Server Networking e Virtual Data Center

Roma, 8 Febbraio 2006

Luciano Pomelli

Consulting Systems Engineer
lpomelli@cisco.com

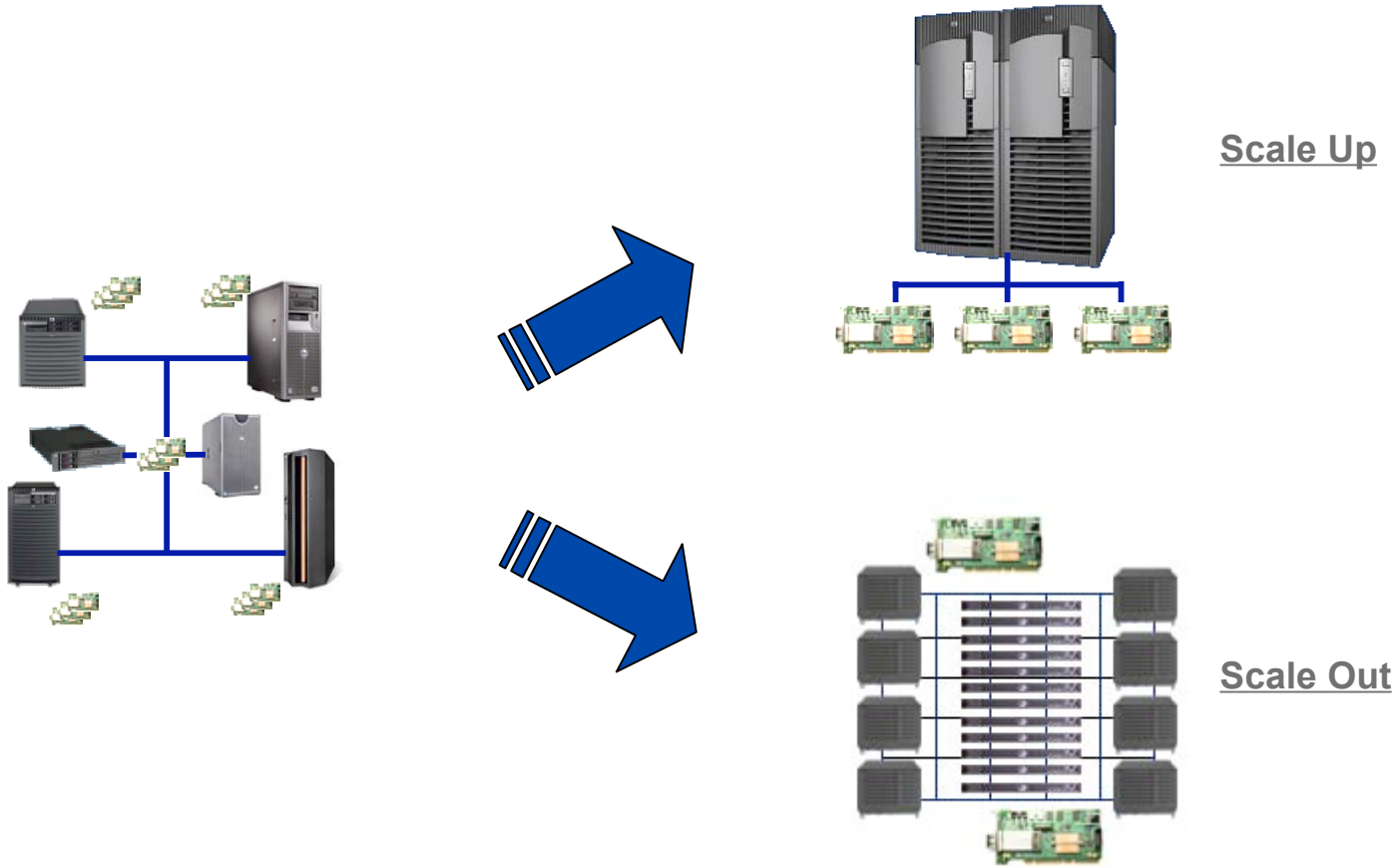
Typical Compute Profile at a Fortune 500 Enterprise

Compute Infrastructure Challenges:

- **Proliferation of disparate server platforms. 1200 servers across 12 data centers and 200 branches**
  - **35% Intel/Windows, 15% Intel/Linux 20% SUN Solaris, 10 HPUX, 20% IBM AIX**
- **Low average CPU utilization—Windows and Linux (15%) UNIX (30%)**
- **Continued use of Mainframes. High hardware and software maintenance costs**
- **Need for lower cost data analysis compute and storage resources**
- **Need faster server deployments for research projects and application development teams**
- **Piloting Blade Servers—challenges with cooling, server I/O, network connectivity**

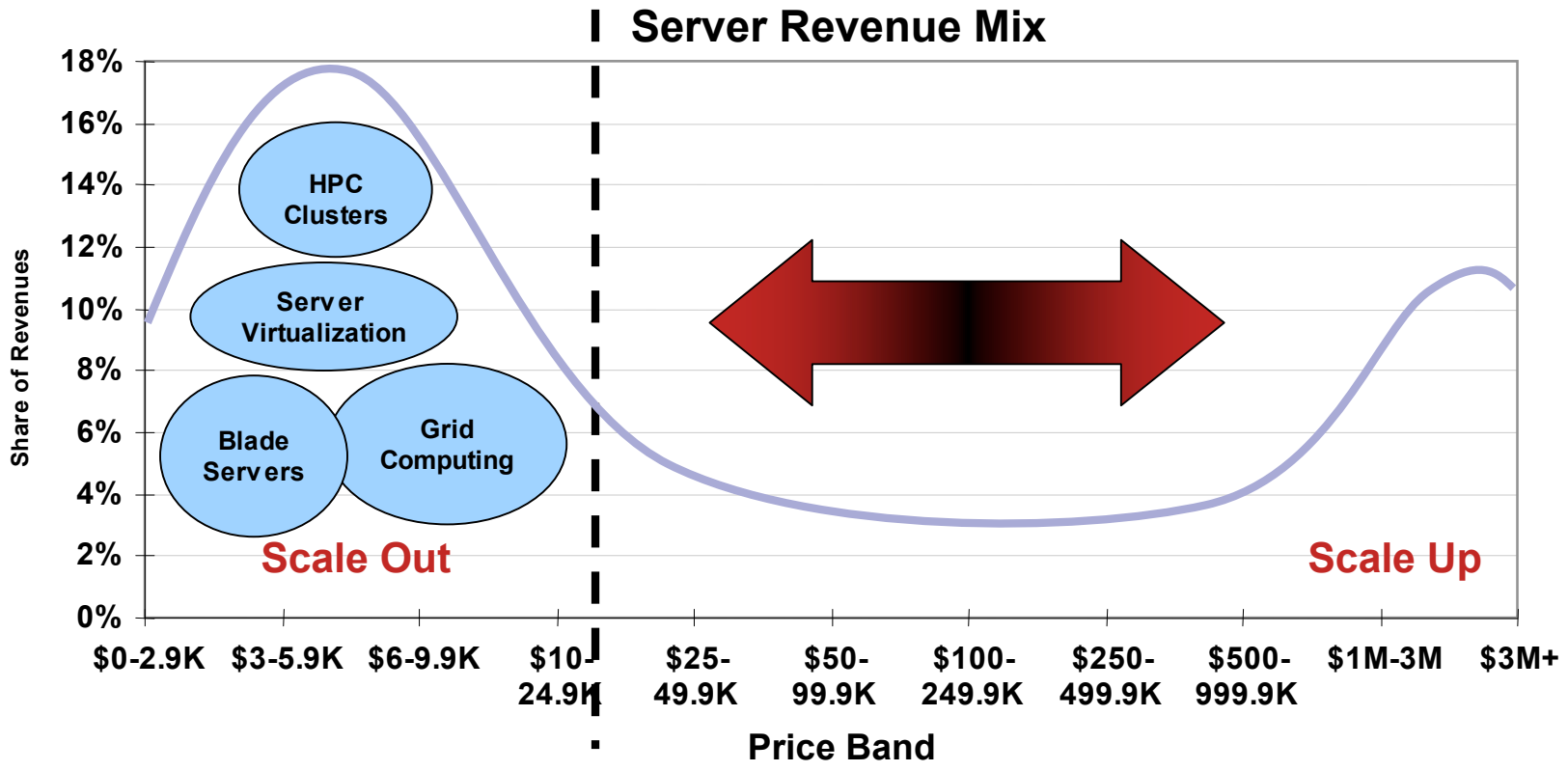
Server Optimization
“Scale Up” or “Scale Out”

Consolidate several mid-sized systems with fewer large servers



Evolve mid-sized systems to a system of standardized servers (blades, 1RU, small form-factor)

Industry Trend Towards Server "Scale-out"

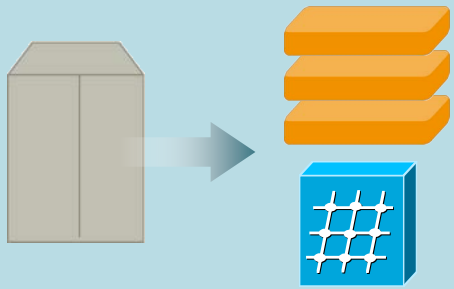


Scale out servers are low-cost, 1 rack unit (RU) or blade servers purchased as needed.

Cisco Computing Networking strategy

Performance

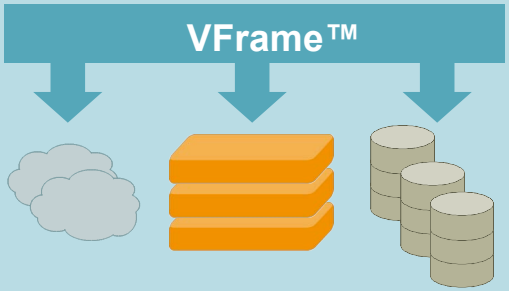
Server Clustering



- High Performance Computing (HPC)
- “Enterprise-class” HPC
- Database scalability

Economics

Utility or Grid Computing



- Application provisioning
- Server Virtualization
- Data Center Virtualization

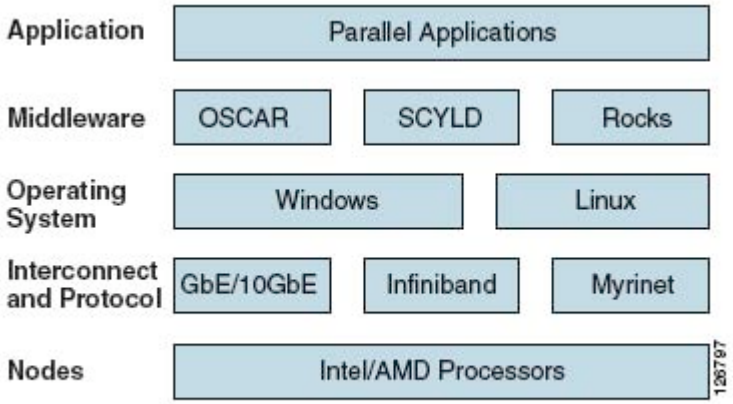
High Performance Computing e Clustering Networking

Cluster types

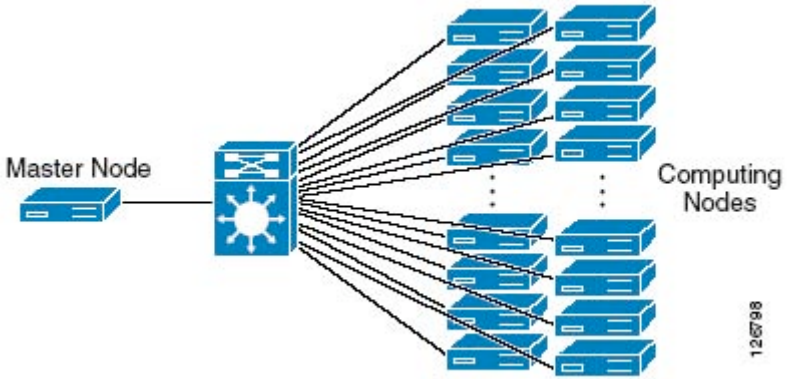
- **High Availability Clusters**
- **Application Clusters (Load-balancing)**
- **Parallel Processing Clusters**

- **High Performance Clusters**
  - **Bio-molecular**
  - **Computational Fluid Dynamics**
  - **Data Mining**
  - **Financial Analysis**
- **Enterprise Clusters**
  - **On Line Transaction Processing (OLTP)**
  - **Decision Support System (DSS)**
  - **Oracle, IBM DB2, MySQL, Microsoft SQL, Exchange**

High Performance Clusters

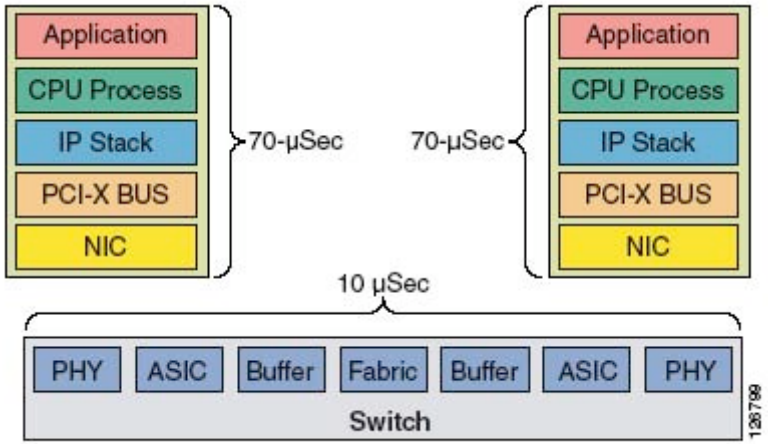


Cluster Components

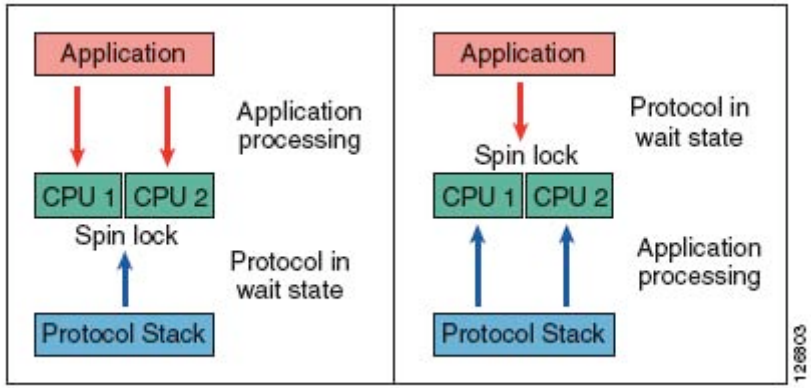


Cluster Topology

Performance impact



Latenza end-to-end dei messaggi



Utilizzo della CPU

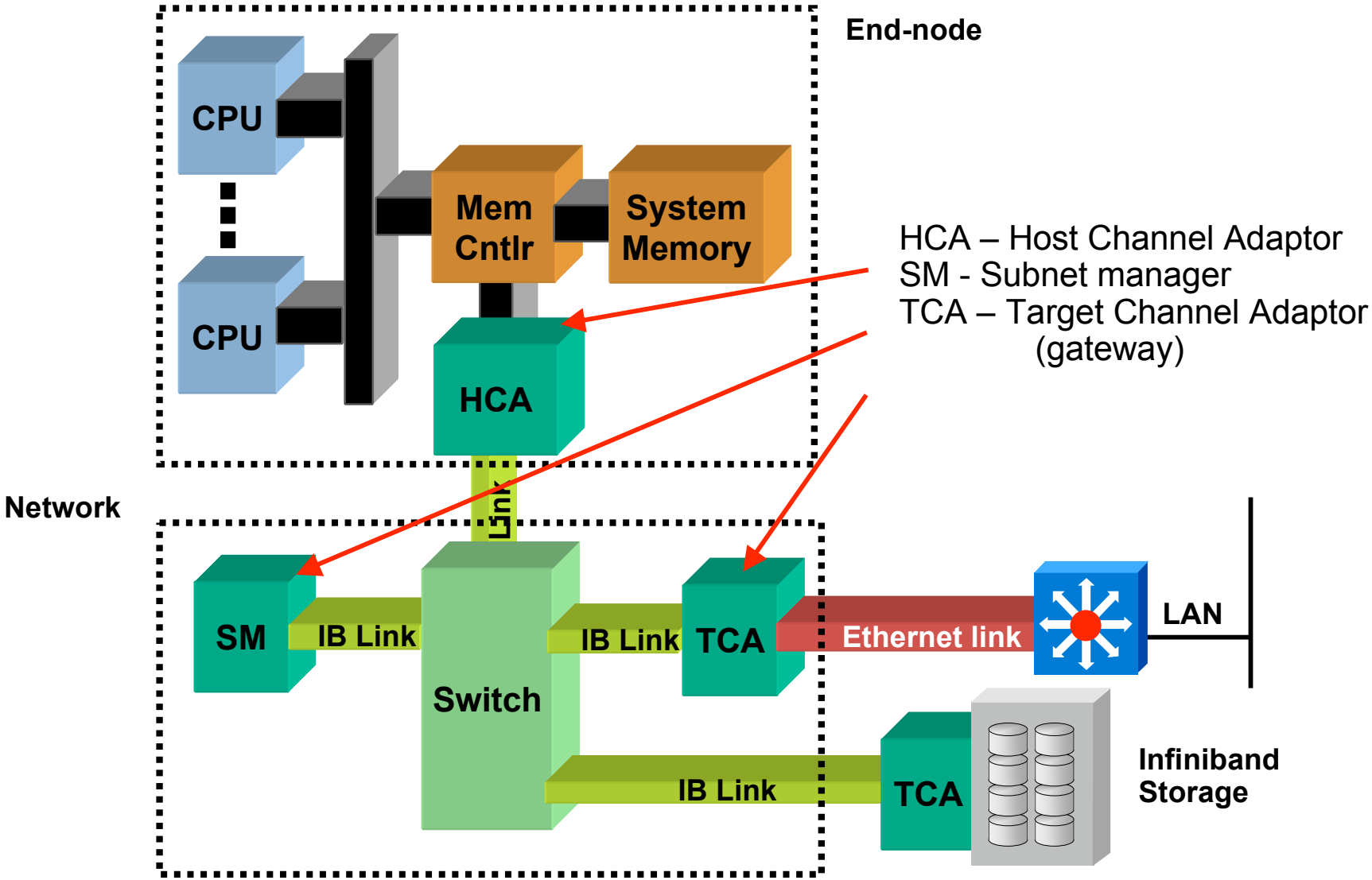
What is InfiniBand?

- InfiniBand is a high speed – low latency technology used to interconnect servers, storage and networks within the datacenter
- Standards Based – InfiniBand Trade Association

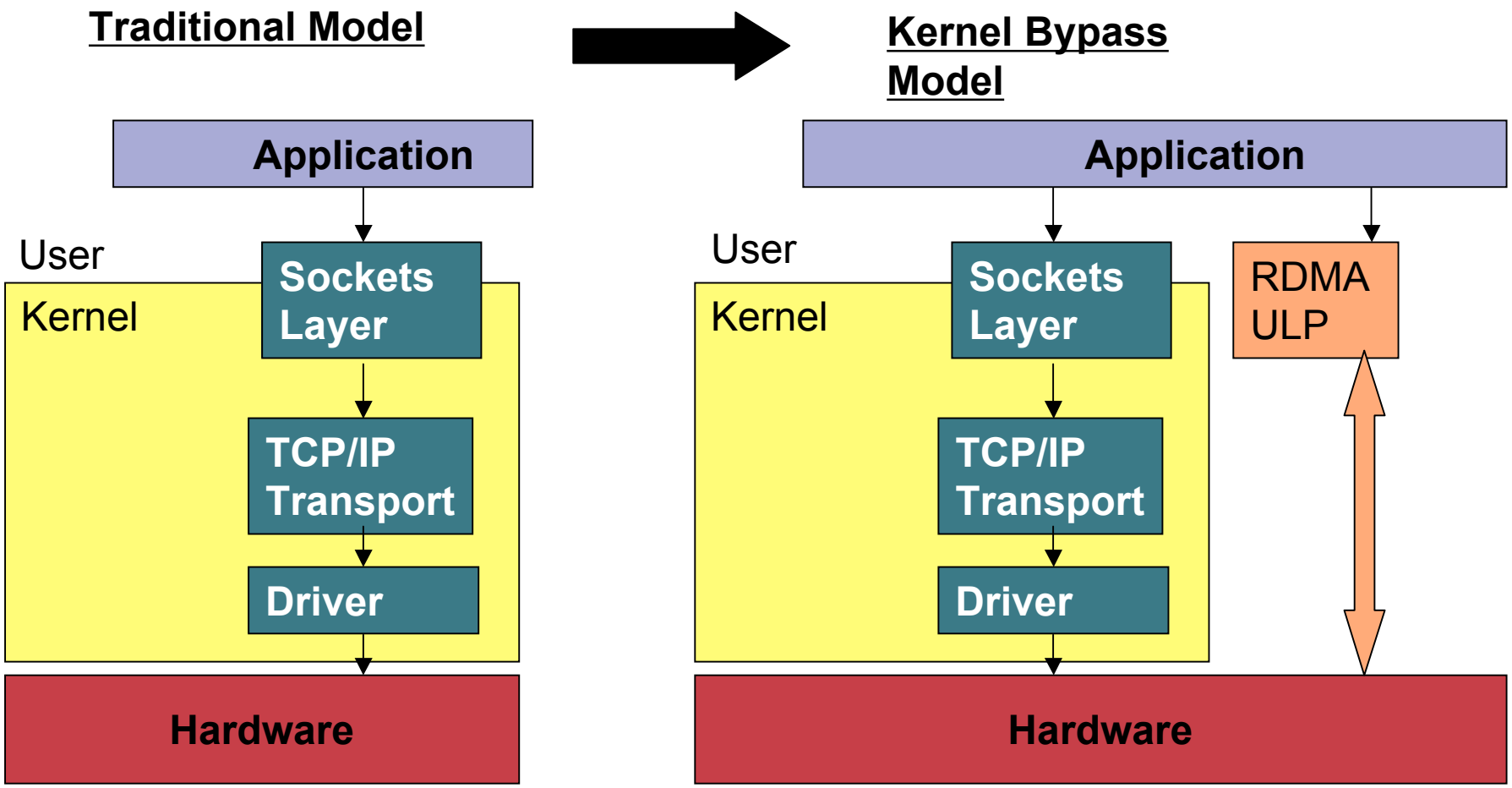
  http://www.infinibandta.org
- Scalable Interconnect based on Link Speed and Link Aggregation:

Links	SDR	DDR	QDR
1X	2.5 Gb/s	5 Gb/s	10 Gb/s
4X	10	20	40
12X	30	60	120

InfiniBand architecture

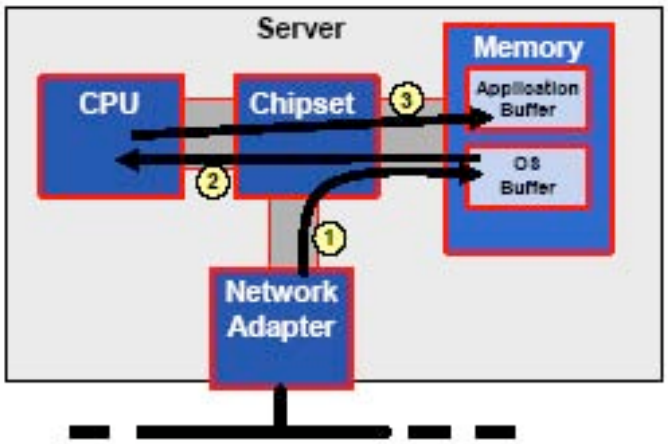


IB Technology: Kernel Bypass



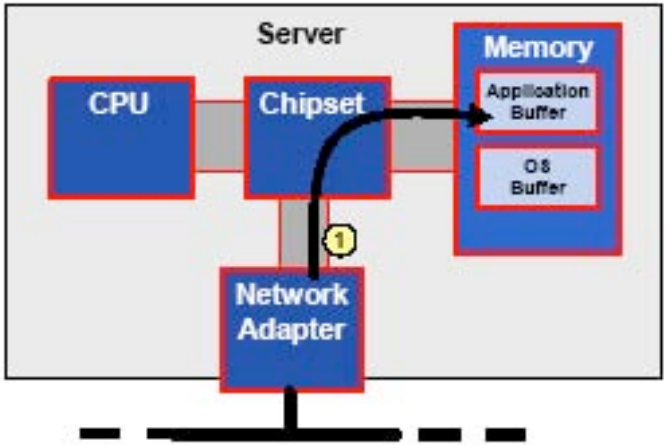
RDMA and OS bypass advantage

IPC using TCP over Ethernet



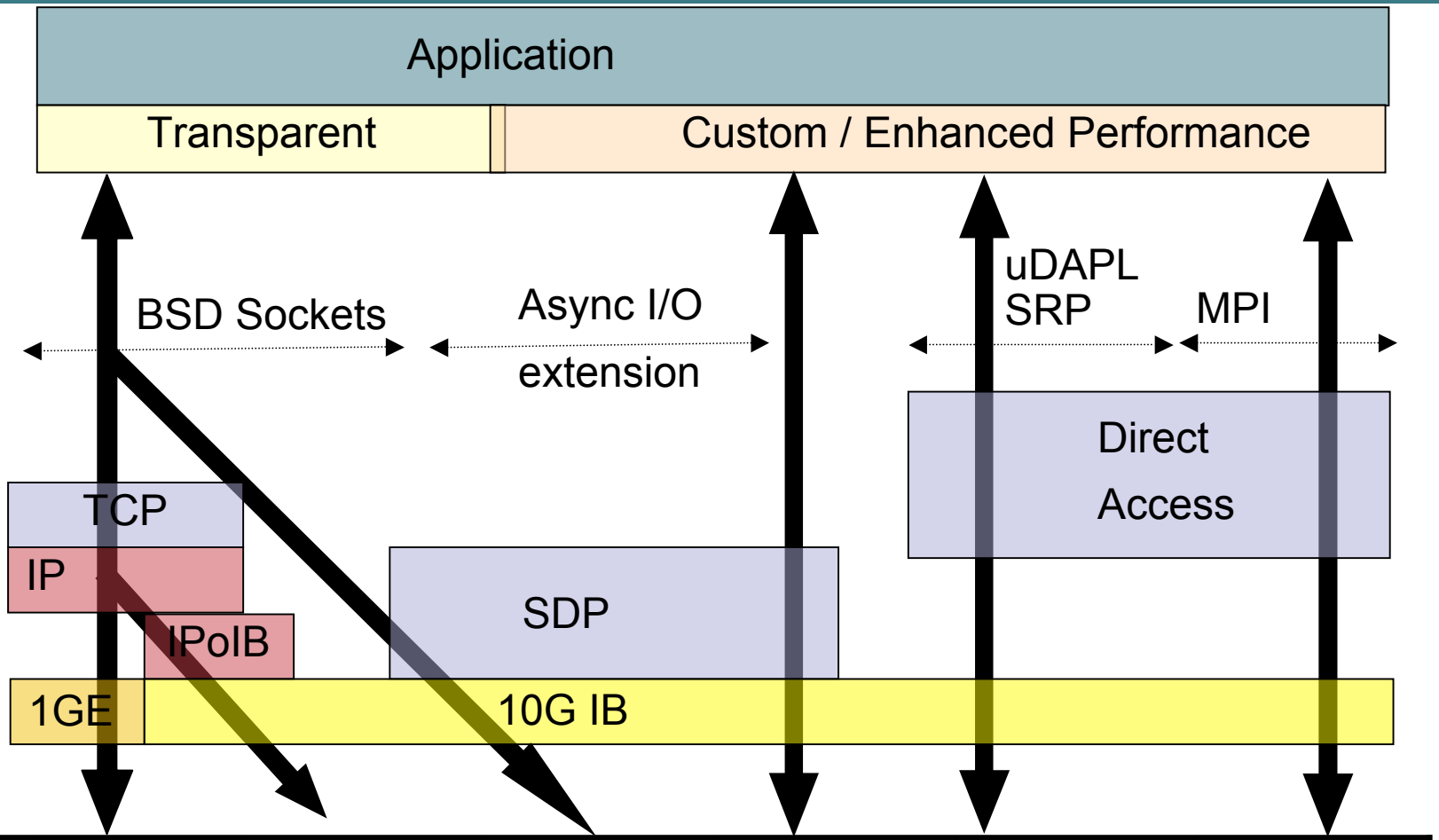
- Data traverses bus 3 times
- CPU involved in Data Movement

IPC using uDAPL or SDP over IB



- Direct Access from HCA to Application buffer
- Data traverses bus 1 time
- CPU NOT involved in Data Movement

InfiniBand Performance
Measured Results



Throughput	1-3 Gb/s	4.1Gb/s	4.5 Gb/s	7.9 Gb/s	8 Gb/s	8 Gb/s
Latency	40-60 usec	30 usec	18 usec	18 usec	8 usec	5 usec

Unmatched Price / Performance

InfiniBand Offers the Best Price / Performance for HPC

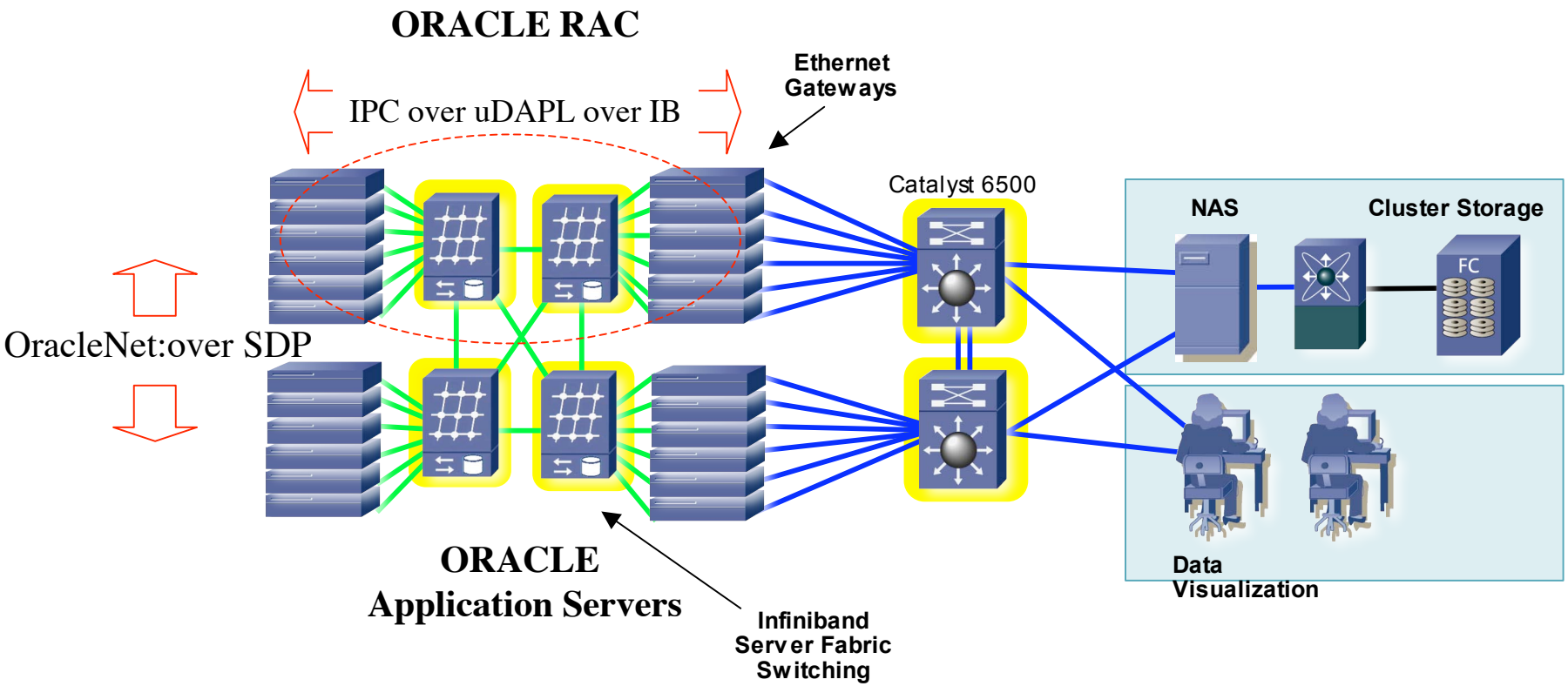
	InfiniBand PCI-Express	Myrinet D	Myrinet E	10GigE	GigE
Data Bandwidth (Large Messages)	**950MB/s**	**245MB/s**	**495MB/s**	**900MB/s**	**100MB/s**
MPI Latency (Small Messages)	**5us**	**6.5us**	**5.7us**	**50us**	**50us**
HCA Cost (Street Price)	**$550**	**$535**	**$880**	**$2K-$5K**	**Free**
Switch Port	**$250**	**$400**	**$400**	**$2K-$6K**	**$100-$300**
Cable Cost (3m Street Price)	**$100**	**$175**	**$175**	**$100**	**$25**

- *Myrinet pricing data from Myricom Web Site (Dec 2004) utilizing Myrinet's new switch*

*** InfiniBand pricing data based on Topspin avg. sales price (Dec 2004)*
**** Myrinet, GigE, and IB performance data from June 2004 OSU study*
*****10GigE and GigE Cost and Performance data from Cisco Internal document*

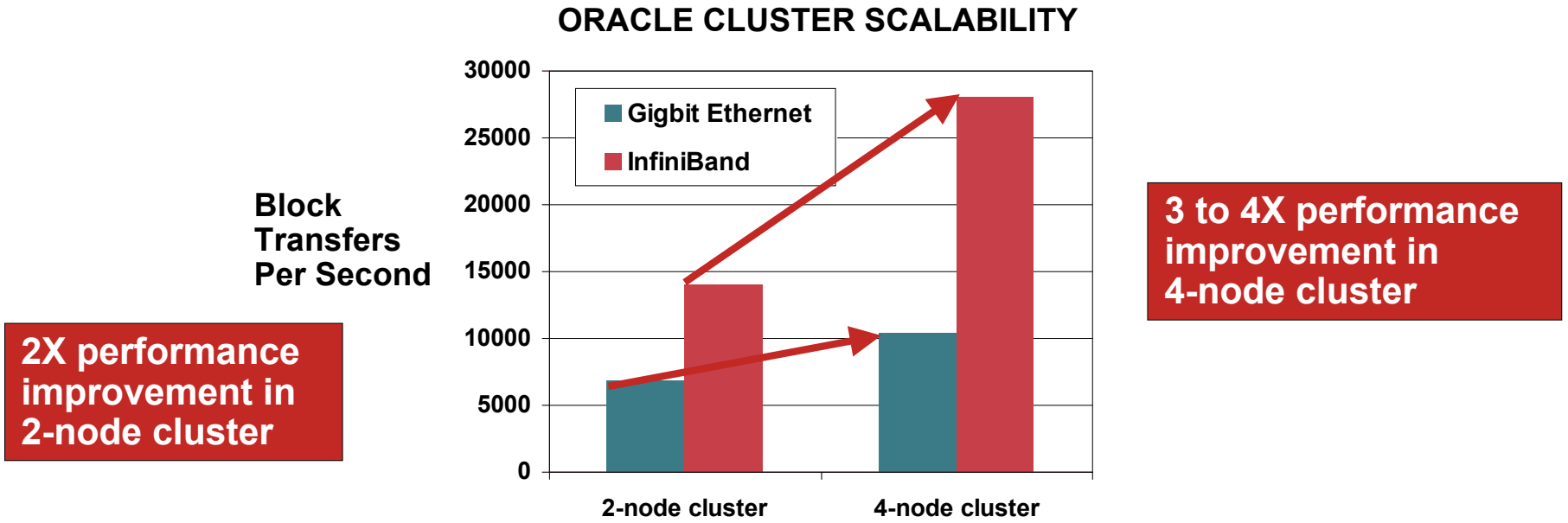
- *Note: MPI Processor to Processor latency – switch latency is less*

Oracle RAC and Cisco Infiniband example



Oracle 10G Infiniband linear scalability

Horizontal Database Clusters



- **Advantages of InfiniBand over Gigabit Ethernet interconnect**
  - **2 to 4 times cluster performance improvement**
  - **Much higher application scalability**

Source: Oracle whitepaper: Achieving Mainframe-Class Performance on Intel Servers Using InfiniBand Building Blocks

The Cisco SFS Product Line

Server Fabric Switch

Multifabric

SFS 3001 (TS90)

(12) 4XIB + 1 Gw

SFS 3012 (TS360)

(24) 4XIB + 12 Gws

(6) GE Gateways

InfiniBand

SFS 7000 (TS120)

(24) 4XIB

SFS 7008 (TS270)

(96) 4XIB

Blade Server

IBM BladeCenter

- HCA (2) 1XIB PCI-X
- Embedded switch (14) 1XIB (Internal) + (1) 4XIB and (1) 12XIB (External)

Dell 1855

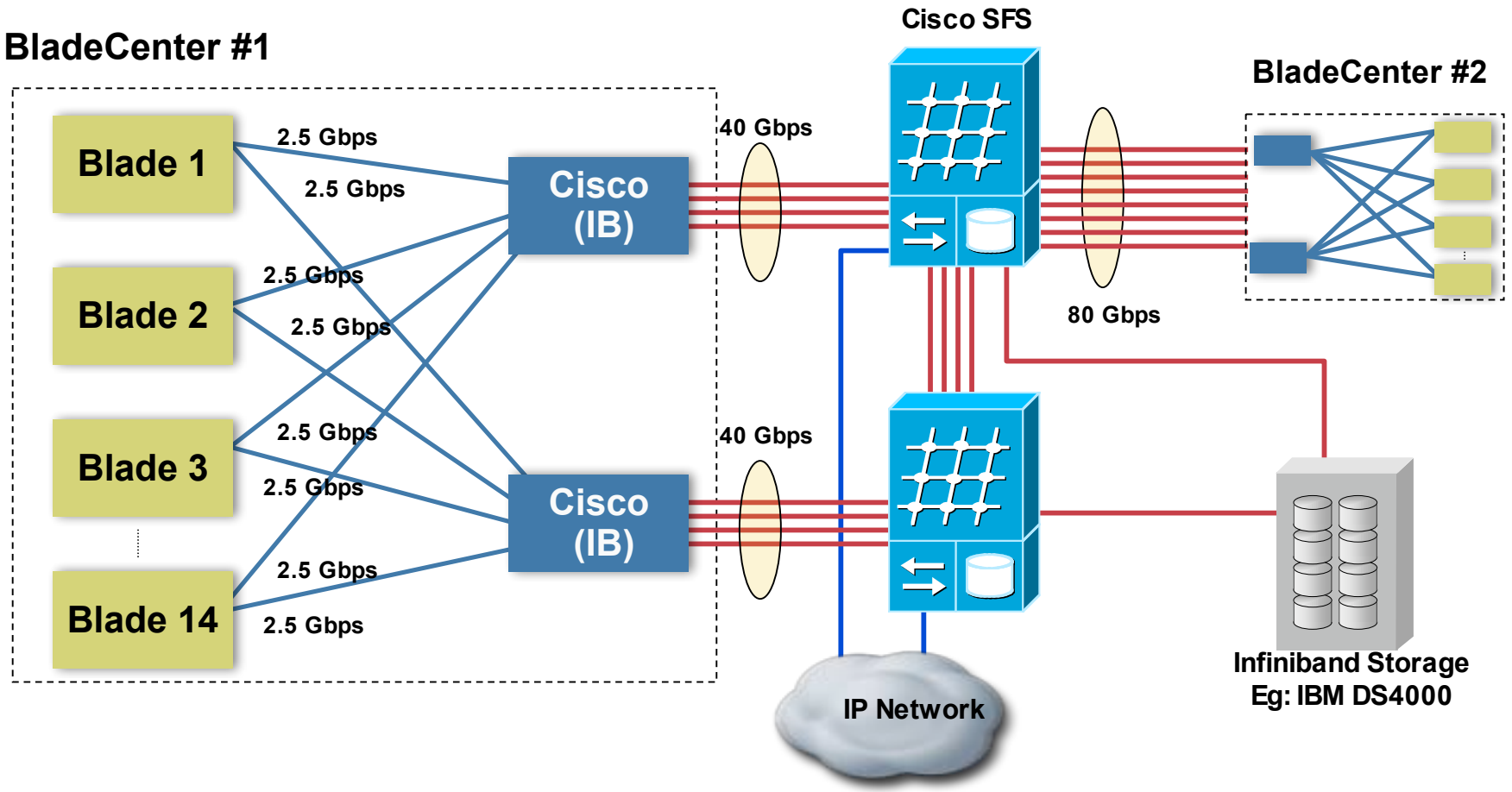
- HCA (2) 4XIB PCI-ex
- Passthru Module (10) 4XIB

HCA

- (2) 4XIB PCI-X
- (2) 4XIB PCI-ex

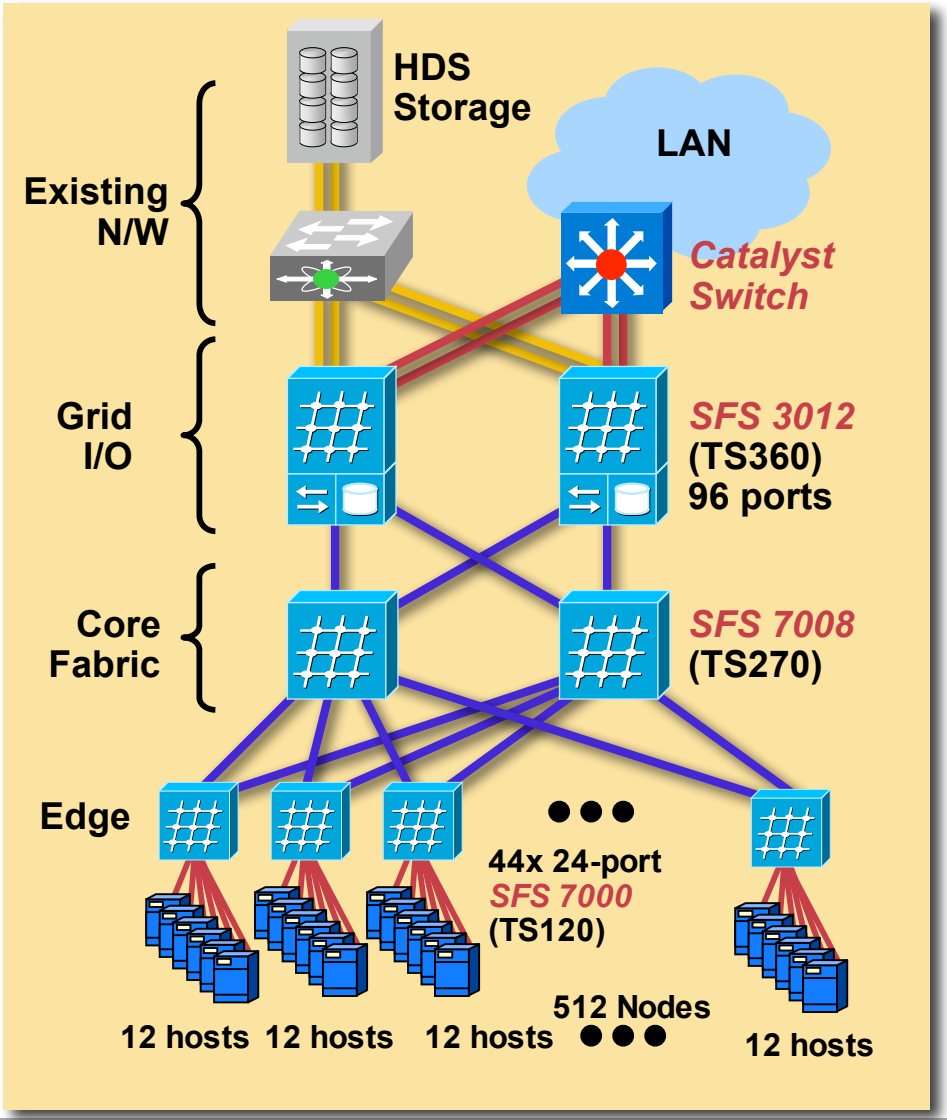
- Remote Boot
- Linux Host Driver
- Windows Host Driver

Blade Server Infiniband connectivity



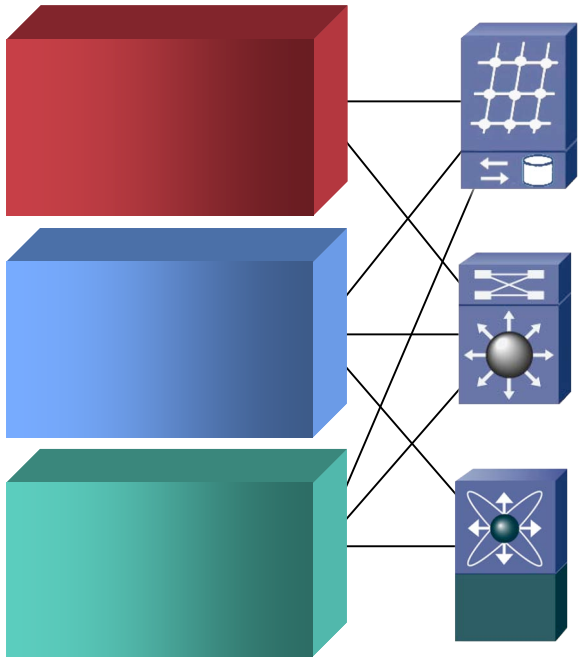
Case Study: Large Wall Street Bank
Enterprise Grid Computing

- **Application:**
  - Replace proprietary platforms with standards-based components
  - Build scalable “on-demand” compute grid for financial applications
- **Environment:**
  - 500+ Intel Servers per slice
  - Topspin Server Switch with Ethernet and Fibre Channel Gateways
  - Hitachi RAID Storage
  - SAN Switches
  - Ethernet Switches
- **Benefits:**
  - 20X Price/Performance Improvement over four years
  - 30-50% Application Performance Improvement
  - Standards-based solution for on-demand computing
  - Environment that scales using 500-node building blocks



Utility Computing and Data Center Virtualization

The Promise of Utility Computing



- Virtual I/O
- Cluster computing w/ RDMA
- Compute Virtualization

- LAN, WAN, Internet Access
- Networked application services
- Integrated security

- Storage Virtualization
- Fabric-assisted Applications
- Data Replication Services

Application or Policy

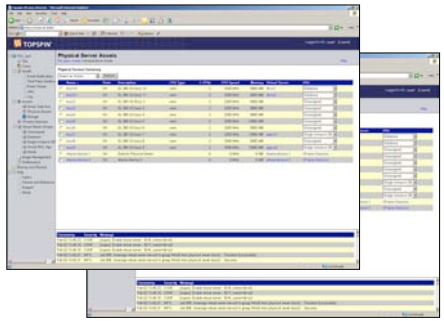
- **Racks of compute resources virtualized in the data center**

  Allocated to applications dynamically

  Network, storage and compute services applied dynamically to pools of resource

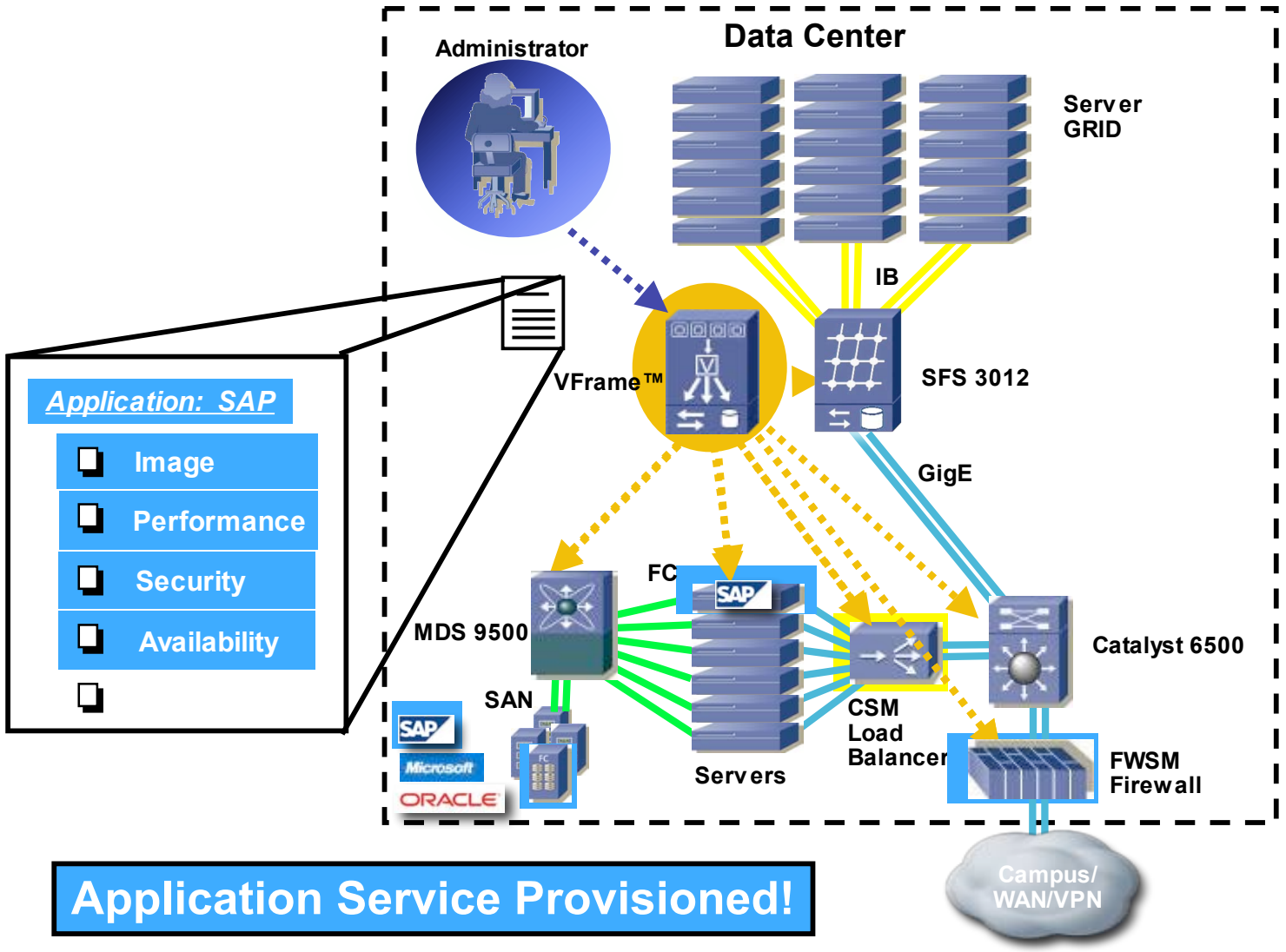
What is VFrame?
Cisco's end-to-end virtualization, provisioning manager

What is VFrame™



- Cisco's data center-wide virtualization software suite
- Enterprise software package that runs in the data center
- Delivers the end-to-end manageability, control, and virtualization benefits of the mainframe on top of today's commodity components and the Cisco IIN
- Provides virtualization, orchestration, and provisioning for the data center resources that sit between the "OS" and the "wire"

VFrame Vision
Cisco Virtual Data Center



Application Service Provisioned!

Define application services and pass policy to VFrame

VFrame translates policies to actions and passes to infrastructure

VFrame identifies right App / OS Image From storage

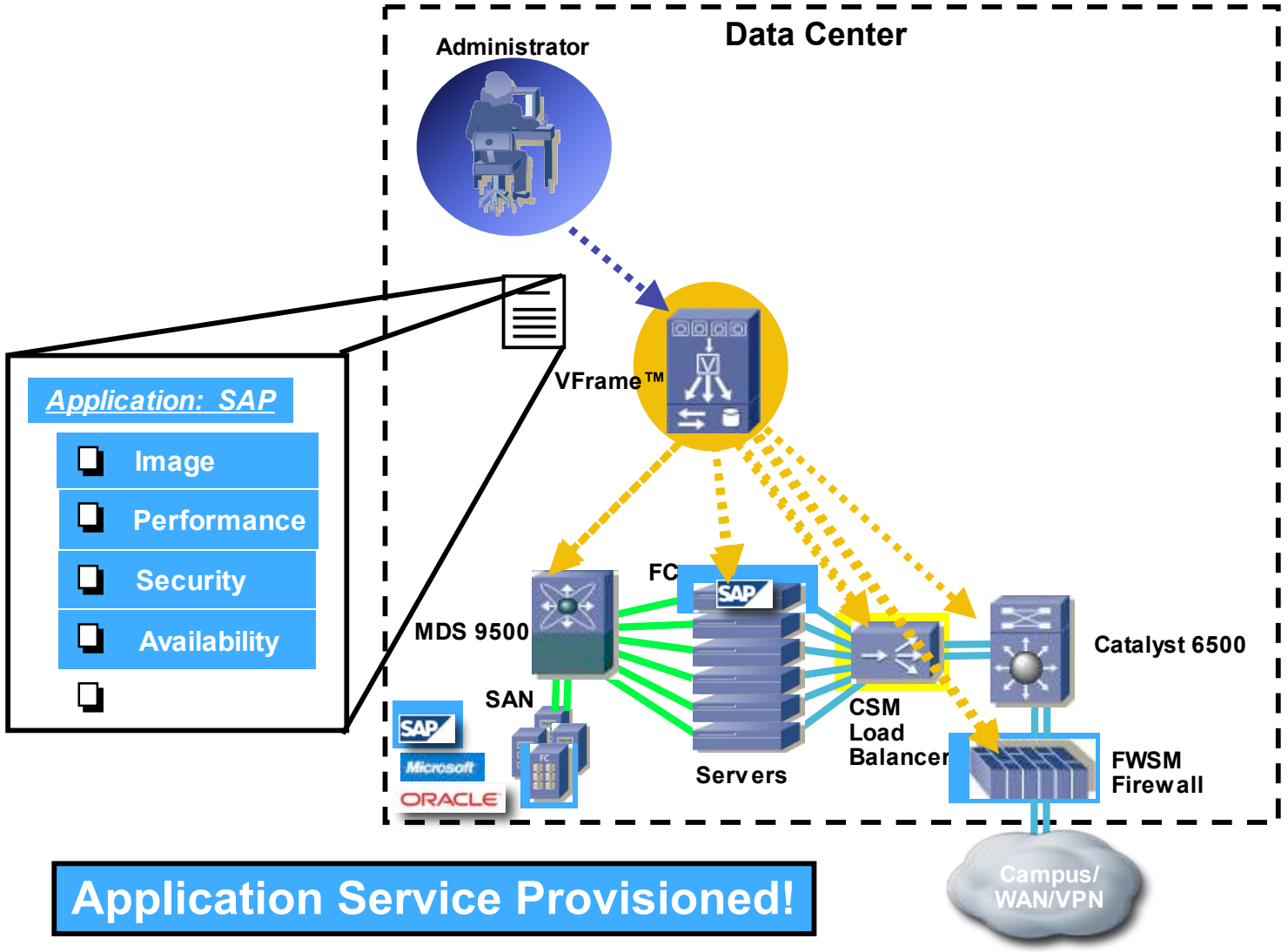
VFrame picks server with right criteria to run application and boots server

VFrame gives new server right VLAN, VSAN and LUN info so it can find/be found by right clients and storage

VFrame provisions security policies to FWSM

VFrame provisions CSM to add new server to load balancing pool

VFrame Vision
Cisco Virtual Data Center



Application Service Provisioned!

Define application services and pass policy to VFrame

VFrame translates policies to actions and passes to infrastructure

VFrame identifies right App / OS Image From storage

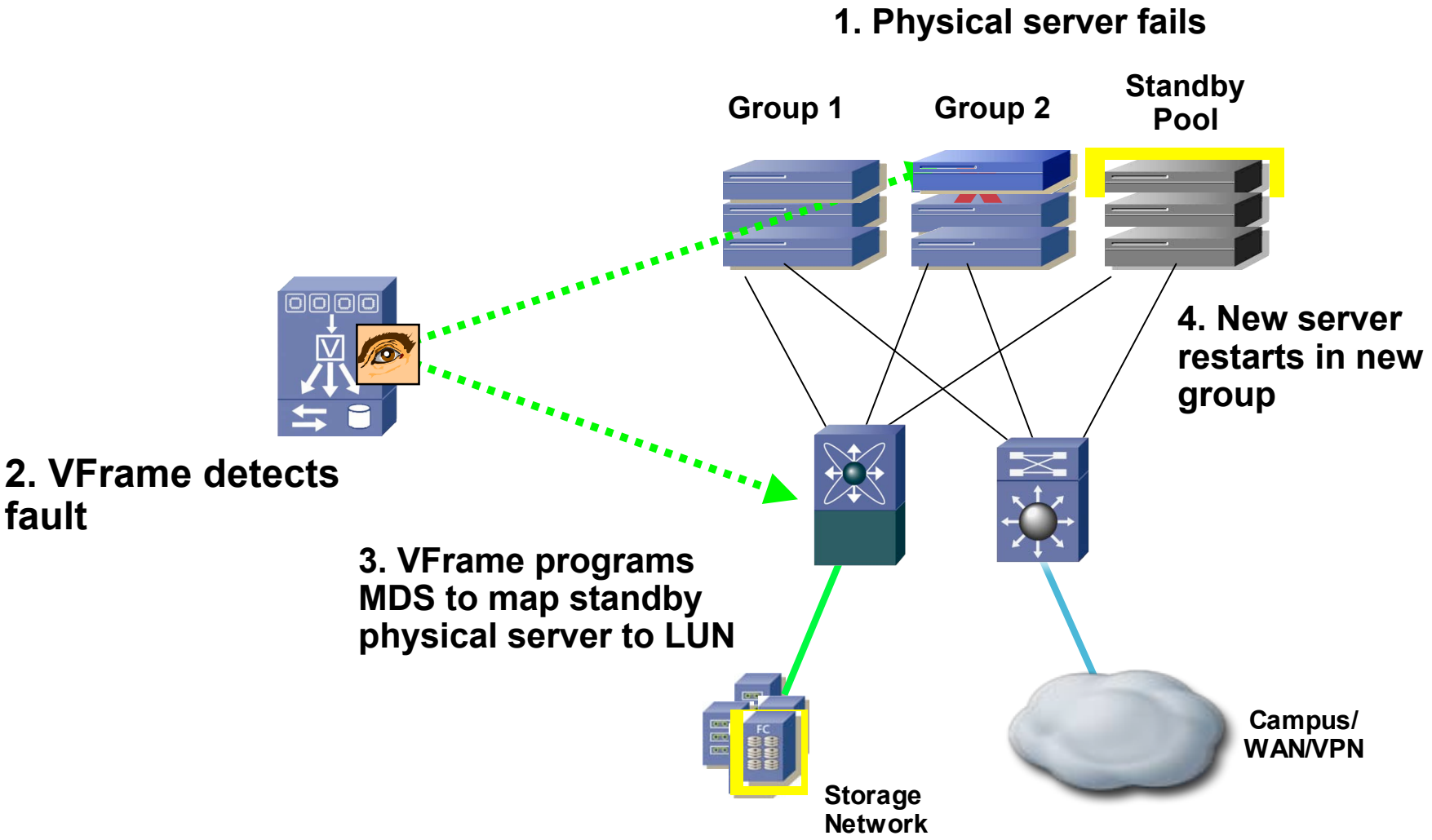
VFrame picks server with right criteria to run application and boots server

VFrame gives new server right VLAN, VSAN and LUN info so it can find/be found by right clients and storage

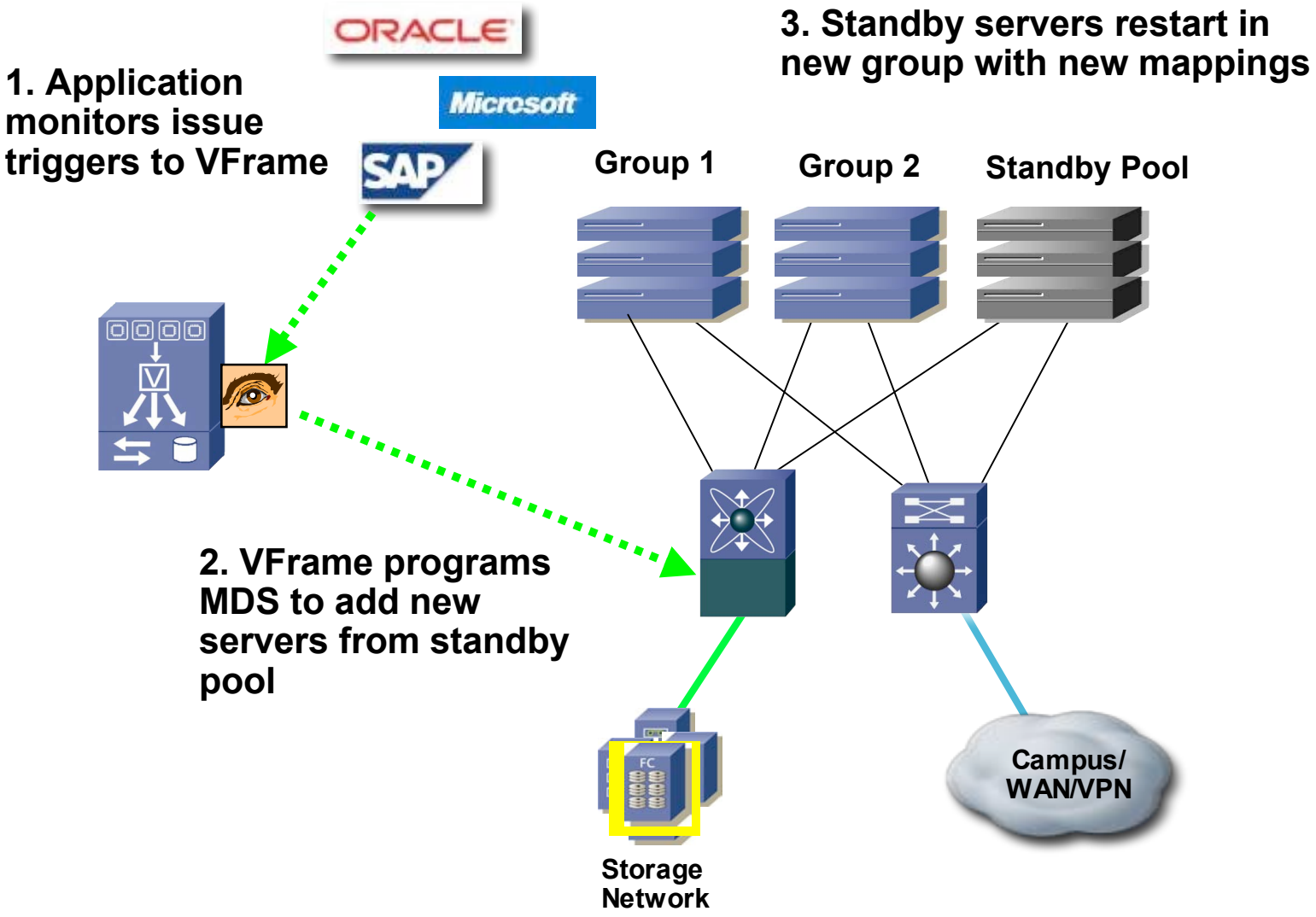
VFrame provisions security policies to FWSM

VFrame provisions CSM to add new server to load balancing pool

Utility Computing example
Server Failover

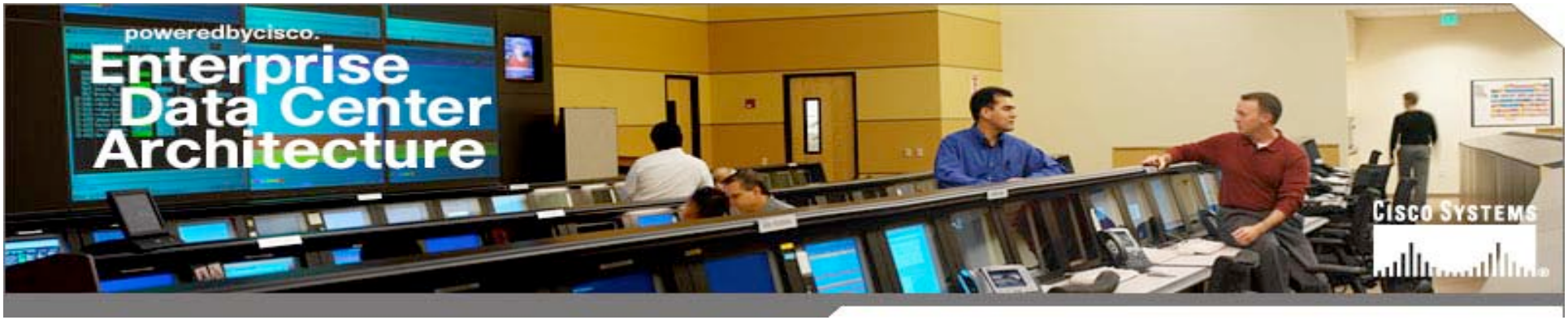


Utility Computing example
Add Capacity on Demand



VFrame Benefits

- Manage the data center from a service-oriented, application-centric perspective
- Eliminate number of layers/devices required to be touched to provision or modify
- Treat the entire data center infrastructure (from the "OS" to the "wire") as one manageable entity of shared virtualized resources (Virtual Mainframe)
- Expose a single orchestration and provisioning interface for all data center infrastructure
- Dramatically reduce TCO



Agenda:		
12.30-13.30	Registrazione e buffet lunch	
14.00-14.45	**Introduzione all'architettura Cisco Enterprise Data Center**	**Luciano POMELLI**
14.45-15.15	**Consolidamento e virtualizzazione dello storage**	**Davide CATTONI**
15.15-15.45	**Ottimizzazione delle prestazioni applicative**	**Andrea VERRI**
15.45-16.15	*Coffee break*	
16.15-16.45	**Server Networking e Virtual Data Center**	**Luciano Pomelli**
16.45-17.15	**Case study**	**Special Guest**
17.15-17.45	Q&A e conclusioni	