

# High Performance Computing

## State of the Industry 2005

Anil Vasudeva  
Principal Analyst & President  
[anil@imexresearch.com](mailto:anil@imexresearch.com)  
408-268-0800

*IMEX*  
RESEARCH.COM

©2000-2005 IMEX Research

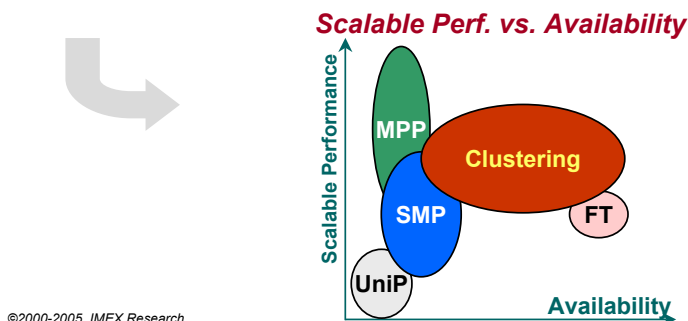
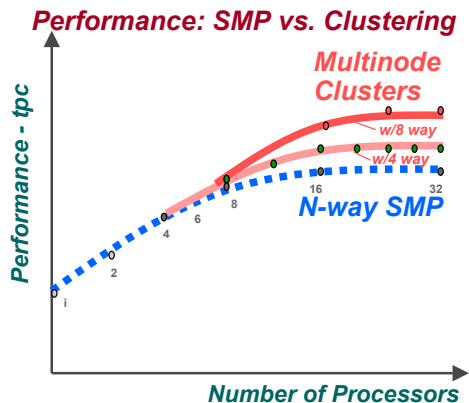
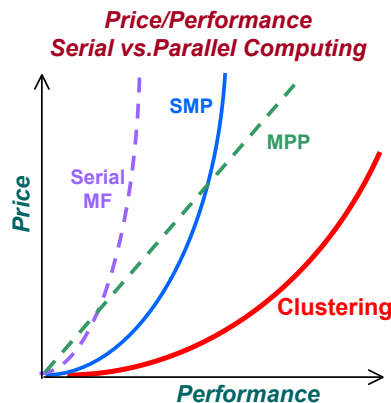
*IMEX*



*Your way to profitable technology markets.*

- **HPC - Markets Drivers & Industry Dynamics**
- **State of HPC Technologies**  
Clustering, Interconnects, Storage, Blade Servers
- **Industry Issues/Technology Hurdles**  
Standards, Management, Fabrics, Thermals, IP Acceleration
- **Futures Trends - Computing & Grids**
- **Holy Grails**  
Volume Driven Economics, IP Evolution, TF/Desktop  
HPC in our Daily Lives (Academia > Hollywood > Wall Street)

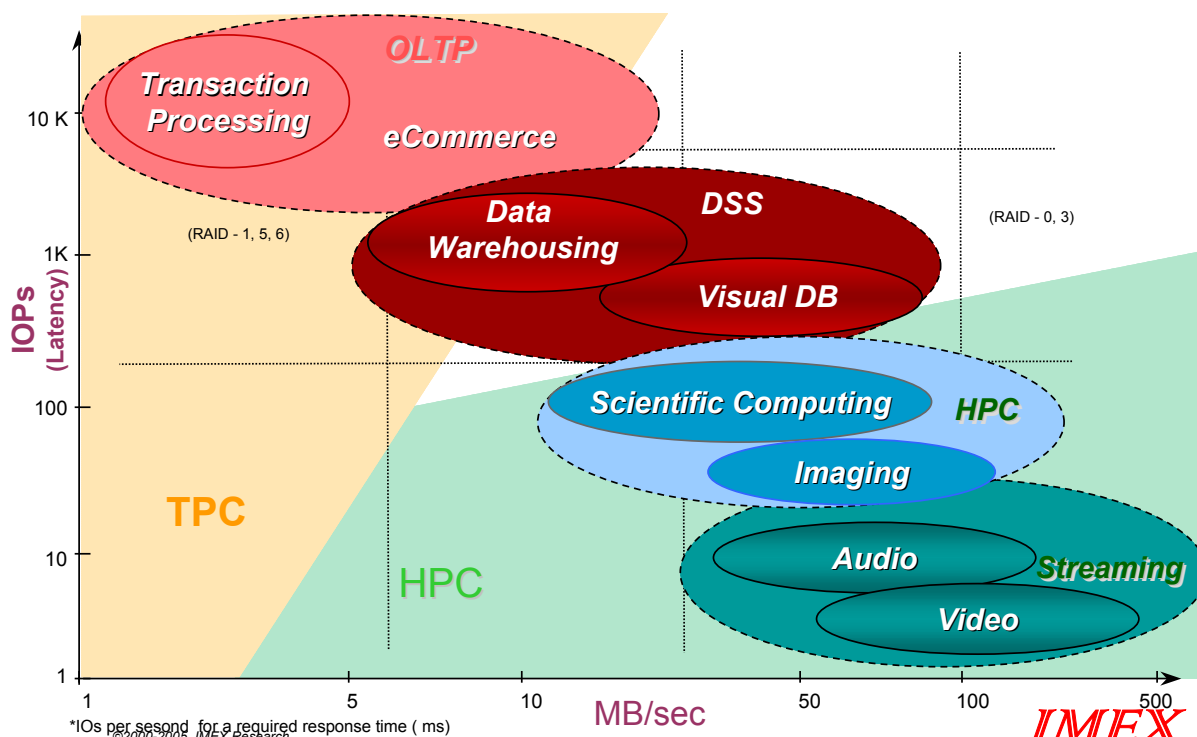
# Server Architectures: Competitive Technologies



©2000-2005 IMEX Research

IMEX

## Market Segments by Applications



IMEX

# ► HPC – 2 decades of rapid progress



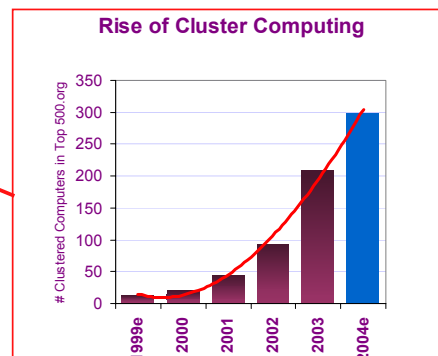
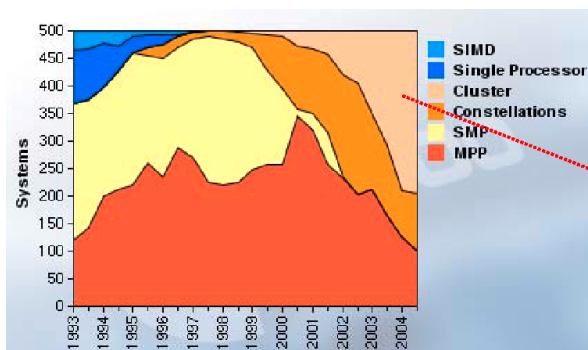
	1991	1998	2005	2010
System	Cray Y-MP C916	Sun HPC 10000	Shuttle@newegg	Blade
Architecture	16x Vector, 4 GB, Bus	24x 333 MHz, Ultra Sparc II, 24 GB, S-Bus	4x 2.6 GHz, x64 Bit, 4 GB, GbE	4x 10GHz, x64 bit, 16GB, 10GbE
OS	UNICOS	Solaris 2.5.1	Win Server 2003 SP2	LongHorn/Linux
Gflops	~10	~10	~10	~100
Price	\$40 million	\$1 Million	<\$4000	~\$1000
Price Reduction	1	1/40	1/10,000	1/400,000
Target Audience	Government National Labs	Large Enterprises	Small>Large Businesses	Every Professional
Applications	Classified, Climate, Physics	Manufacturing, Energy, Finance	Bioinformatics, Materials Sciences	Architects, Wall Street, Hollywood

©2000-2005 IMEX Research

**IMEX**

## ► The rapid rise of Clusters in HPC

	Ten years ago	Five years ago	Today
<b>Largest system</b>	143 Gflops	2.1 Tflops	70.7 Tflops
<b>Teraflop systems</b>	0	2	398
<b>Research/Academic</b>	60%	48%	41%
<b>Industry</b>	24%	46%	55%
<b>Linux clusters</b>	0	6	294



©2000-2005 IMEX Research

**IMEX**

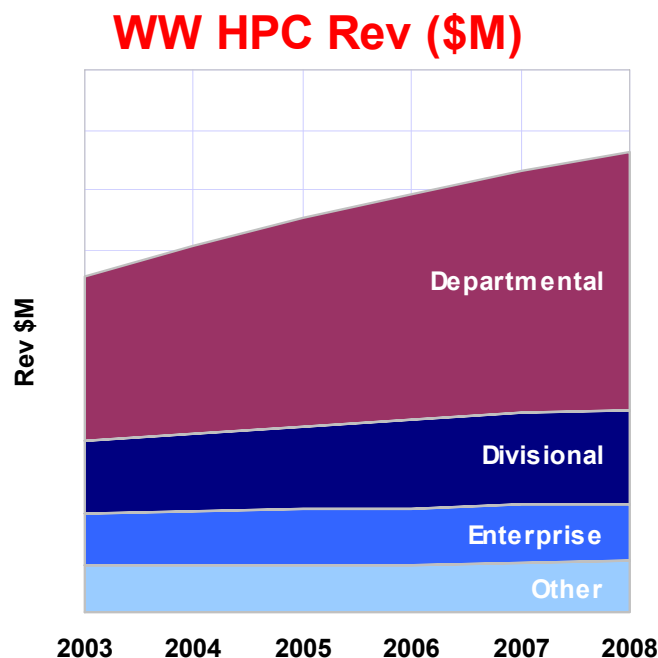
## ► HPC Market Drivers

- **Availability of industry standard HW making HPC mainstream**
  - HSV Servers @\$200, Storage @25c/GB, Add-On GbE NICS at \$10, GbE Switches at \$2-5/port
  - Migration to Embedded ASICs on Blades
- **Rapid procurement, installation & System Integration simplifying acquisition**
- **Cluster ready apps accelerating market growth**
  - Engineering (LSTC, Ansys..)
  - Bioinformatics (Blast, Gaussian..)
  - Finance (Matlab, Excel...)
  - Oil & Gas (Eclipse, Promagic..)
  - Government (Uexplore, NOAA Hysplit..)

©2000-2005 IMEX Research

**IMEX**

## ► HPC Servers Rev - 11% of Total Server Market

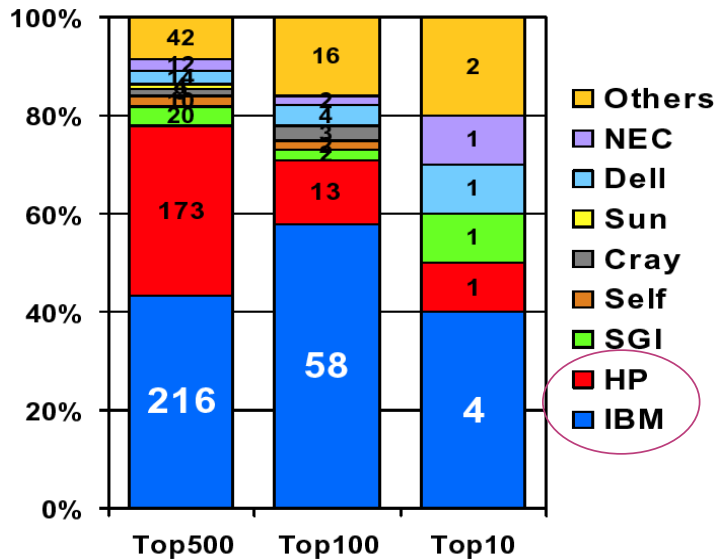


©2000-2005 IMEX Research

**IMEX**

## Market Leaders

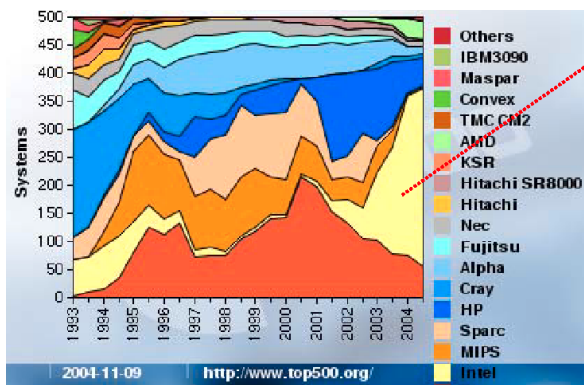
Top500 November 2004



©2000-2005 IMEX Research

IMEX

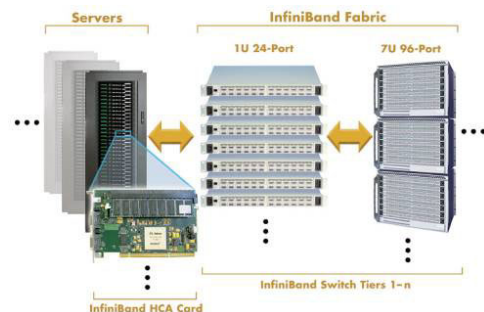
## ► Rise of Industry Standard Architecture in HPC



X86 Architecture (Intel & AMD) using Linux and Microsoft based clusters dominate the Processor/System, followed by HP in High Performance Computing Architecture.

### TeraFlop HPC Using Off-the-Shelf Modules

Intel servers with PCI Express, Infiniband HBAs & Switches, SATA/SAS HDD or AMD Opteron Servers, GbE HBA/Switches, SATA



©2000-2005 IMEX Research

IMEX

# ► Factors Affecting Performance

- **Applications Configurations affect Linpack significantly**

(3 metrics have the most affect)

- Problem Size
- Size of Blocks
- Topology

- **Tightly Coupled MPI Applications**

- Very sensitive to network performance characteristics
  - internodal communications delay
  - OS Network Stack is a significant factor
- Very sensitive to mismatched node performance
  - Random OS activities can add msec delays to usec type communication line delays

©2000-2005 IMEX Research

**IMEX**

# ► HPC Architecture & Challenges

## Architecture Goals & Criteria

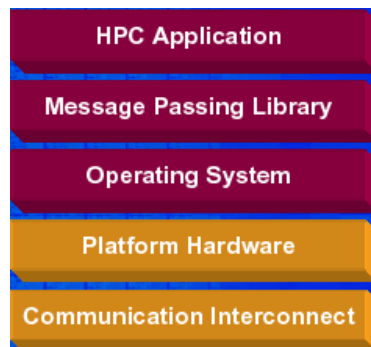
### Goals

- Maximize compute time vs. messaging time
- High constant cross-sectional throughput
- Reduce or hide latency

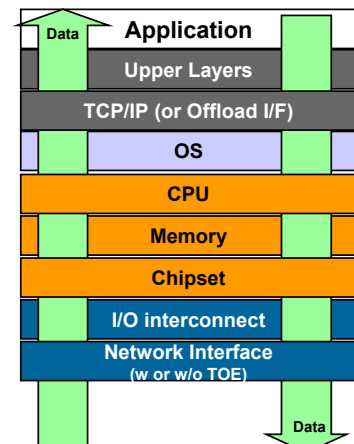
### Metrics

- Bandwidth
- Latency
- System Interface
- SW Stack

## HPC Communications Stack



## Entire Stack Needs Attention

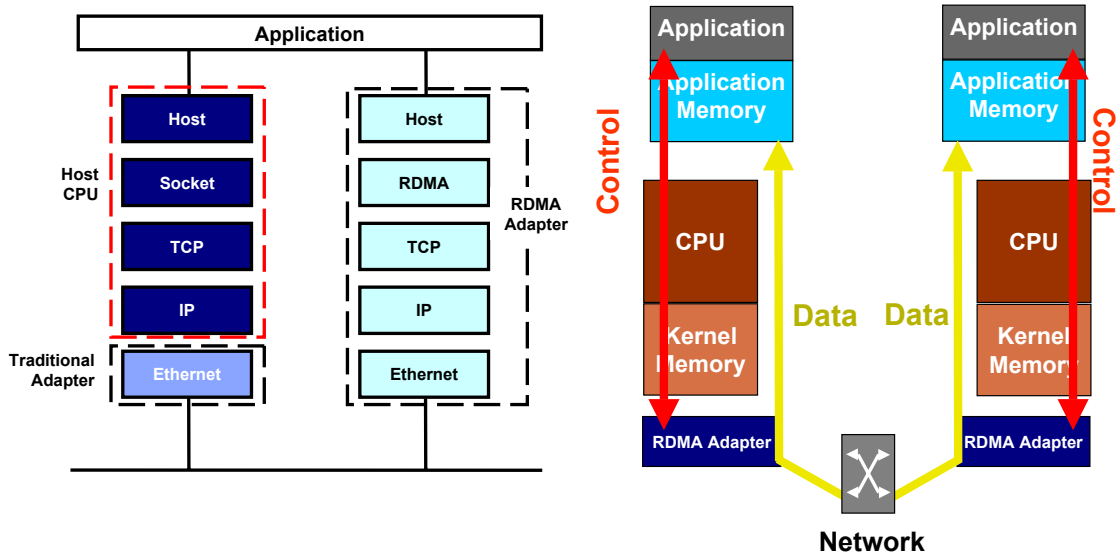


**Challenge is to move data from application to the network and back with minimum latency**

©2000-2005 IMEX Research

**IMEX**

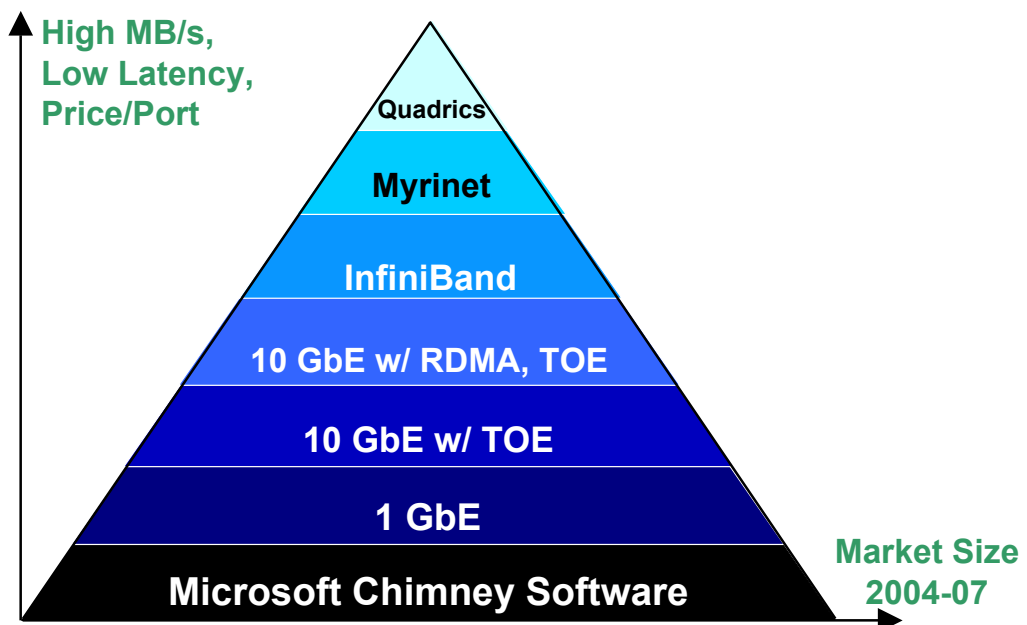
# ► HPC Architecture (TCP Bypass)



©2000-2005 IMEX Research

*IMEX*

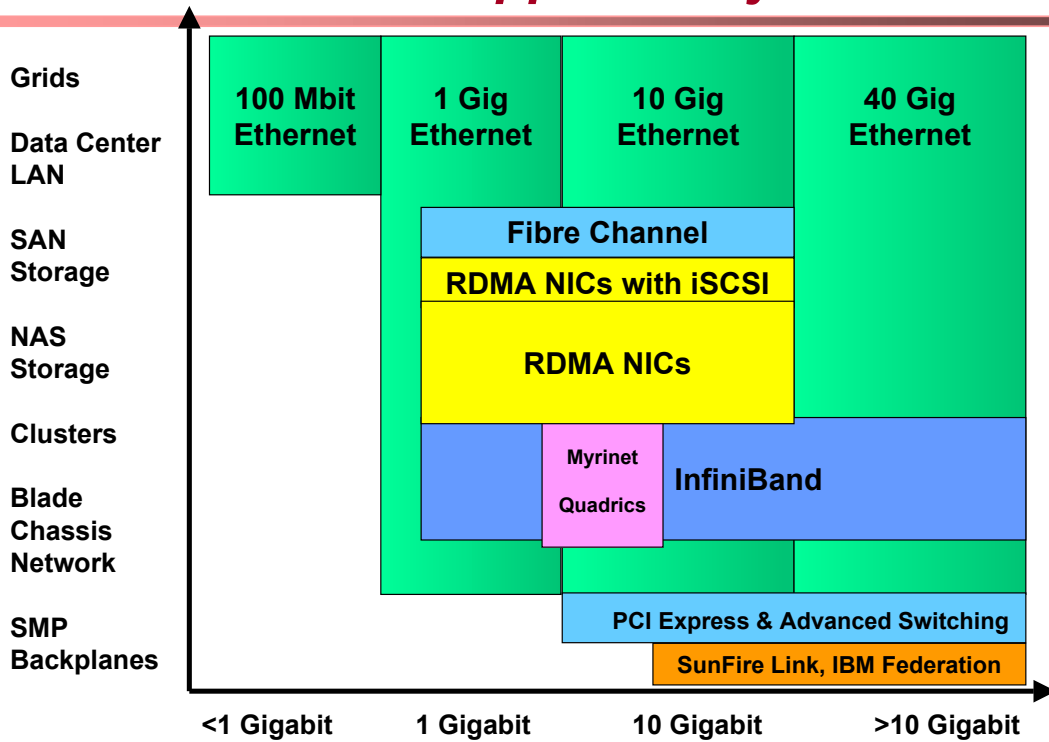
# ► Interconnect Price/Performance Hierarchy



©2000-2005 IMEX Research

*IMEX*

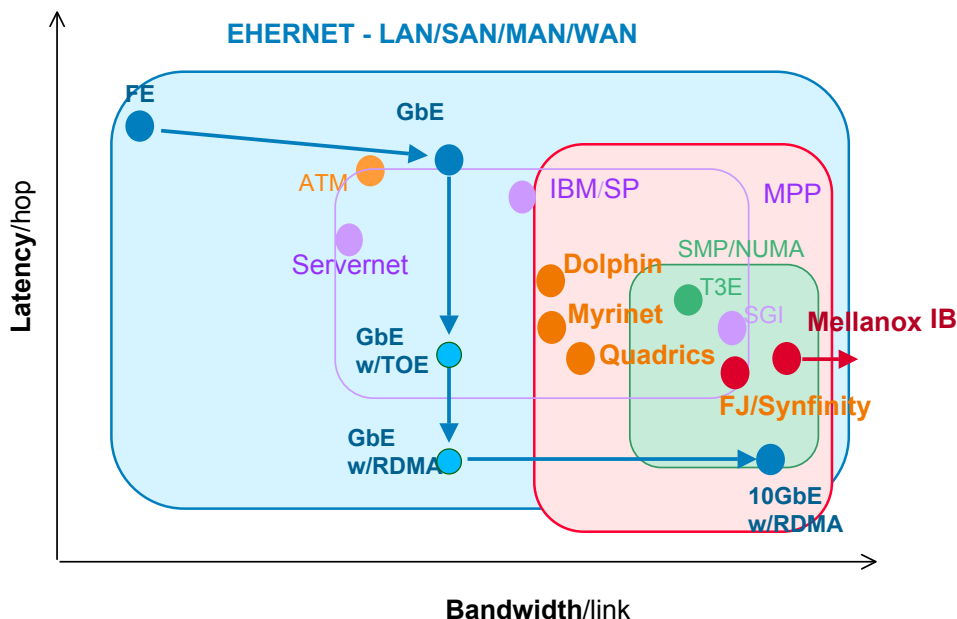
## Interconnects: Applicability



©2000-2005 IMEX Research

IMEX

## State of Interconnect Fabrics



©2000-2005 IMEX Research

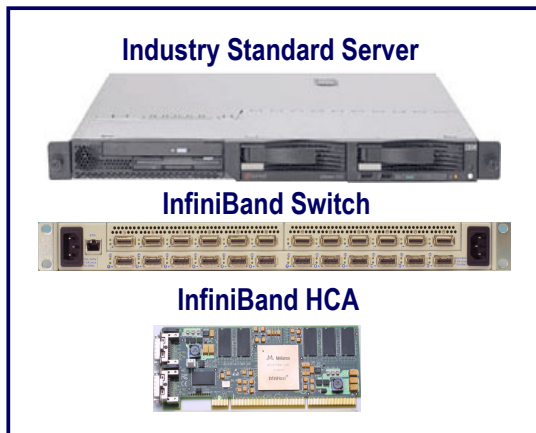
IMEX



# ► InfiniBand Clustering

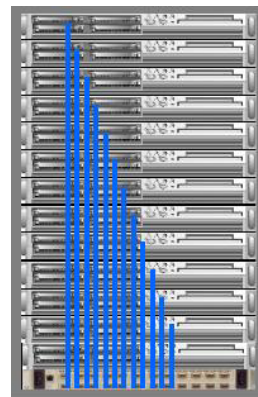
- **InfiniBand and Industry Standard Servers**

- Displacing Mainframe Computers in the Data Center with Three Standard Hardware Components
- Powerful Low Cost Combo Out Performs SMP



©2000-2005 IMEX Research

12-Nodes (24-CPU) Up to 72 GFlops of CPU Power; 24+ GB Memory



Enterprise Availability:  
With Fabric Failover

**IMEX**

# ► InfiniBand Adopters

- **Major providers of InfiniBand Solutions**
- **Enterprise Class Products**
  - InfiniBand to IP
  - InfiniBand to FC
  - Advanced Mgmt
- **Platform support includes**
  - Intel (IA32 + 64EMT, Itanium), AMD Opteron, Power PC, Sparc
- **Operating Systems**
  - Linux, Windows, AIX, HP/UX, Solaris, MAC OS X, VxWorks



NetworkAppliance®



Agilent Technologies



SANMINA-SCI



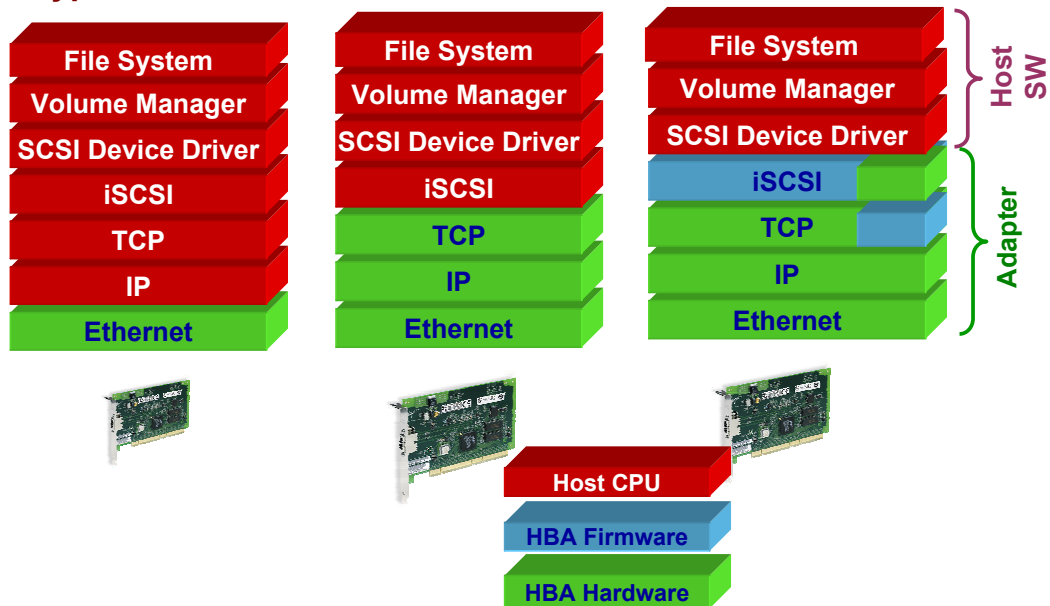
VOLTAIRE

**IMEX**

©2000-2005 IMEX Research

# ► Ethernet - TCP + iSCSI Offload Engines

## Typical 1GbE NIC

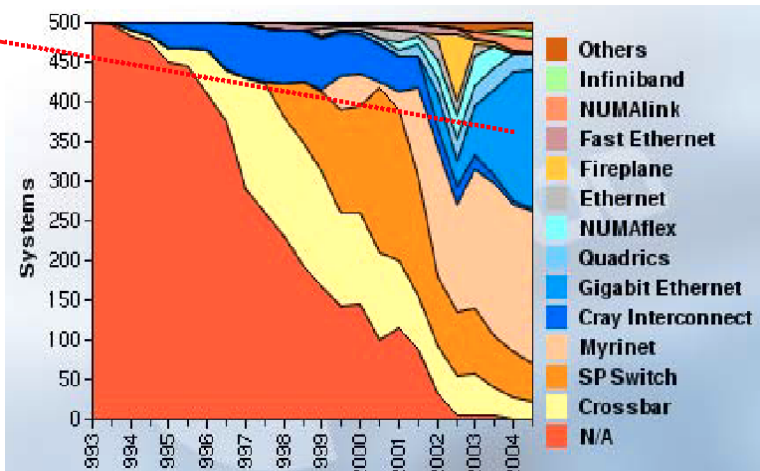


©2000-2005 IMEX Research

IMEX

# ► HPC Interconnect – Leaders

Myrinet for the highly latency sensitive applications and GbE for the majority applications Dominate the Interconnect for HPC. Infiniband is rearing to dominate at the Midrange latency applications



©2000-2005 IMEX Research

IMEX

# ► HPC Tools Availability

## ■ System Hardware

- Servers, blades, clusters & workstations
  - POWER, Intel, AMD Opteron
- Supercomputers
- Visualization/rendering
- High performance interconnects
  - e.g., HPS, Myrinet, Infiniband
- Storage systems, storage virtualization

## ■ System Software

- Linux, Unix, Windows
- System & storage management
  - e.g., CSM, PSSP, GPFS
- DB2, WebSphere, Tivoli, Rational
- Grid & on demand middleware

## ■ Special-purpose Systems

- Blue Gene/L
- Gov't & research partnerships (e.g., ASCI)

## ■ Applications & Tools

- ISV & in-house software
- Compilers, schedulers, libraries, tools
  - 3rd party, ESSL/PESL, LoadLeveler
- Open source & public domain codes
- ACTC tools

## ■ Services

- IGS services practices & consulting
- Hosting & utility services
- IBM Global Financing

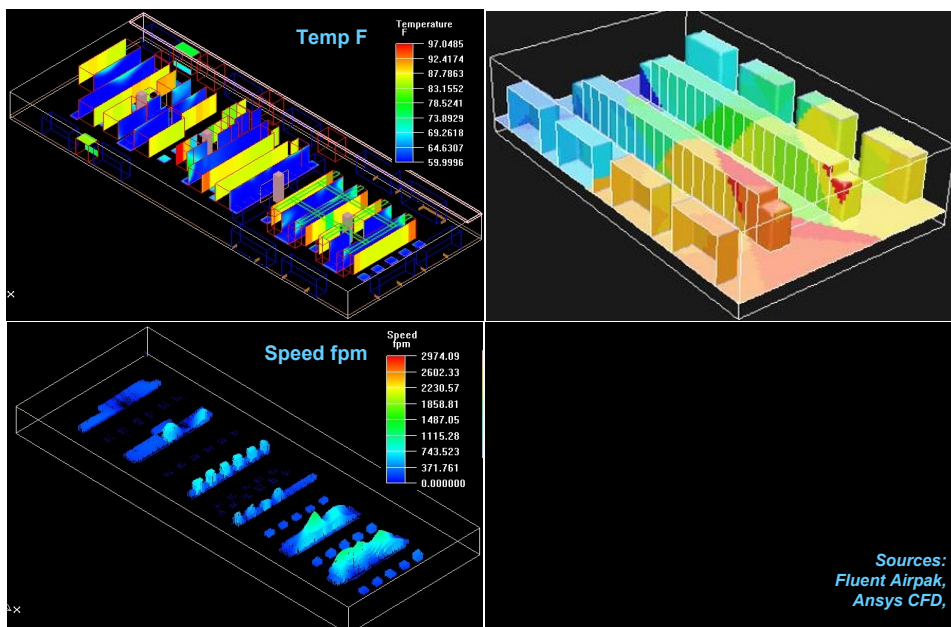
## ■ Solutions

- Deep Computing Capacity on Demand
- Departmental Supercomputing Solutions
- Visualization framework
- Infrastructure solutions (e.g., Grid)
- Industry-specific solutions (e.g. AEIF)

©2000-2005 IMEX Research

**IMEX**

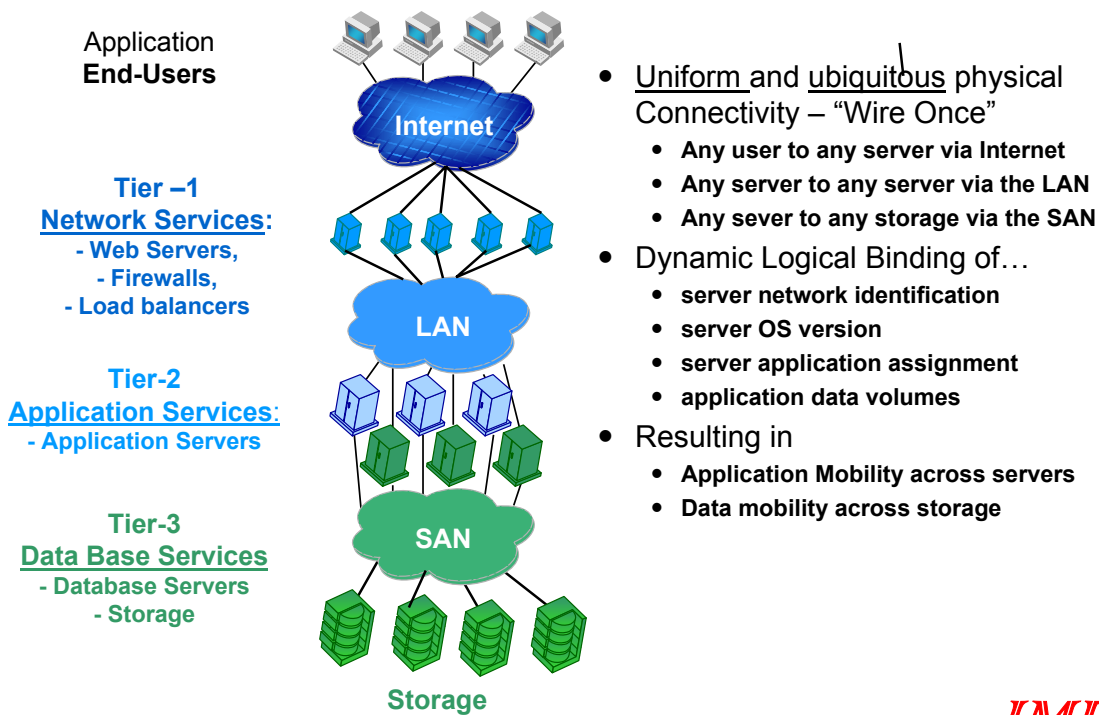
# ► Data Center Environmentals Prep



©2000-2005 IMEX Research

**IMEX**

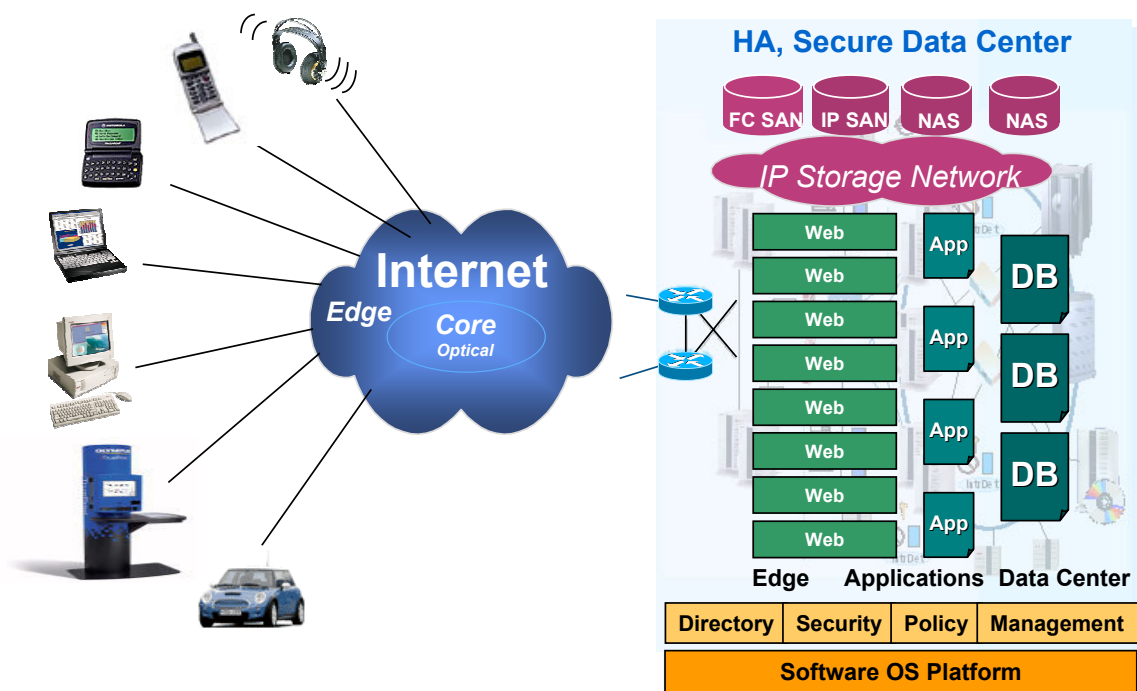
# ► 3 Tier Computing Infrastructure



©2000-2005 IMEX Research

**IMEX**

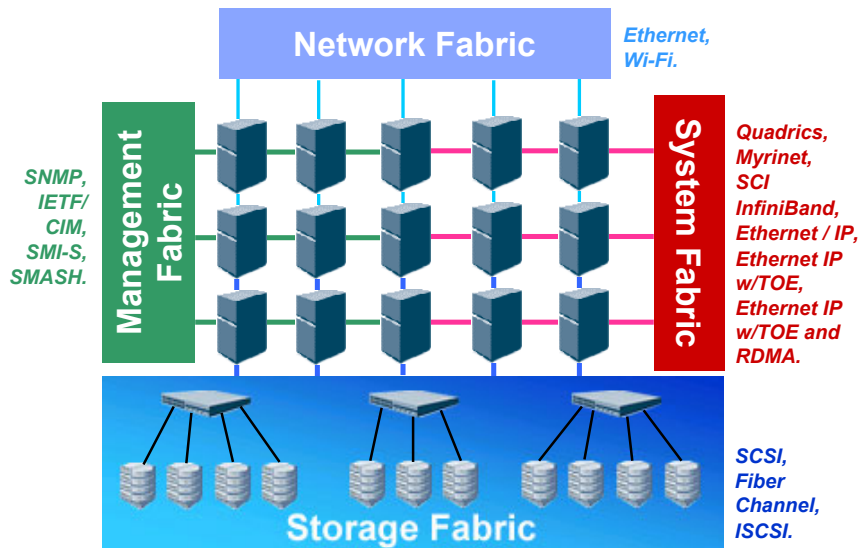
# ► E2E Internet & 3 Tier Computing



©2000-2005 IMEX Research

**IMEX**

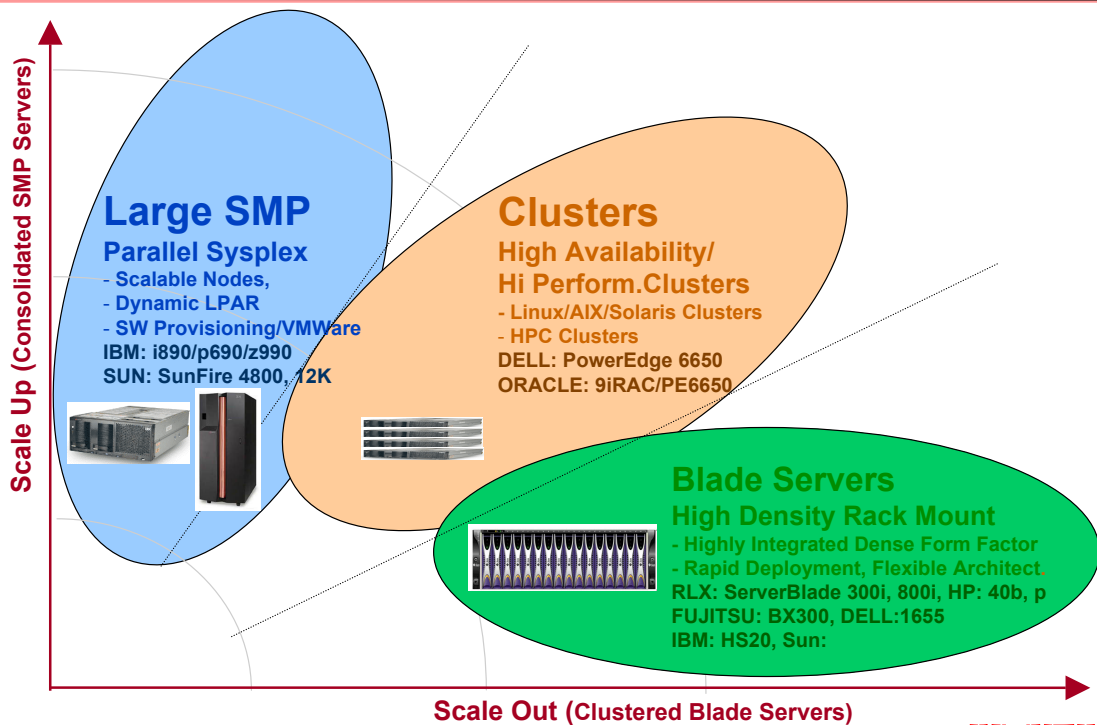
# ► Fabric based Clusters Architecture



©2000-2005 IMEX Research

**IMEX**

# ► Server Platforms - Migrations



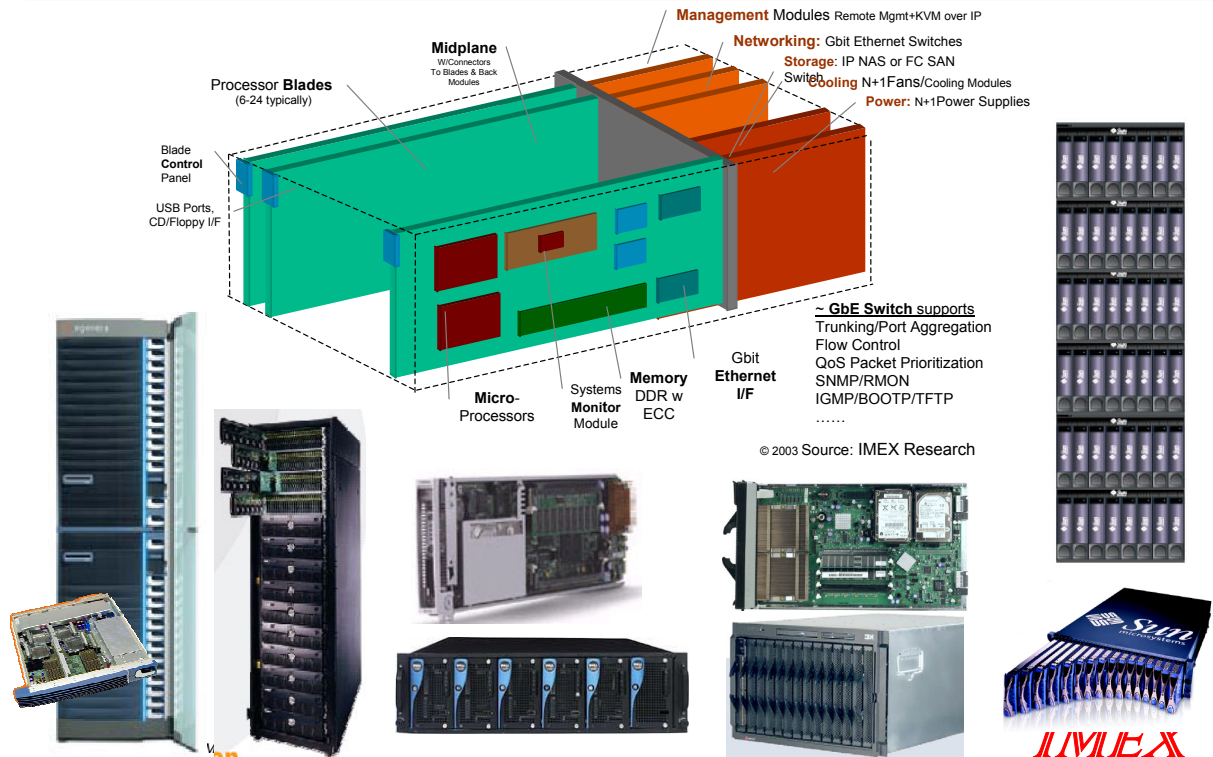
©2000-2005 IMEX Research

**IMEX**

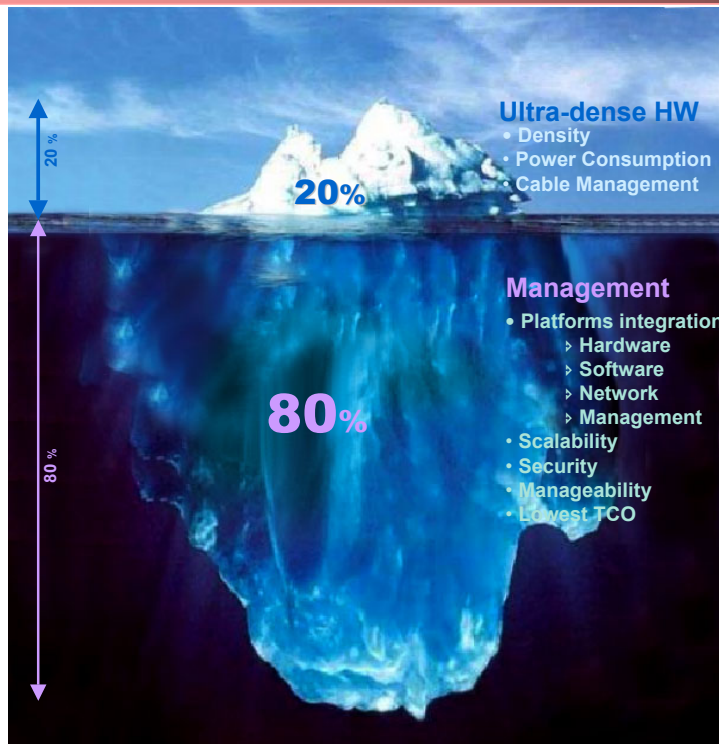




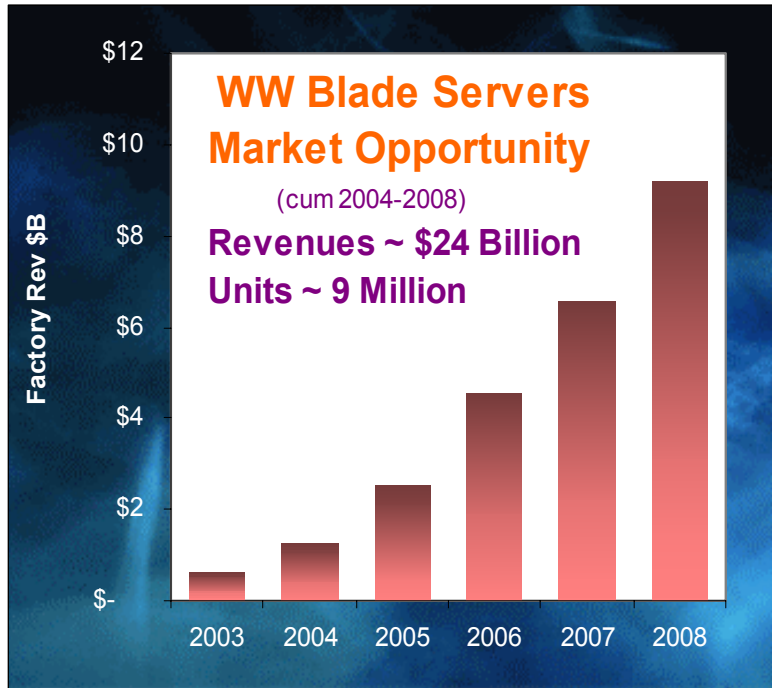
# Blades Servers - Infrastructure



## Management – Blade Servers



# ► Blade Servers Market Opportunity

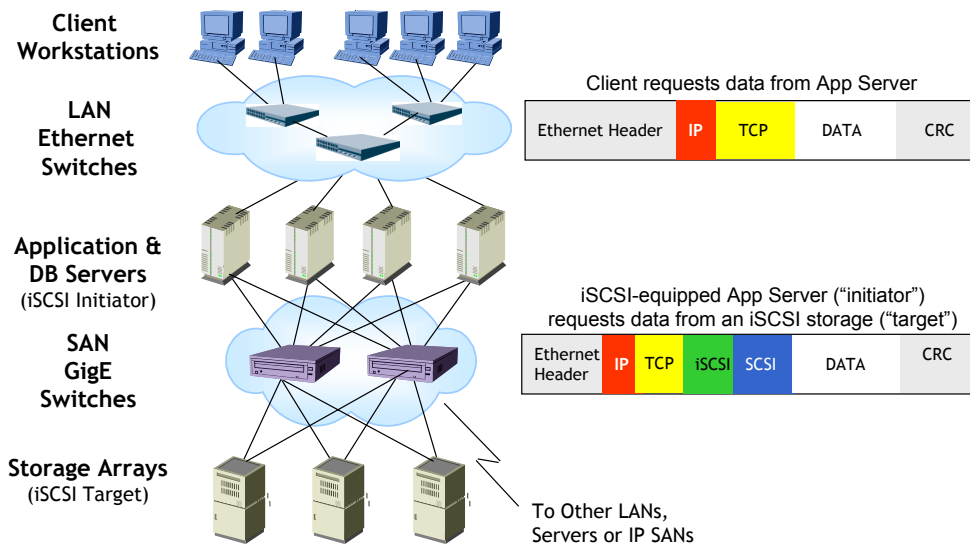


©2000-2005 IMEX Research

Data: IMEX Research 2005

**IMEX**

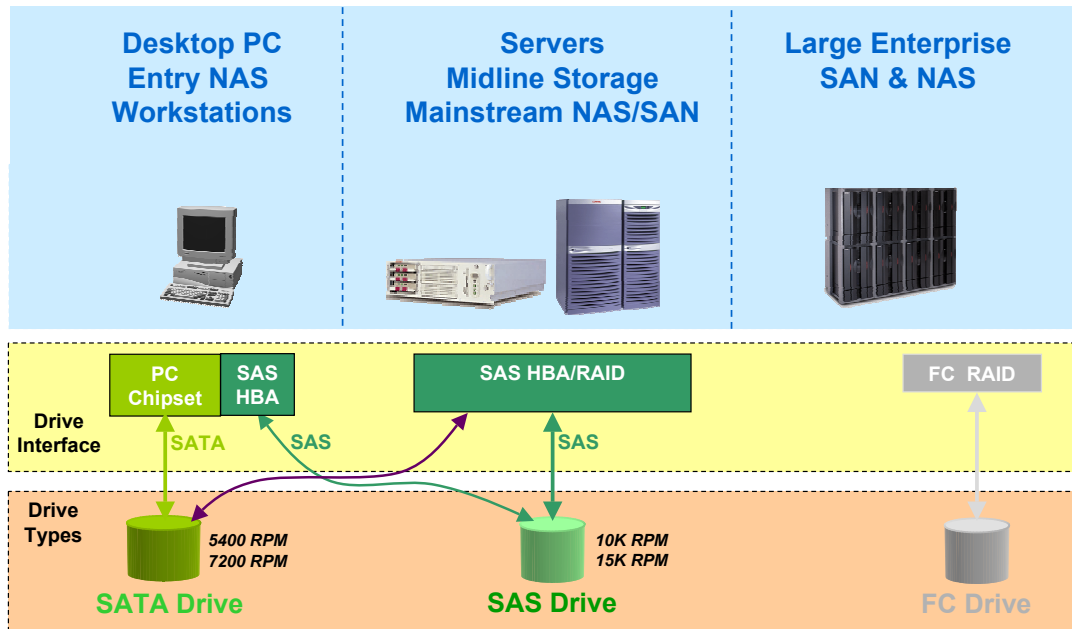
## ► The iSCSI SAN



©2000-2005 IMEX Research

**IMEX**

## ► Tiered Storage by Price/Performance



©2000-2005 IMEX Research

**IMEX**

## ► Grid Computing

- **Grid is a Catch-all Marketing Term**
  - Means different things to different constituencies
  - Desktop Cycle-Stealing
  - Managed HPC Clusters
  - Virtualization of Data Center Resources
  - Outsourcing to "Utility Data Centers"
- **Multiple, opposing requirements**
  - Compute-intensive applications want to ship data to idle processor resources
  - Data-Intensive applications want to ship computations to appropriate large data repositories
- **An evolving entity with many pulls & pushes**
  - Economics, Feasibilities, management of hetero vs homogeneous HW

©2000-2005 IMEX Research

**IMEX**



# ► Grid Challenges: Hype vs. Reality

- **Cost of Local Grids (LAGs\*) – ala Blades cheaper than Wide Area Grids (WAGs\*)**
  - Computing Cost \$1000 > 1 cpu day (10 Terops) = \$1
  - 10 TB network transfer Cost = \$1
  - Internet BW cost \$100/mbps/mo > 1GB network xfr cost ~\$1
  - Result: Local HPC Cluster is 10,000 cheaper than WAN Communication
- **MPI Style Apps work well in LAN Clusters but uneconomical in WAN**
  - Data analysis best done by moving programs to data not data to programs
  - Small data, high compute applications work well across internet
  - Internet is not the CPU back plane

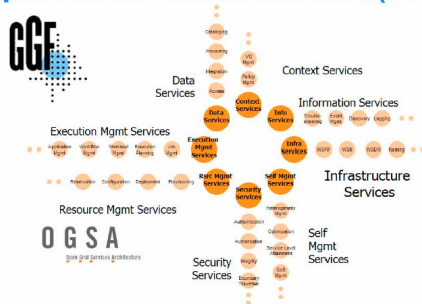
\* Copyright 2005 IMEX Research & IMEX International

©2000-2005 IMEX Research

IMEX

# ► Grid Services Architecture & Standards

## Open Grid Services Architecture (OGSA)



## Open Standards & Implementations Are Moving Forward

### •WS-RF/Notification

- Websphere Emerging Technology Toolkit (ETTK)  
*Permits early implementation of WS-RF / Notification using Websphere hosting*
- Small-footprint prototype of WS-RF / Notification implementation in a portable device (Blackberry)  
*Demonstrates the scalability and ease of implementation of WS-RF / Notification*
- Open source project implementation as part of Apache/Axis  
*Open source implementation of WS-RF / Notification in open-source / lightweight hosting environment*
- Implementation of WS-RF / Notification in next full release of Websphere  
*Integrated with Websphere programming models & development tools*
- Cooperation between Web Services Standards bodies  
*MOU on cooperation around Web Service Management*

alphaWorks

Apache

Websphere

OASIS DMTF W3C IETF

### •Open Grid Service Architecture

- OGSA has become the "flagship" architecture of the Global Grid Forum
- OGSA "Profiles" are being specified which relate OGSA to other web services specifications  
*WS-RF/N, WS-Security, WS-I, WS-Addressing etc.*
- Higher level concrete specification are being developed for important Grid functions  
*Basic Execution Services, Data Services, etc.*



- OASIS
  - WSRF and WS-Notification
  - WS-Security, etc.
  - WSDM
- DMTF
  - Utility Computing
  - CIM
  - Server management
- W3C
  - WS-Addressing
  - Various others

©2000-2005 IMEX Research

IMEX

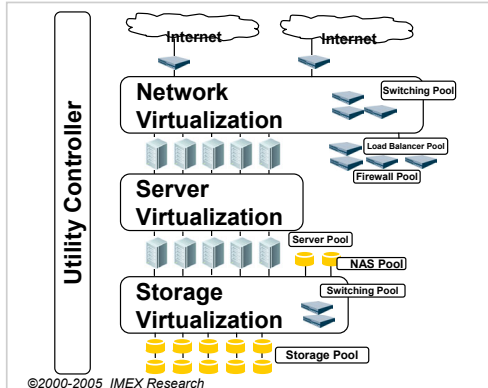
# ► Utility Computing/Web Services

## What a Computer Room will Look Like a Year from Now



## Autonomics

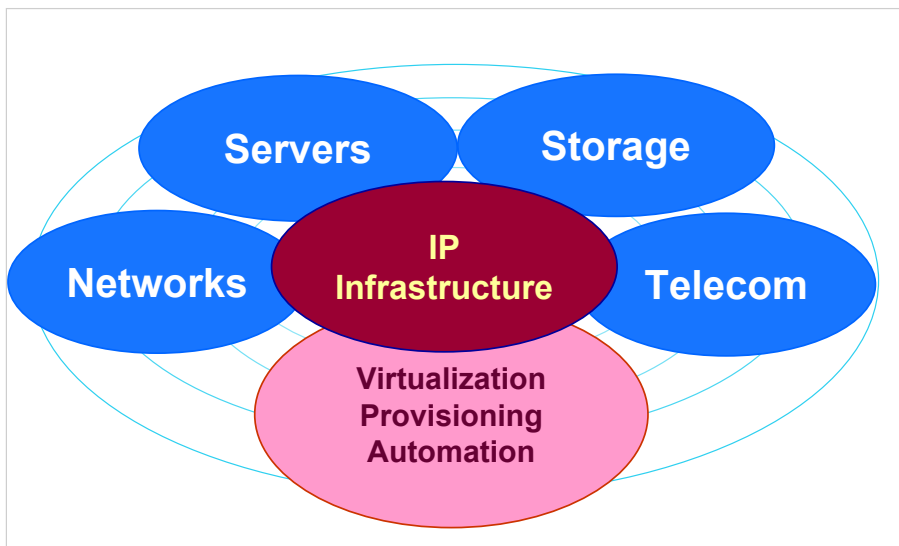
- Self Healing
- Self Optimizing
- Universal Identity
- Single System Image
- ....



- **HP** - Utility Data Center
- **Sun** - N1/SunONE/Orion
- **Microsoft** - .Net
- **IBM** - OnDemand/Autonomics

**IMEX**

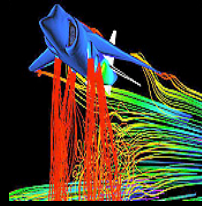
# ► Future - IP Based Infrastructure



**IMEX**

# HPC – From Academia to Wall St to Hollywood

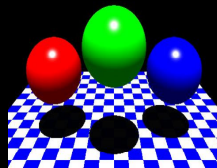
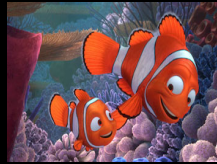
## High Performance Computing



▶ 100+ Teraflops

▶ Throughput = 100 GB/s

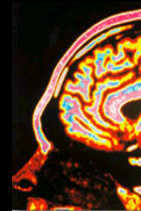
## Commercial Visualization



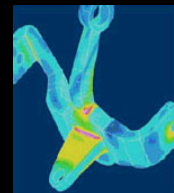
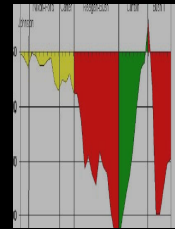
▶ Rendering (Texture & Polygons)

▶ Throughput = 1.2 GB/s

## Bioinformatics



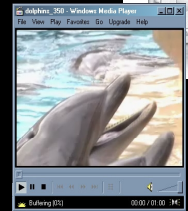
## Decision-Support Systems



▶ Data rate & capacity

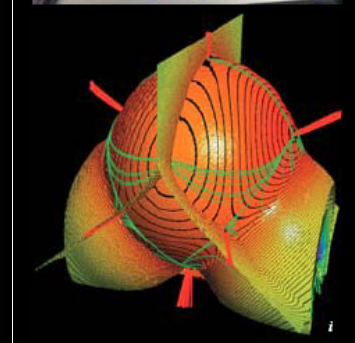
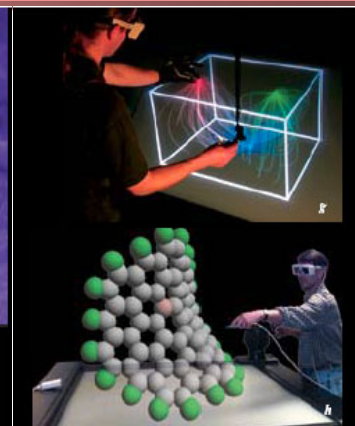
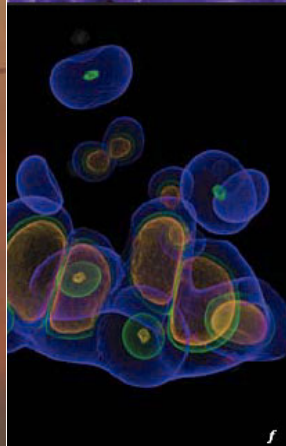
▶ Throughput : DSL/Cable

## Entertainment Audio/Video OnDemand



©2004–05 IMEX Research.com

## ▶ HPC: From Homeland Security to Bioinformatics





## ► *HPC From Academia to Hollywood*



## High Performance Computing State of the Industry 2005

Anil Vasudeva  
Principal Analyst & President  
[anil@imexresearch.com](mailto:anil@imexresearch.com)  
408-268-0800

*IMEX*  
RESEARCH.COM