmable

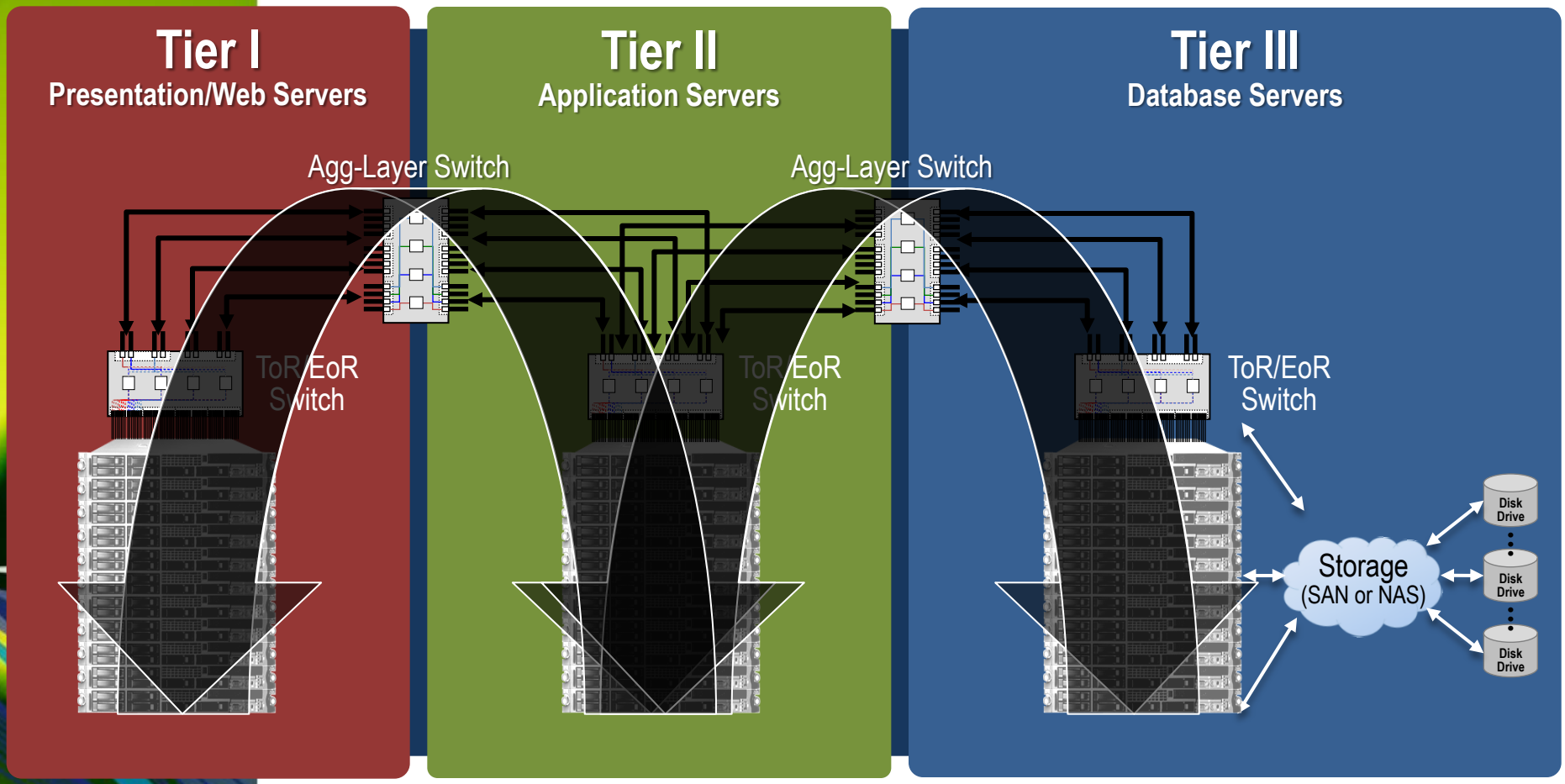
### *Elasticity:*

- Dynamic resource pooling – Grow & Shrink
- Migration
- Parking

# Traditional 3-Tier Enterprise Datacenter
## Logical View



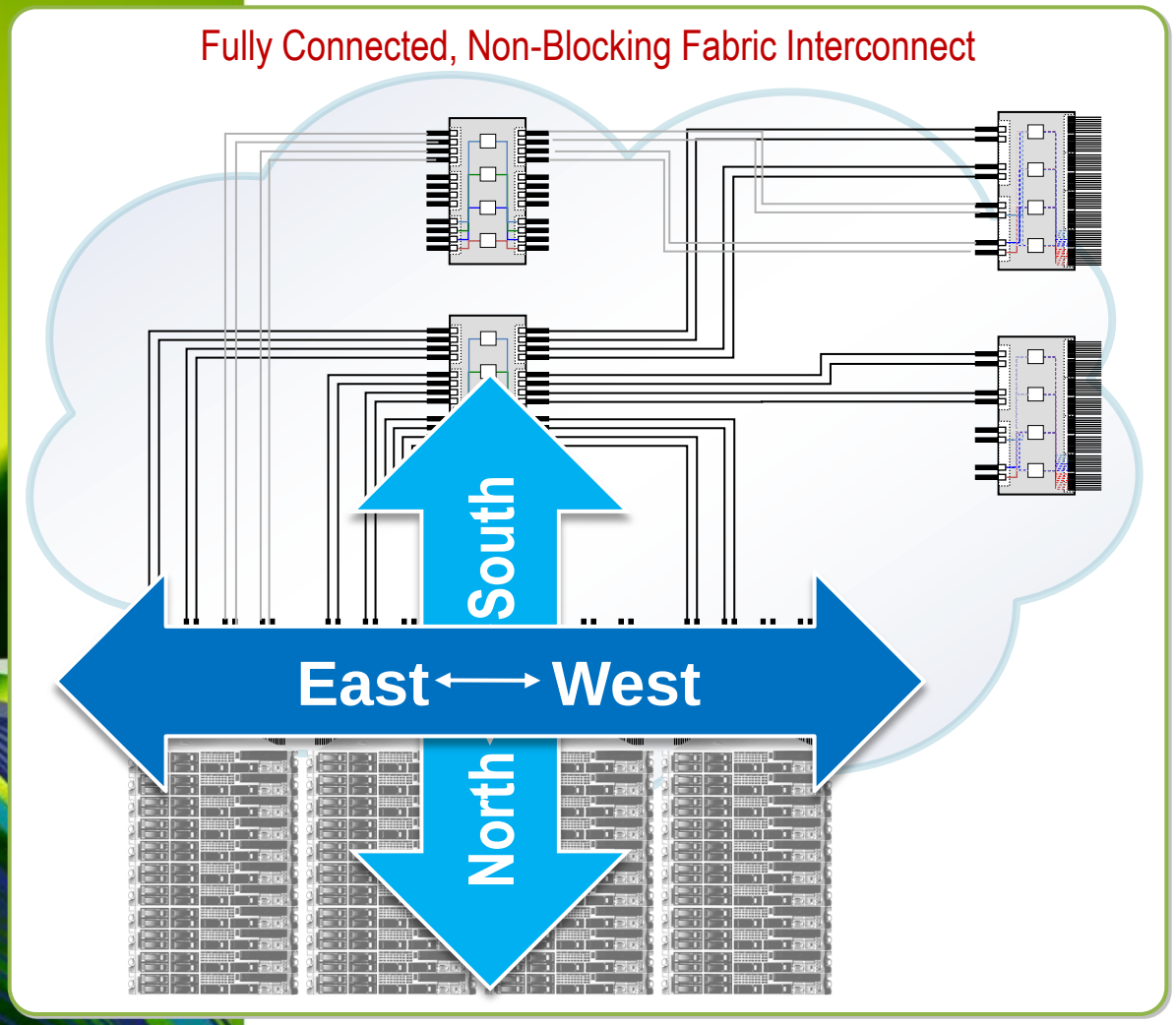
# Traditional 3-Tier Enterprise Data Center
## Physical View




Traffic was predominantly North➔South – almost no traffic within each tier

Provisioning is a "Physical" exercise and difficult to re-configure

# The New "Virtual Data Center"
# Flat network with full cross-sectional B/W from any server to any server




- Any Server can be in any "Tier" of the Data Center
  - Provisioning is completely Virtual
- Fast re-configuration is possible with minimal manpower
- No assumptions can be made regarding North➔South vs. East➔West traffic patterns

10GbE converged controller and fabric facilitate virtual data centers

# Traditional Approach vs. Public Cloud

## Compute Islands

OLTP

HPC

Development/Test

Web 2.0

High OPEX/CAPEX costs

Low flexibility & scalability

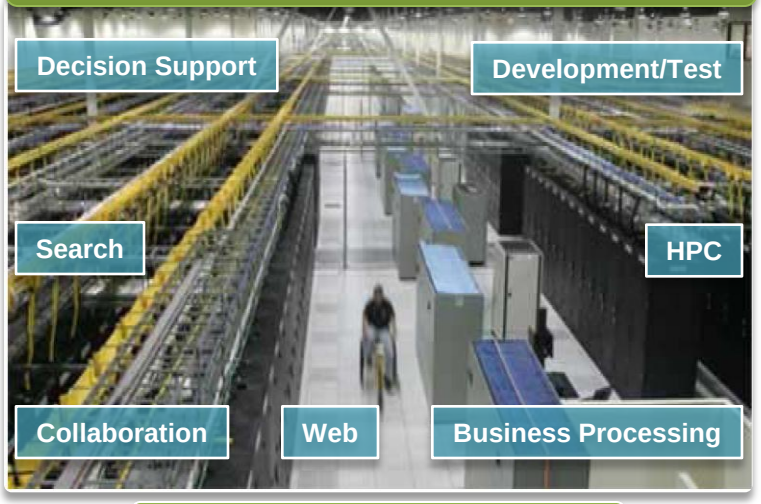
*vs.*

## Global Computing

Secure | Dynamic Mgt | APIs

Virtual | Reliable | Agile

**Internet / Intranet**




Homogenous with arbitrary scale

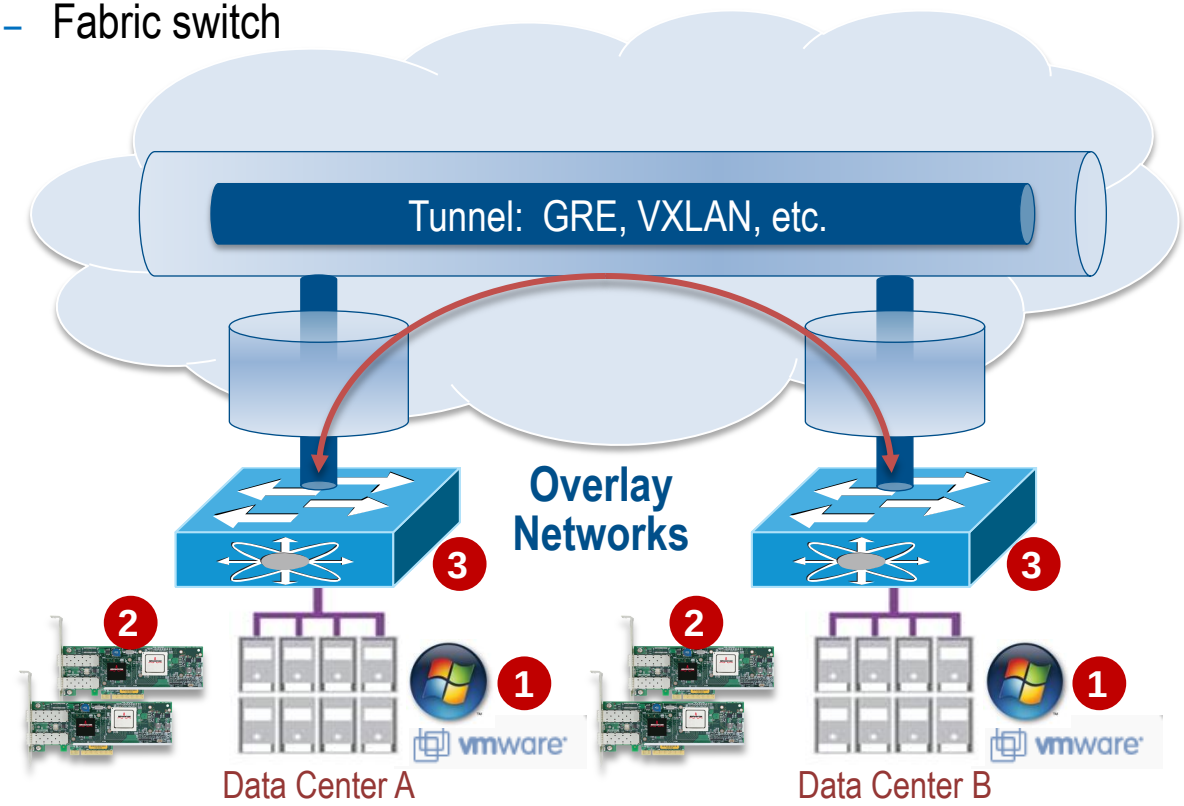
Unified fabric

Extreme efficiency & performance

BROADCOM. Connecting everything®

# Convergence and Overlay Networks
## (also known as Network Tunnels)

- Tunnels form "Virtual Networks" that fully abstract the Physical Network, extending existing constructs (VLANs, IP subnets, etc.)
- Tunnels can be originated or terminated in:
  - Host software
  - Network controller
  - Fabric switch



# Server Virtualization is Driving 10G

## Vectors Driving Convergence

### 1. VM Density
16+ VMs per server common
Usually requiring 1GbE per VM
CPU performance @ or >10GbE

### 2. Best Practices
4-8x GbE LAN / NAS / iSCSI
2x GbE Management
2x GbE VMotion

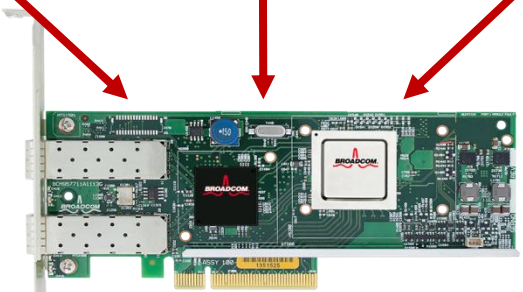
### 3. Network Storage
85% of VM installs use
Network storage
(allows for VM migration)

**Adapter, cable consolidation**
**Reduced power and cost**
**Improved performance**

**NIC Partitioning w/ QoS allows**
**1Gb deployment model w/**
**better performance, cost and Mgmt**

**Enhances virtualized deployment**
**iSCSI and NAS 10GbE performance**
**Converged fabric – Cost/Power/Mgt**

**Networking**
Direct Queues
NIC Partitioning
SR-IOV



(Converged NIC)

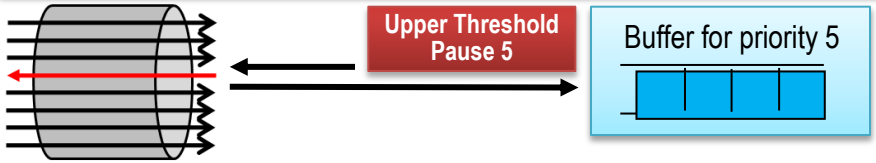
**Storage**
NAS, iSCSI & FCoE

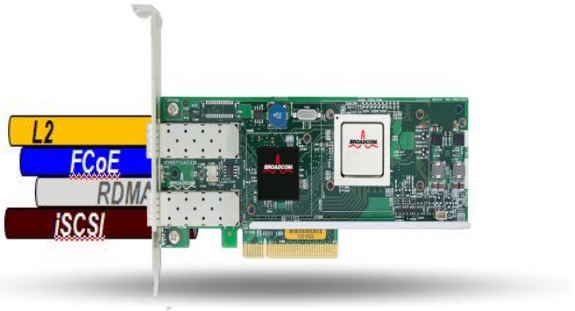
Server virtualization is a huge driver for 10Gb+ Adoption

# Technologies that Facilitate Convergence

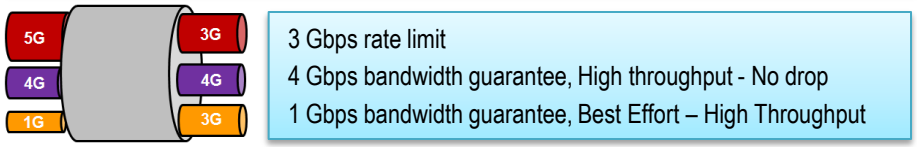
## PFC, ETS, QCN, and DCBx

- **Priority-based Flow Control (PFC): IEEE P802.1Qbb**
  - ✓ Provides link layer flow control by priorities – 8 priorities
  - ✓ PFC prevents drops during the onset of congestion



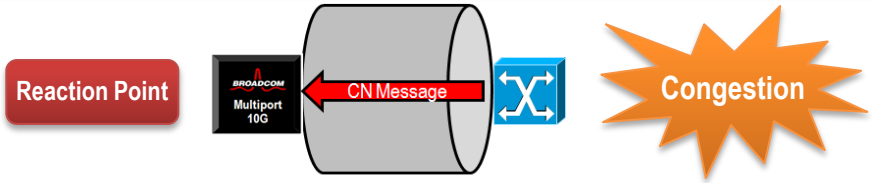




- **Enhanced Transmission Selection (ETS): IEEE P802.1Qaz**
  - ✓ Provides minimum bandwidth guarantees
  - ✓ Unused BW is available for others




**"DCBx"**
**Uses LLDP to exchange PFC/ETS configuration**

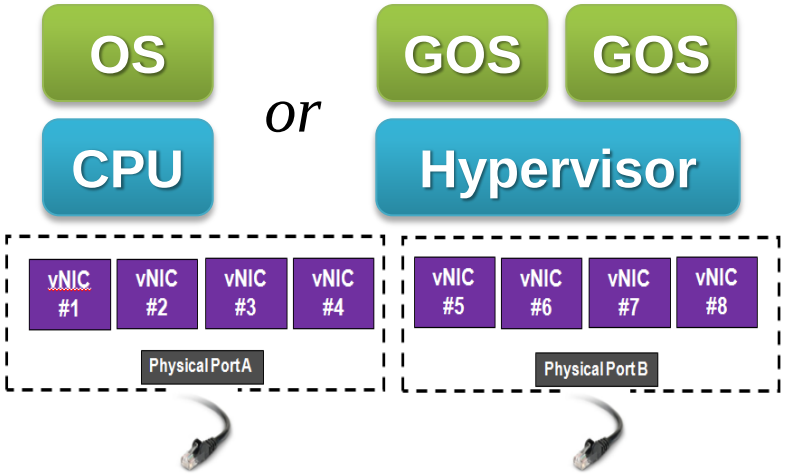
- **Quantized Congestion Notification (QCN): IEEE P802.1Qau**
  - ✓ Reduce congestion duration by slowing traffic at the src
  - ✓ End nodes monitor buffer utilization for desired equilibrium

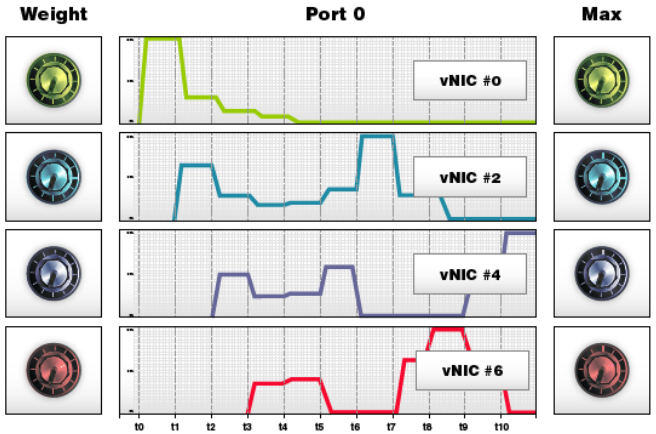


**Critical new technologies that allow traffic isolation and QoS**

BROADCOM
Connecting everything

# Virtualization: NIC Partitioning



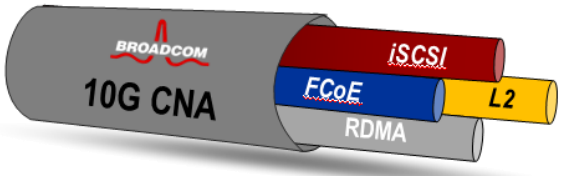




## 1. Independent NIC partitions

- ✓ Four or more per physical port
- ✓ Supports single OS or Hypervisor environments
- ✓ No OS or BIOS changes required
- ✓ Switch independent and switch dependent modes of operation

## 2. Fine-Grain B/W Control

- ✓ Control for fixed Bandwidth caps and Bandwidth weighted priority

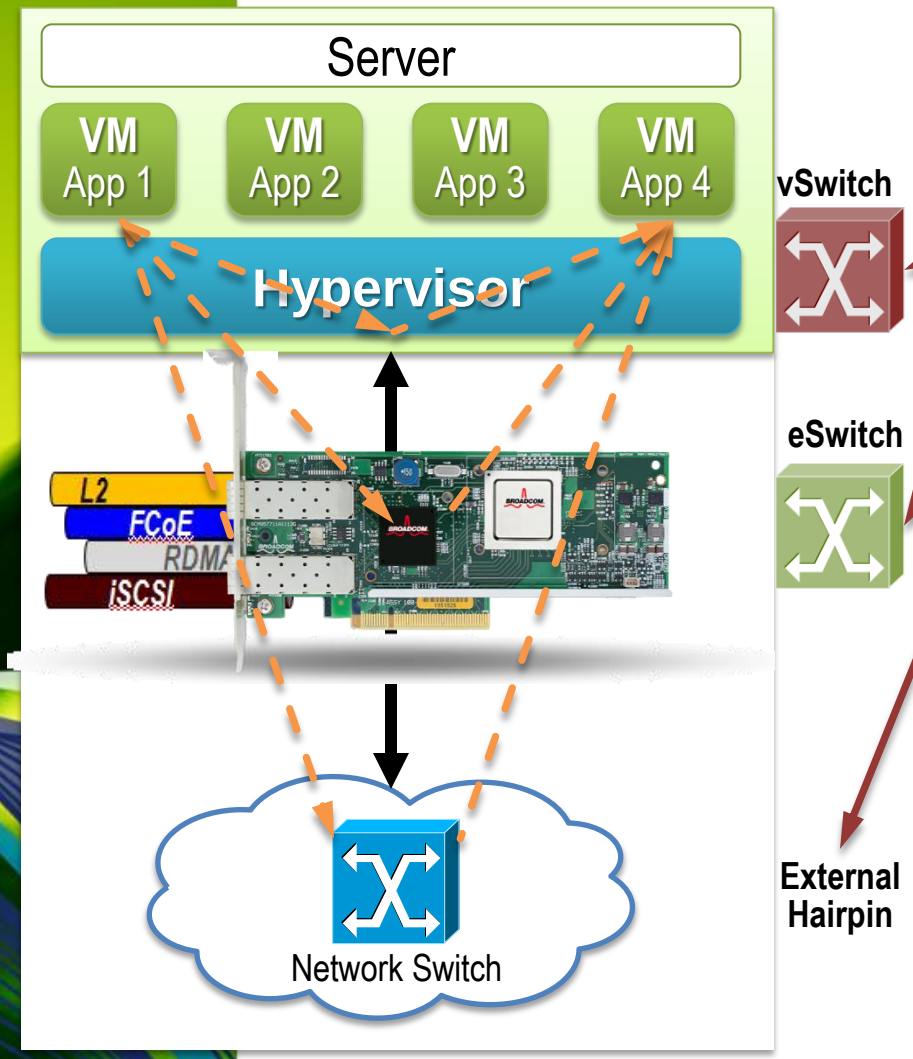



## 3. Convergence

- ✓ Networking and storage

NIC partitioning: Multiple logical pipes on a single big pipe with QoS

# Server Virtualization:
## Managing the New Switching Edge




**New switching points inside the Ecosystem:**

*NIC Partitioning and SR-IOV Enable Internal Switching between vNICs and vHBAs*

### Benefits

- Flexibility and availability
- Speed of innovation

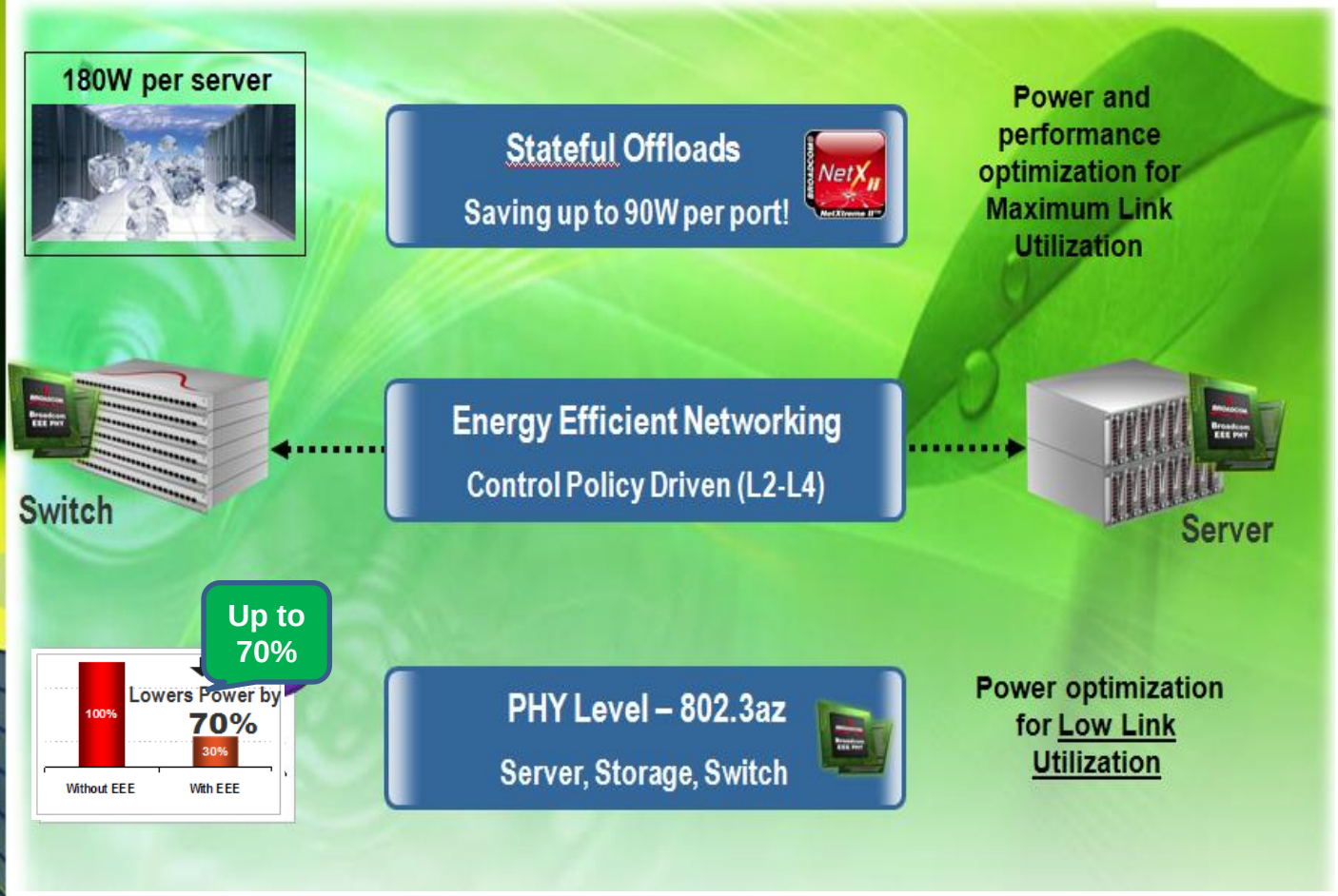
### Challenges

- Management
- Policy enforcement
- Performance (for vSwitch)

# Convergence and Power Management

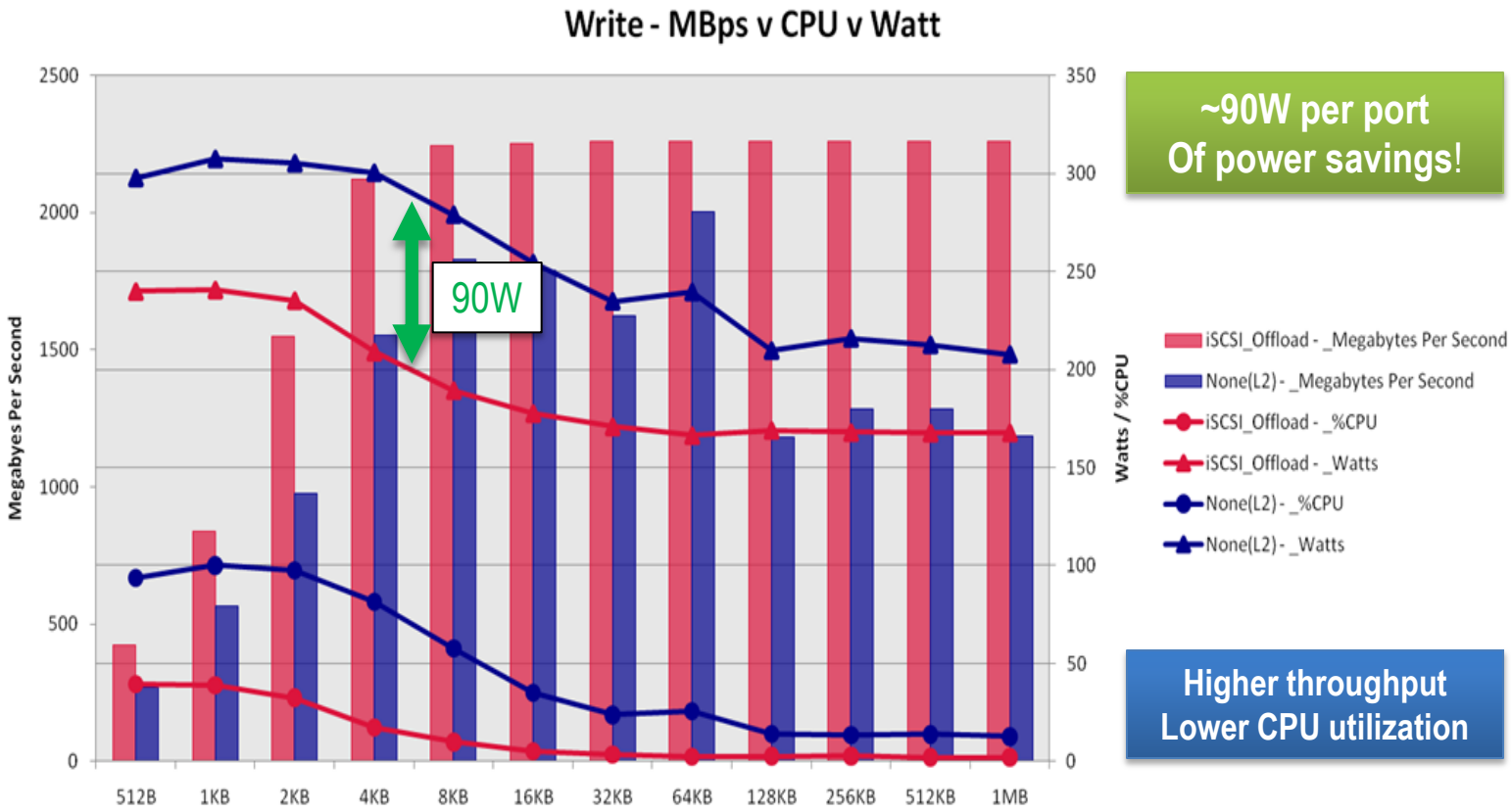
## Offloads, EEN, and EEE

In addition to consolidating Controllers and Switch ports, state-of-the-art Converged Controllers Conserve Power with Offloads!




Reduced power across full spectrum of data traffic patterns

# Power Management in Action
## iSCSI Hardware Offload Example




**~90W per port Of power savings!**

**Higher throughput Lower CPU utilization**

### Offload Provides Data Centers

- ✓ Improved PUE (Power Utilization Efficiency)
- ✓ Returns CPU to be used for applications, not I/O

# Three Tenants of Virtualized Performance

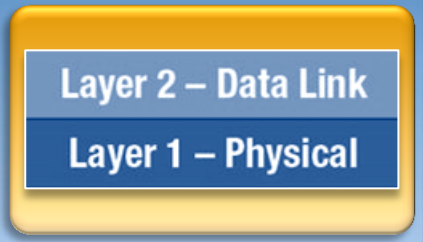
All three elements are required to provide SLAs

1. **Handling *Consolidated* Server Workloads**
   - Increased workload with #of VMs per Server
2. **Handling *Converged* I/O Workloads**
   - Networking + Storage (NAS, iSCSI or FCoE)
3. **Handling *Bursts* of Traffic**
   - There's no such thing as an "Average" traffic burst
     - When an application *steps on the gas,* it wants immediate response!

# Market Musts for a Consolidated Controller:
## Converged and Virtualized Networking Platform

Performance and Efficiency Across all Protocols and Environments

### L2 Networking




- PCIe Gen2 and Gen3
- Stateless offloads
- Multiple teaming
- Data Center Bridging (DCB)
- RX/TX coalescing
- Deep RSS
- Management

### Virtualization

- MSI-X (Multi-core)
- NIC Partitioning
- NetQueue / Multi-Queue
- SR-IOV
- Internal eSwitching

### Consolidation - Convergence

- NAS (NFS & CIFS)
- iSCSI HBA
- FCoE HBA
- TOE
- Low Latency/RDMA
- 1588 and 802.1AS
  - Forensics and troubleshooting

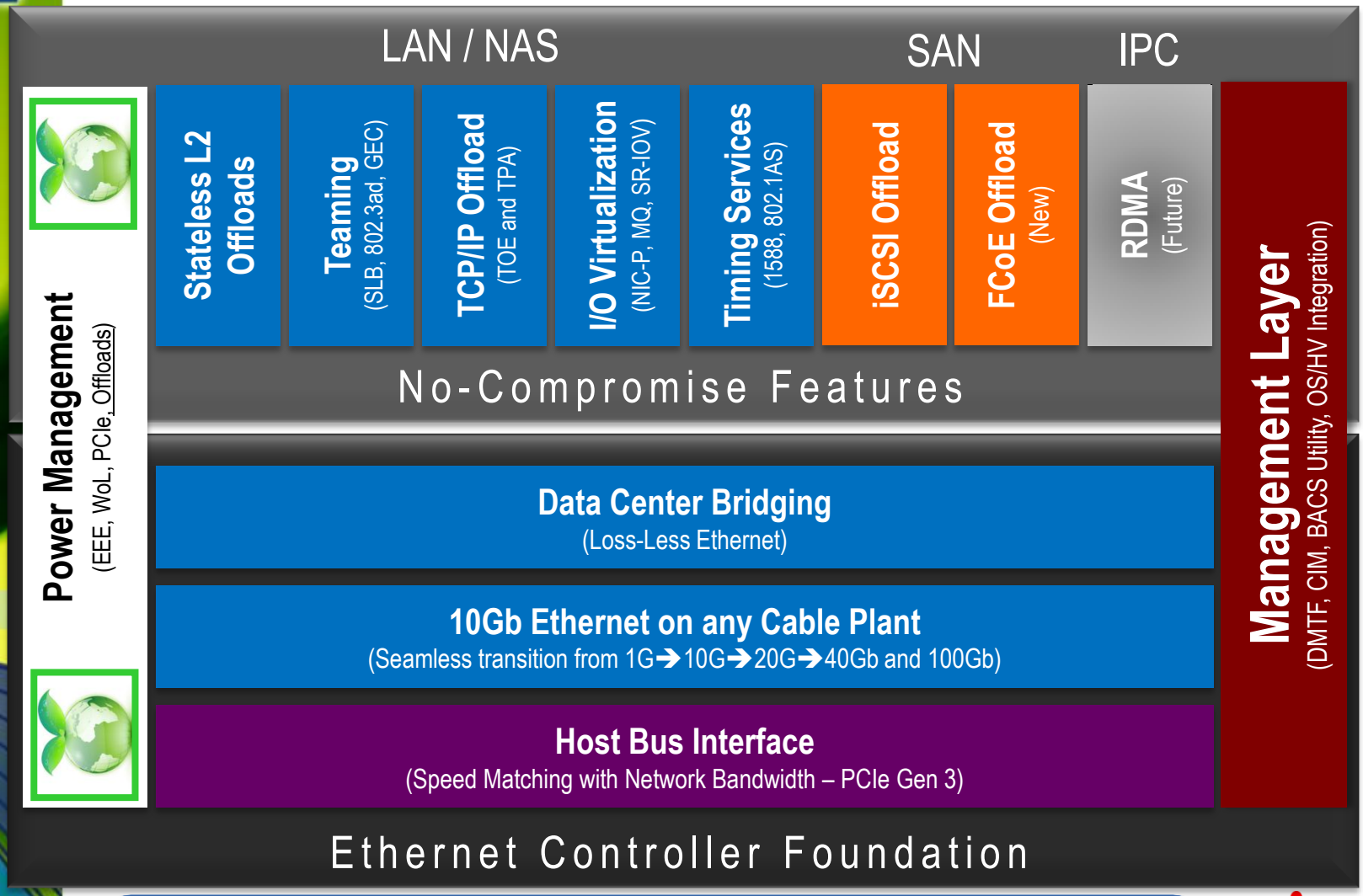
### Power Management

- Hardware offloads
  - **90W/port savings**
- PCIe ASPM and CLKREQ#
- WoL (Wake on LAN)
- Energy Efficient Ethernet
- Passive cooling

Industry Standard Management (DMTF, SNIA, EVB, DCBX, NC-SI, etc.)

COMPUTERWORLD SNIA SNW
BROADCOM Connecting everything

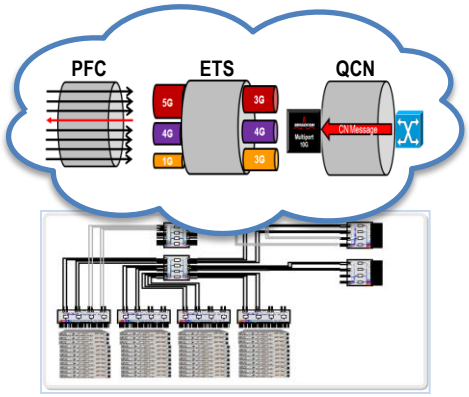
# Ideal Converged Controller Architecture

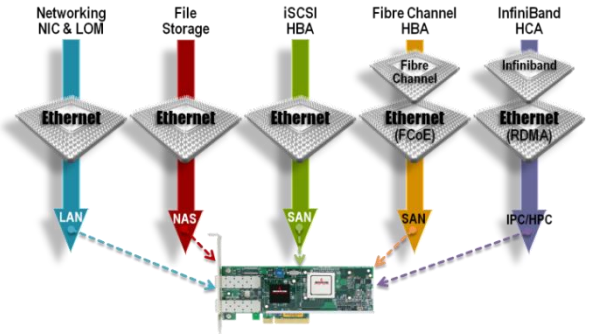



Common platform architecture and compatibility 1Gb➔10Gb➔40Gb

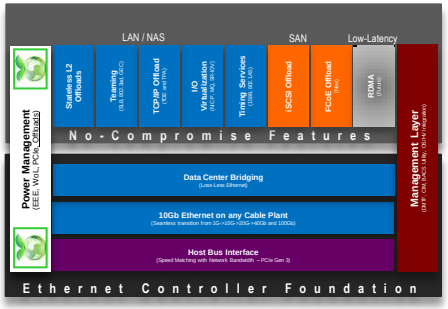
# Conclusions and Call to Action

- Consolidation and convergence require new network architectures to be deployed
  - Loss-less Ethernet (DCB)
  - Fat and flat L2/L3 fabrics
- Storage and networking are, and HAVE been converging
  - Server virtualization and Cloud are accelerating the process
  - NAS is already converged with L2 and iSCSI has begun
  - FCoE is just the next workload to migrate to Ethernet
    - Short term consolidation over Ethernet; longer term convergence
- No one knows when full convergence will occur, so…
  - Future-proof with controller and switch architectures that are convergence-ready today!




**Consolidation vs. Convergence**




Future-proof with convergence-ready architectures today!

BROADCOM. Connecting everything.

# *Thank You !*

Have a question?
Contact me at:
**gscherer@broadcom.com**

# Useful Links

Hypervisor vSwitch: http://openvswitch.org/papers/hotnets2009.pdf

Multi-Queue: http://www.broadcom.com/collateral/wp/Virtualization-WP100-R.pdf

SR-IOV: http://www.pcisig.com/specifications/iov/single_root/

EVB: http://wikibon.org/wiki/v/Edge_Virtual_Bridging

DCB: http://en.wikipedia.org/wiki/Data_center_bridging

iSCSI: http://www.vmware.com/pdf/vsphere4/r41/vsp_41_iscsi_san_cfg.pdf

http://www.broadcom.com/collateral/wp/iSCSI-WP200-R.pdf

http://www.broadcom.com/collateral/wp/iSCSI-VMware-WP100-R.pdf

FCoE: http://fcoe.com/

EEE: http://www.itu.int/dms_pub/itu-t/oth/09/05/T09050000010005PDFE.pdf

COMPUTERWORLD SNIA

SNW

# Transforming the Information Infrastructure: Build, Manage, Optimize.

FALL 2011