High Performance Computing

State of the Industry 2005

Anil Vasudeva
Principal Analyst & President
anil@imexresearch.com
408-268-0800

IMEX RESEARCH.COM

©2000-2005 IMEX Research

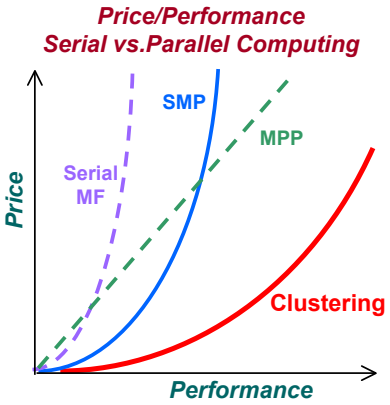
Your way to profitable technology markets.

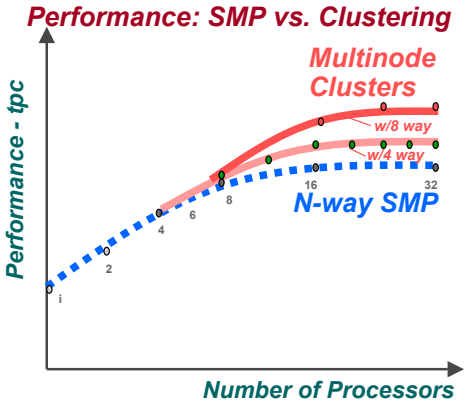
- **HPC - Markets Drivers & Industry Dynamics**
- **State of HPC Technologies**
  Clustering, Interconnects, Storage, Blade Servers
- **Industry Issues/Technology Hurdles**
  Standards, Management, Fabrics, Thermals, IP Acceleration
- **Futures Trends - Computing & Grids**
- **Holy Grails**
  Volume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

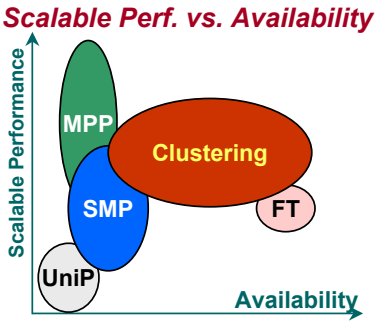
©2000-2005 IMEX Research

IMEX

Server Architectures: Competitive Technologies



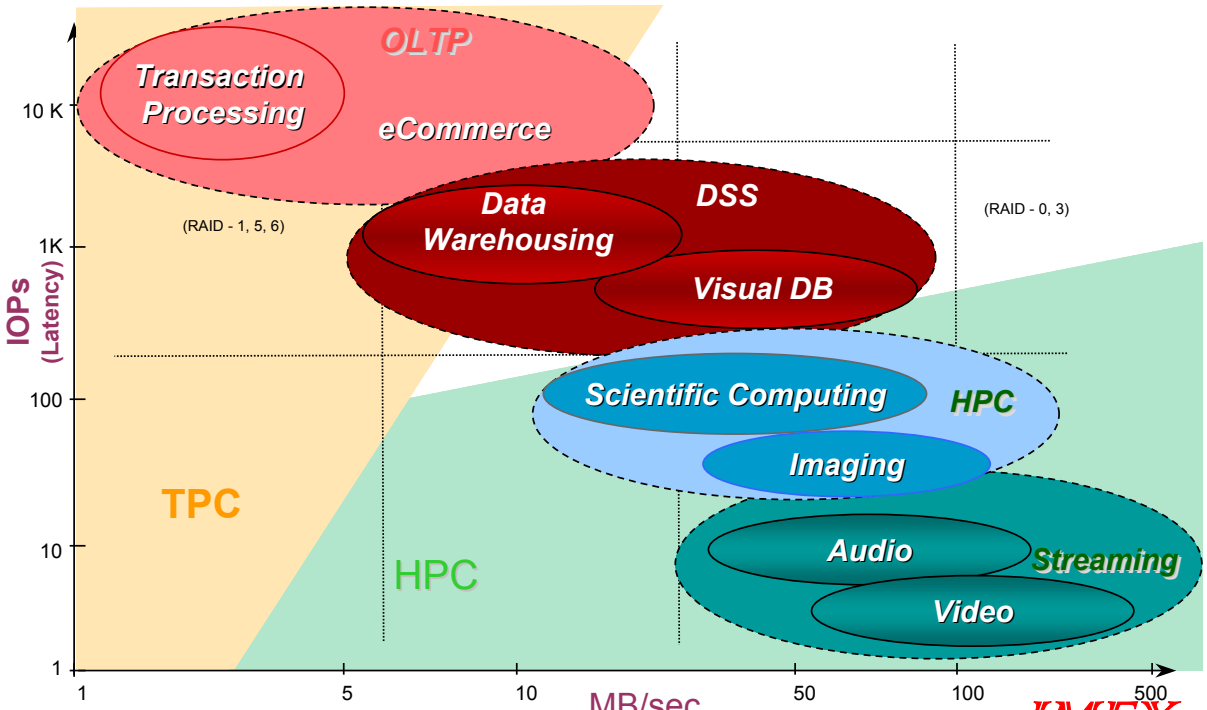




©2000-2005 IMEX Research

IMEX

Market Segments by Applications



*IOs per second for a required response time (ms)

©2000-2005 IMEX Research

IMEX

HPC – 2 decades of rapid progress

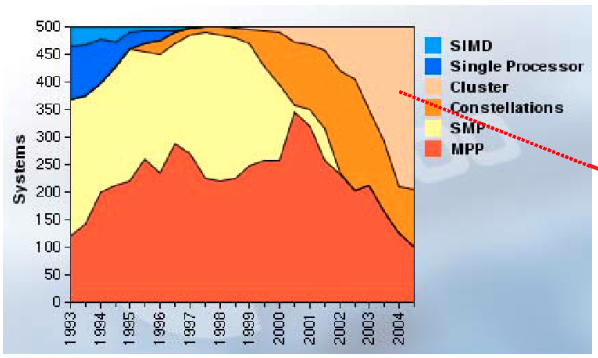
	1991	1998	2005	2010
System	Cray Y-MP C916	Sun HPC10000	Shuttle@newegg	Blade
Architecture	16x Vector, 4 GB, Bus	24x 333 MHz, Ultra Sparc II, 24 GB,S-Bus	4x 2.6 GHz, x64 Bit, 4 GB, GbE	4x 10GHz, x64 bit, 16GB, 10GbE
OS	UNICOS	Solaris 2.5.1	Win Server 2003 SP2	LongHorn/Linux
Gflops	~10	~10	~10	~100
Price	$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

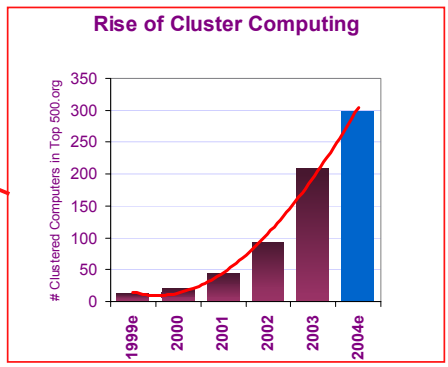
©2000-2005 IMEX Research

IMEX

The rapid rise of Clusters in HPC

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





©2000-2005 IMEX Research

IMEX

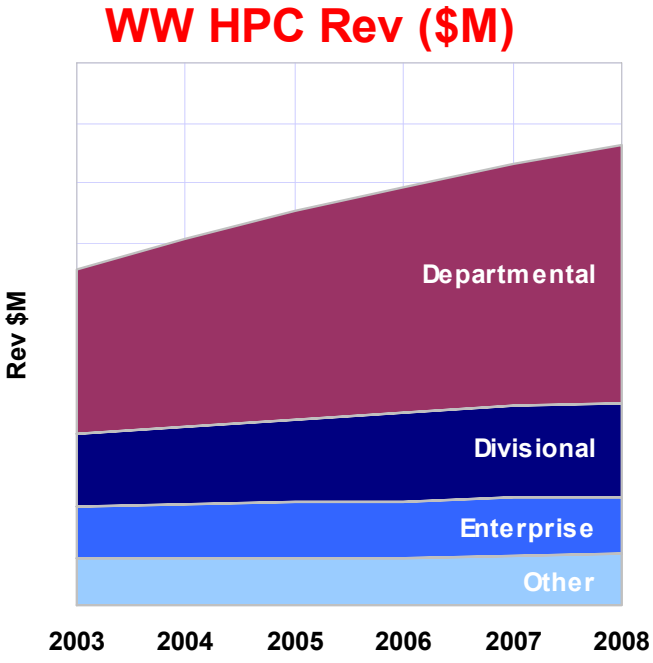
HPC Market Drivers

- **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

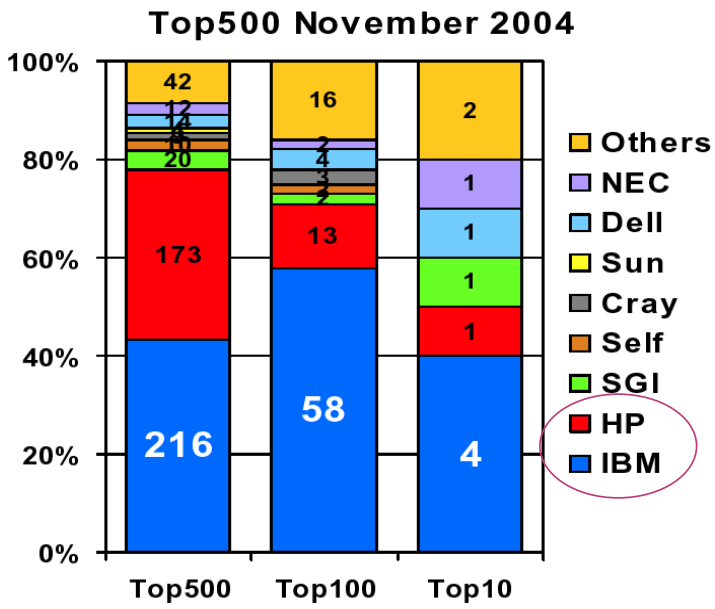


©2000-2005 IMEX Research

IMEX

HPC Market Leaders 2004

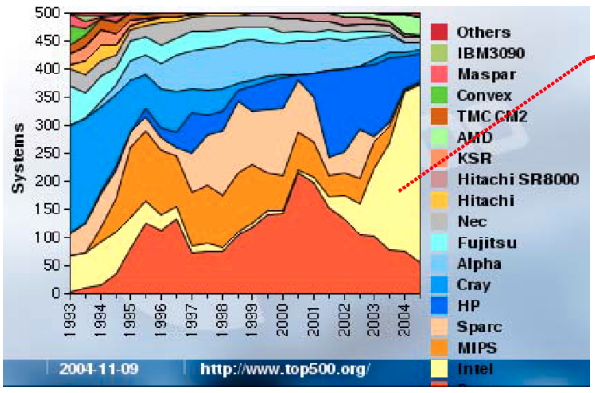
Market Leaders



©2000-2005 IMEX Research

IMEX

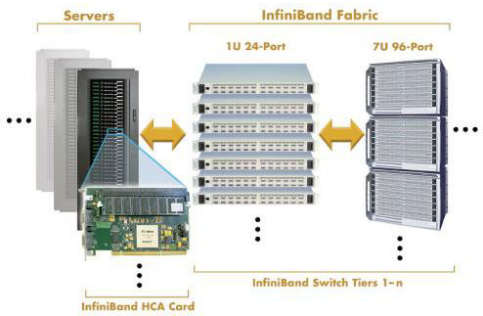
Rise of Industry Standard Architecture in HPC



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

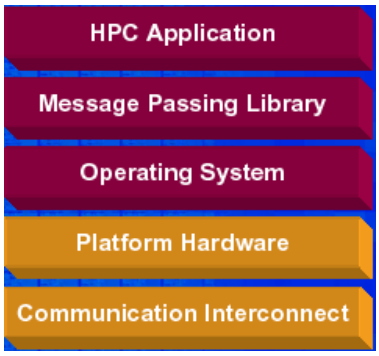
Goals

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

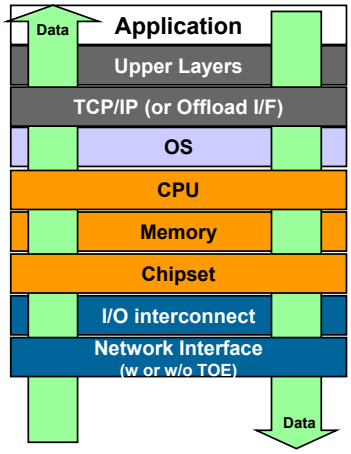
Metrics

- Bandwidth
- Latency
- System Interface
- SW Stack

HPC Communications Stack



Entire Stack Needs Attention

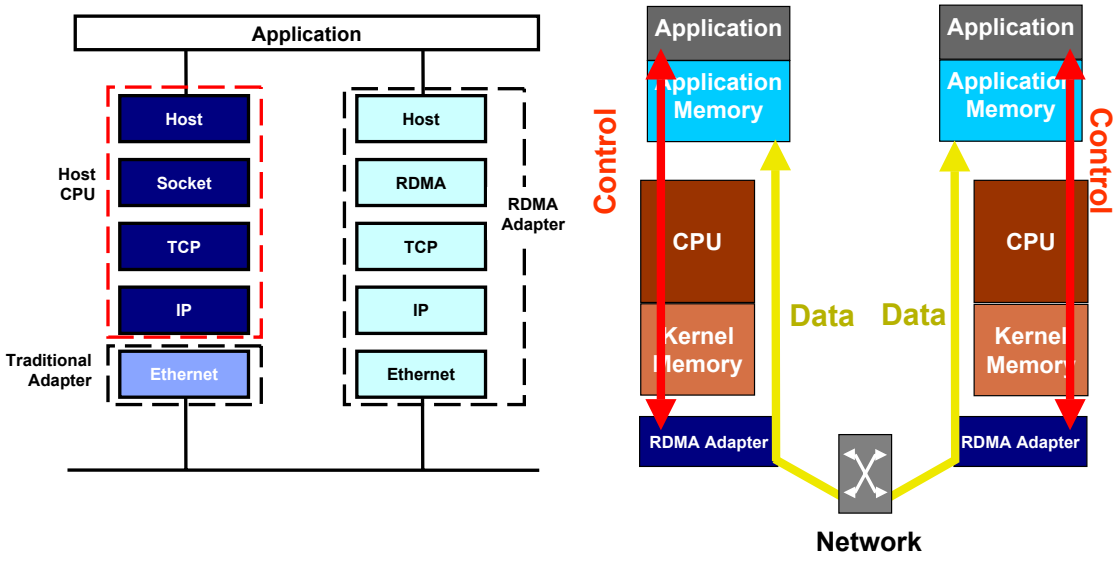


Challenge is to move data from application to the network and back with minimum latency

©2000-2005 IMEX Research

IMEX

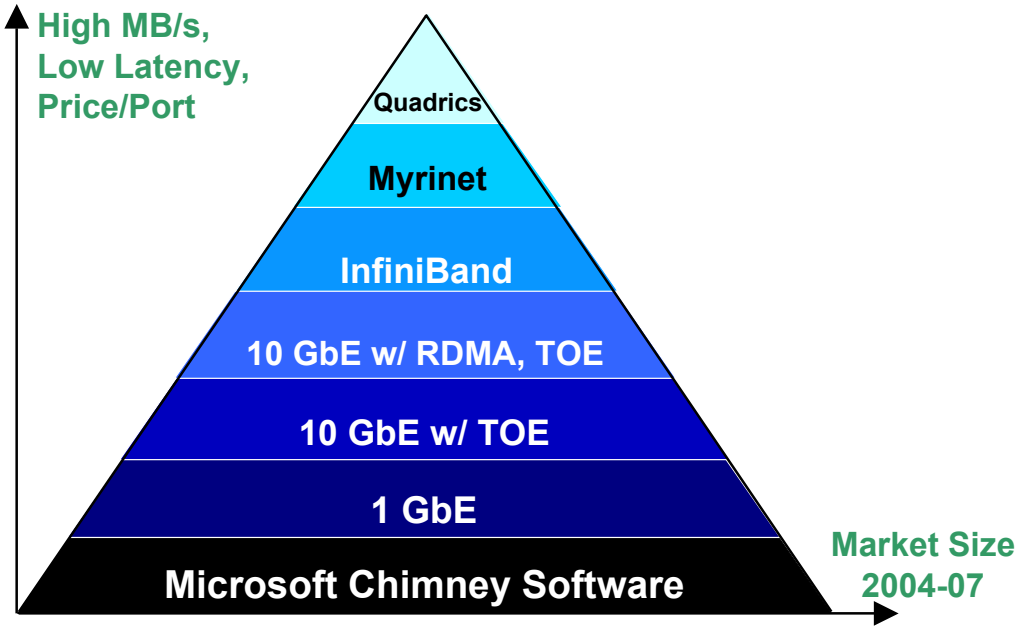
HPC Architecture (TCP Byepass)



©2000-2005 IMEX Research

IMEX

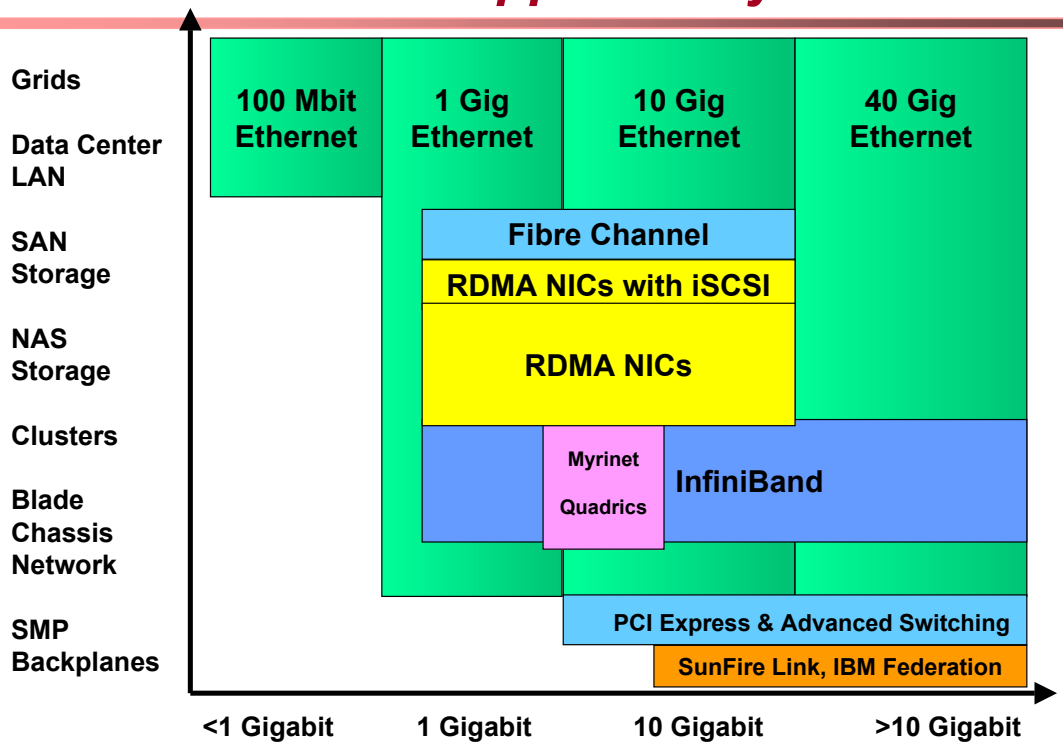
Interconnect Price/Performance Hierarchy



©2000-2005 IMEX Research

IMEX

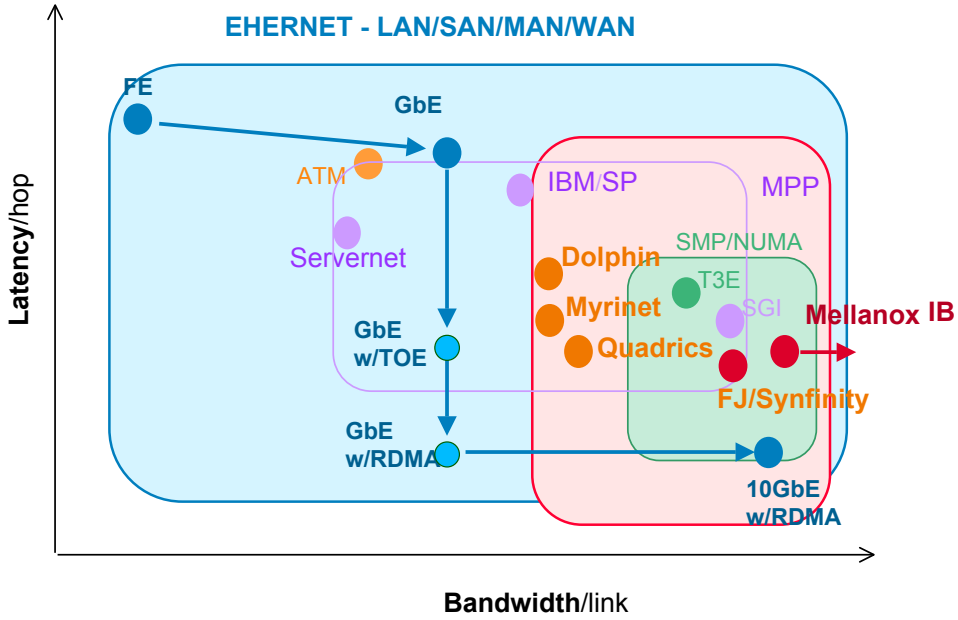
Interconnects: Applicability



©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

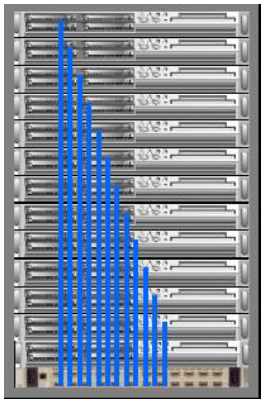
Industry Standard Server

InfiniBand Switch

InfiniBand HCA

=>

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



Enterprise Availability: With Fabric Failover

©2000-2005 IMEX Research

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, HPUX, Solaris, MAC OS X, VxWorks

DELL

NetworkAppliance

hp

Agilent Technologies

FUJITSU

InfiniCon systems

APPRO HPC Cluster Solutions

Linux Networx

IBM

SANMINA-SCI

NEC

HITACHI

SBS Technologies

sgi

TOPSPIN

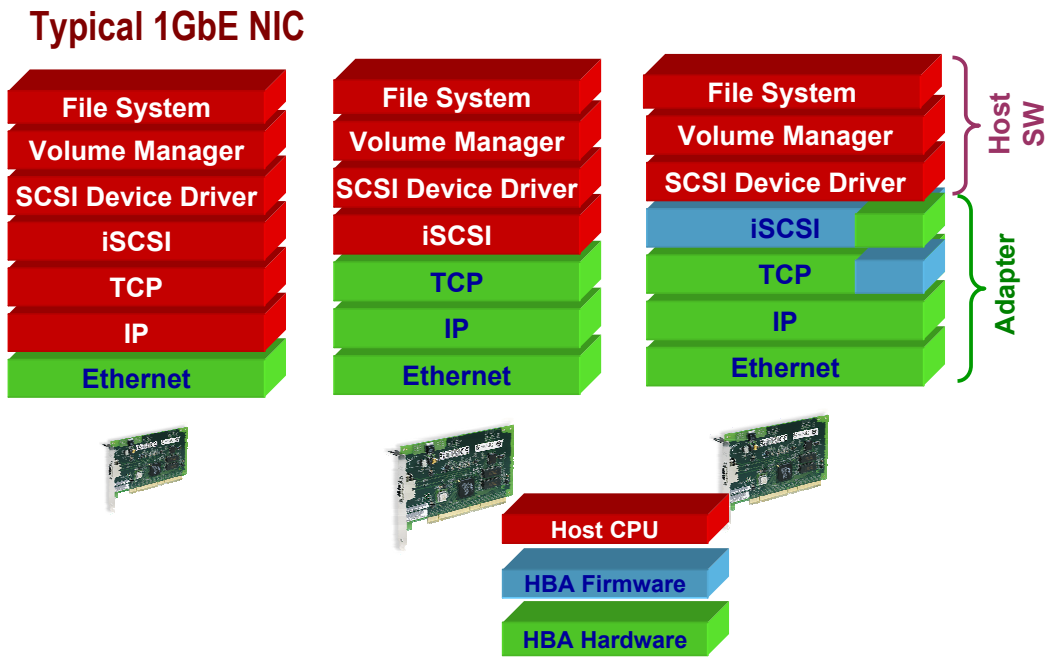
Sun microsystems

VOLTAIRE

©2000-2005 IMEX Research

IMEX

Ethernet - TCP + iSCSI Offload Engines

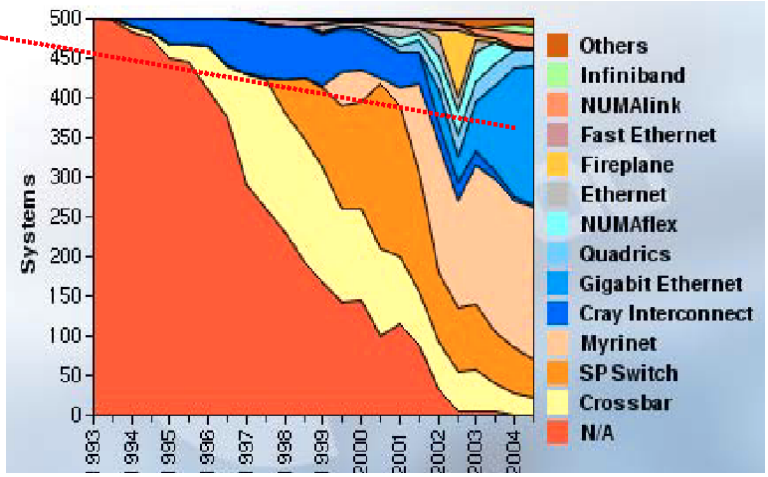


©2000-2005 IMEX Research

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

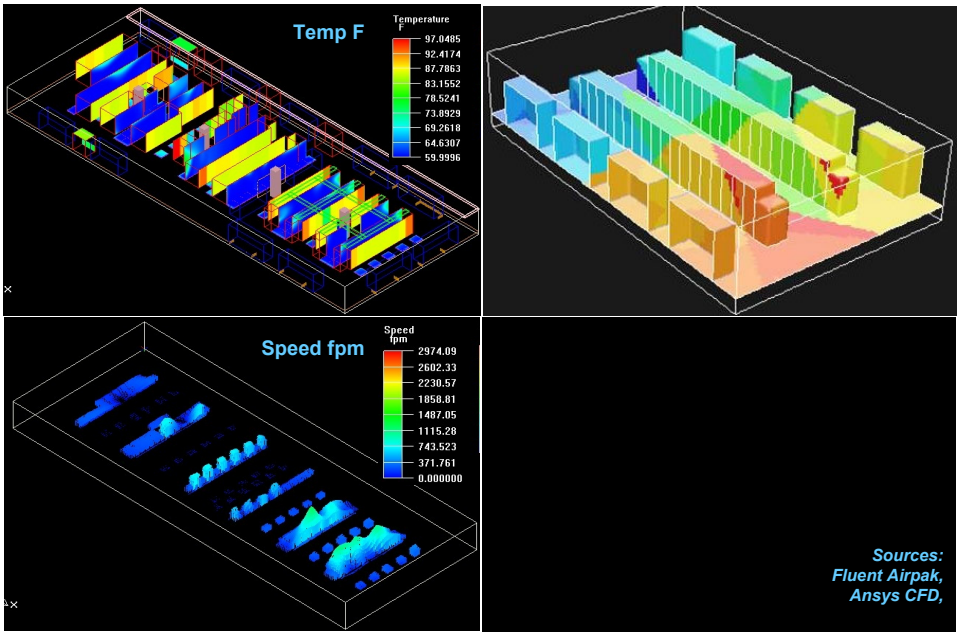
HPC Tools Availability

- *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/PESSL, 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

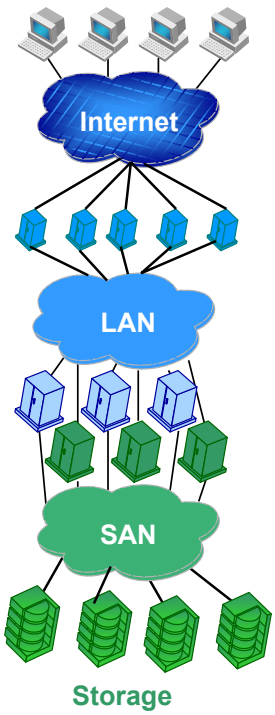
3 Tier Computing Infrastructure

Application
End-Users

Tier –1
Network Services:
- Web Servers,
- Firewalls,
- Load balancers

Tier-2
Application Services:
- Application Servers

Tier-3
Data Base Services
- Database Servers
- Storage

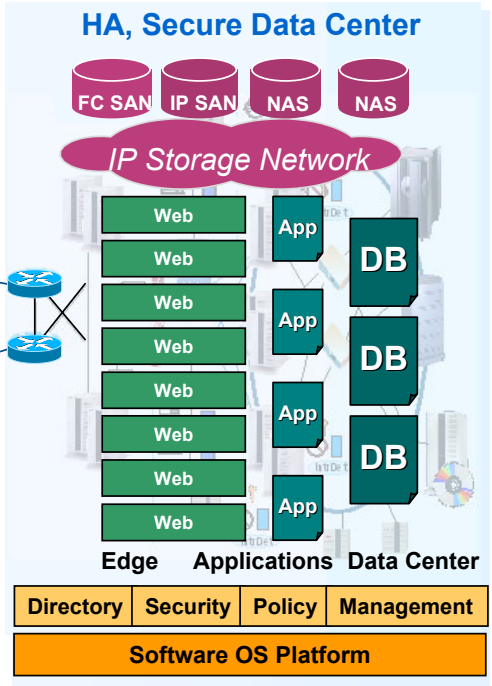


- Uniform and ubiquitous physical Connectivity – "Wire Once"
  - **Any user to any server via Internet**
  - **Any server to any server via the LAN**
  - **Any sever to any storage via the SAN**
- Dynamic Logical Binding of…
  - **server network identification**
  - **server OS version**
  - **server application assignment**
  - **application data volumes**
- Resulting in
  - **Application Mobility across servers**
  - **Data mobility across storage**

©2000-2005 IMEX Research

IMEX

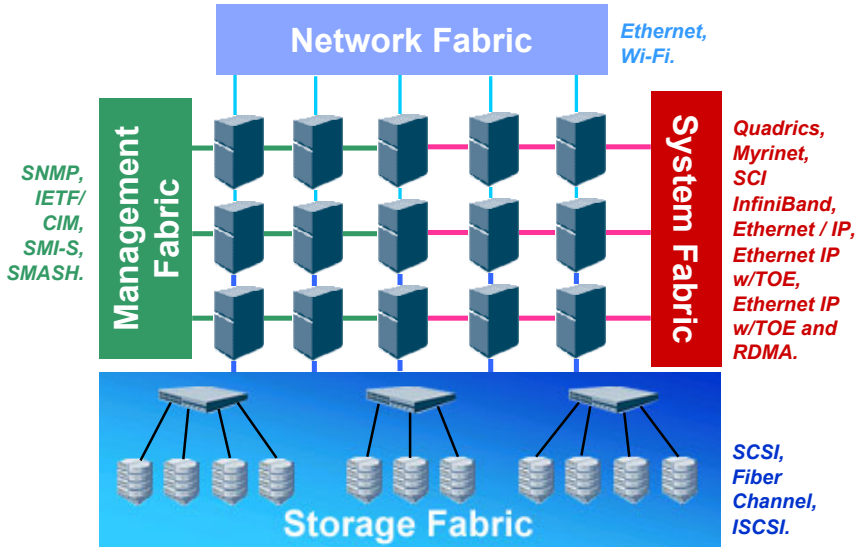
E2E Internet & 3 Tier Computing



©2000-2005 IMEX Research

IMEX

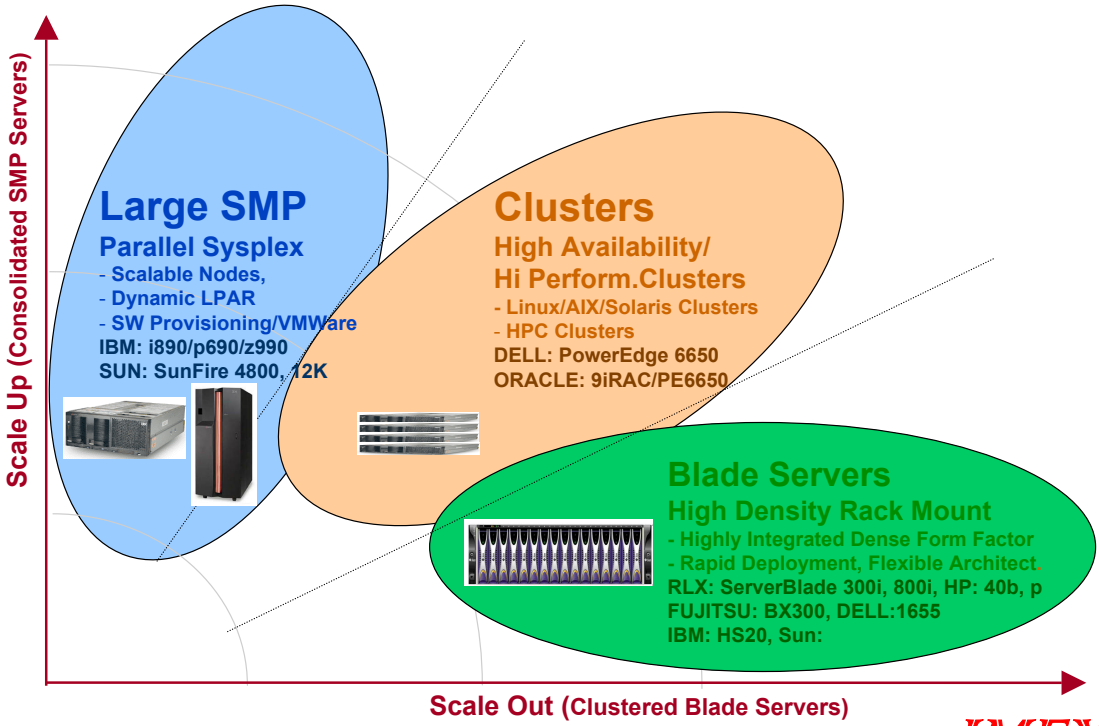
Fabric based Clusters Architecture



©2000-2005 IMEX Research

IMEX

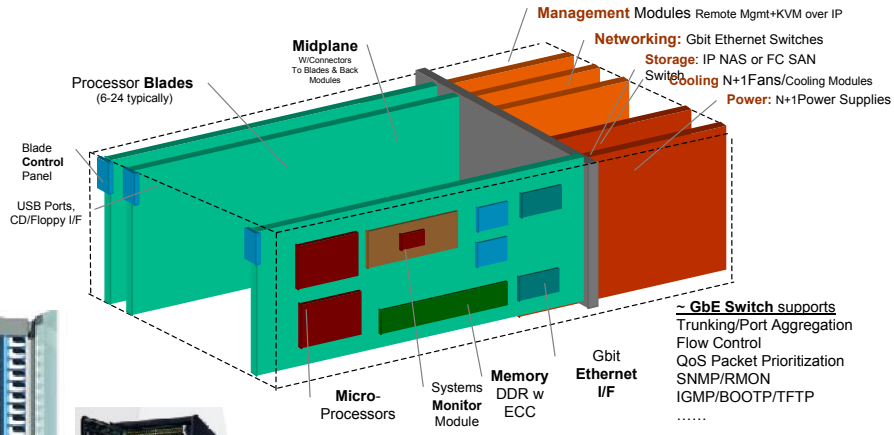
Server Platforms - Migrations



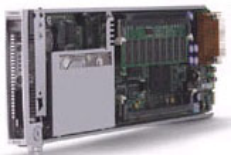
©2000-2005 IMEX Research

IMEX

Blades Servers - Infrastructure

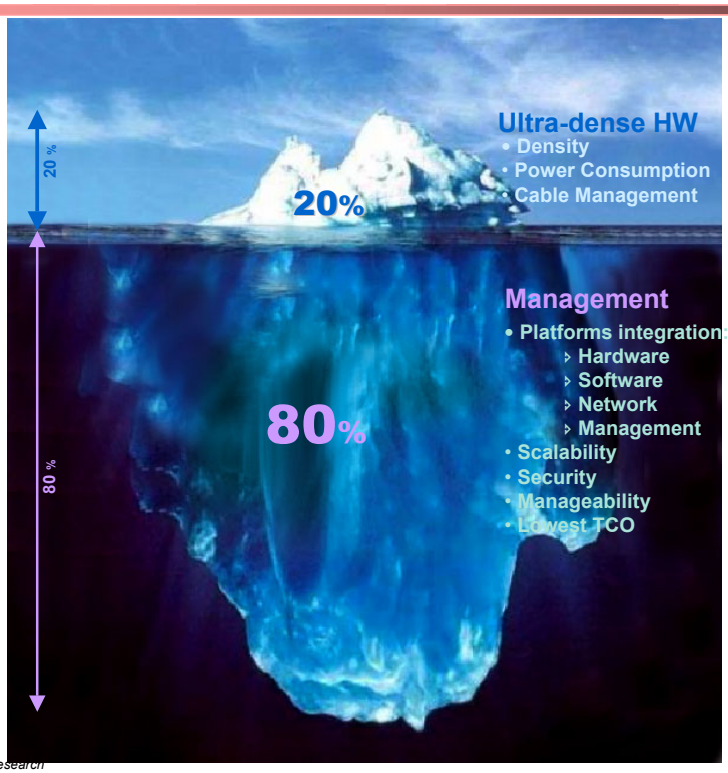


© 2003 Source: IMEX Research



IMEX

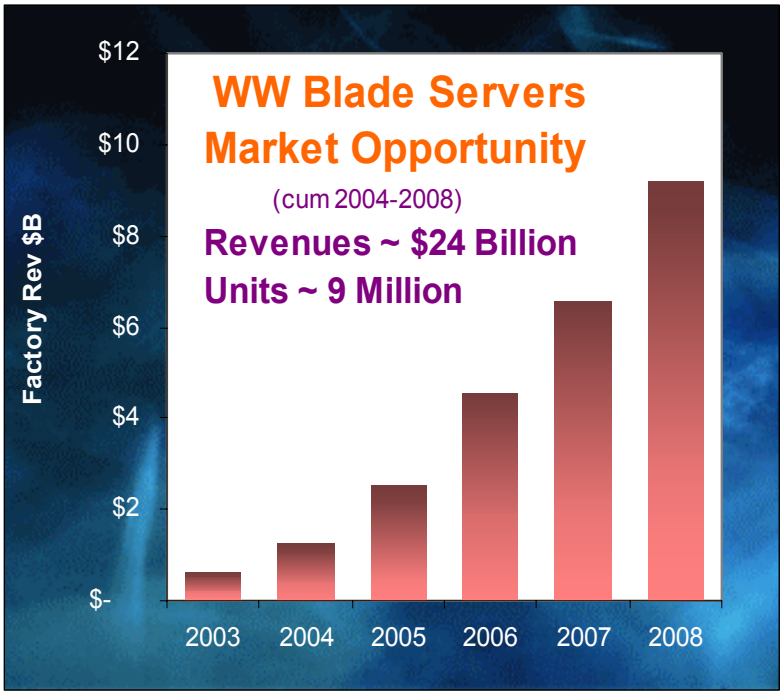
Management – Blade Servers



©2000-2005 IMEX Research

IMEX

Blade Servers Market Opportunity

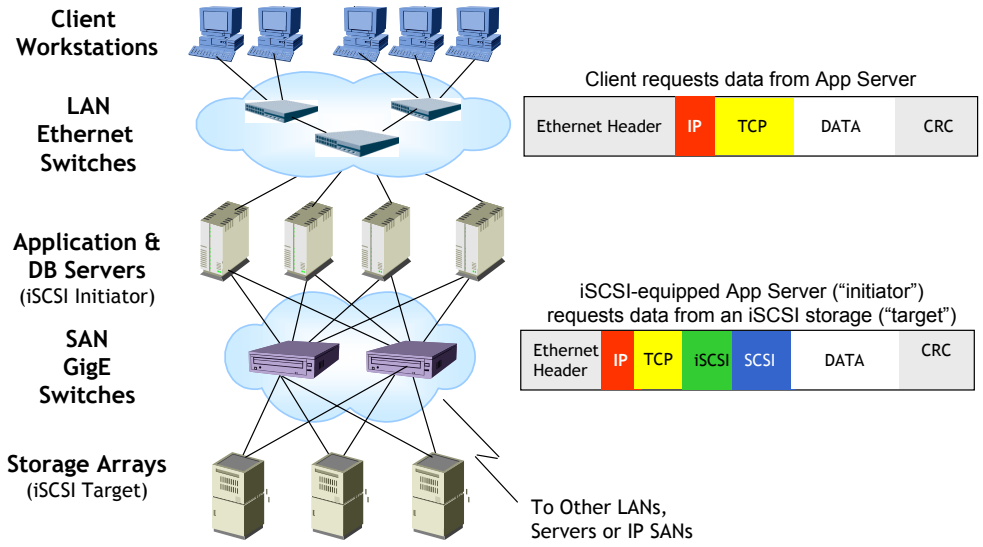


Data: IMEX Research 2005

©2000-2005 IMEX Research

IMEX

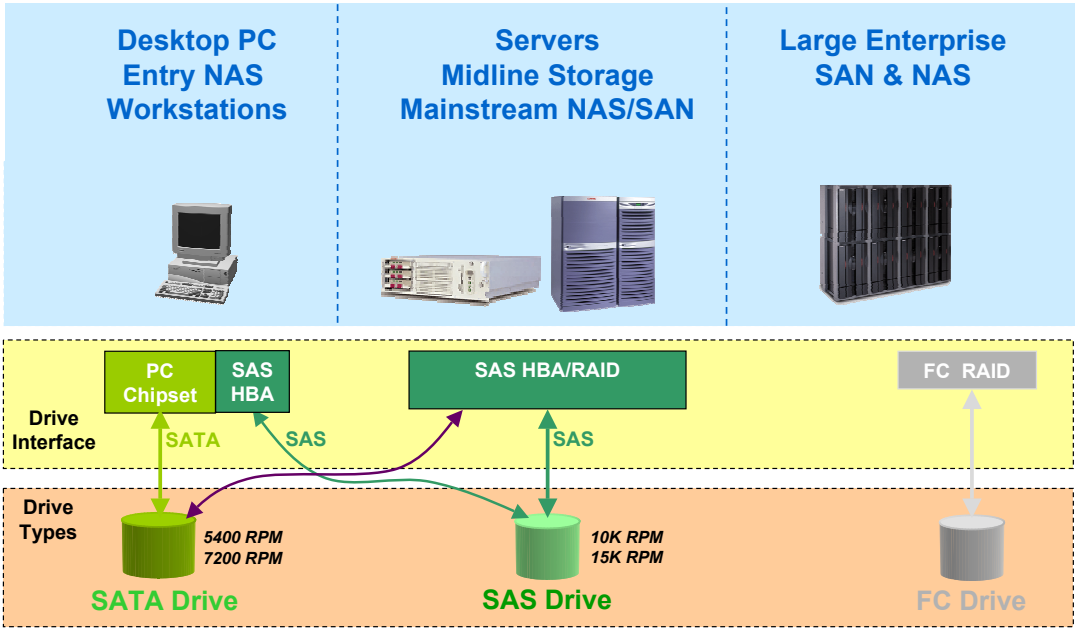
The iSCSI SAN



©2000-2005 IMEX Research

IMEX

Tiered Storage by Price/Performance



Source:Maxtor

©2000-2005 IMEX Research

IMEX

Grid Computing

- **Grid is a Catch-all Marketing Term**
  - Means different things to different constituencies
  - Desktop Cycle-Stealing
  - Manahed HPC Clusters
  - Virtualization of Data Center Resources
  - Outsourcing to “Utility Data Centers”
- **Multiple, opposing requirements**
  - Compute-intensive applications want to ship data tp 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 woek well across internet
  - Internet is not the CPU back plane

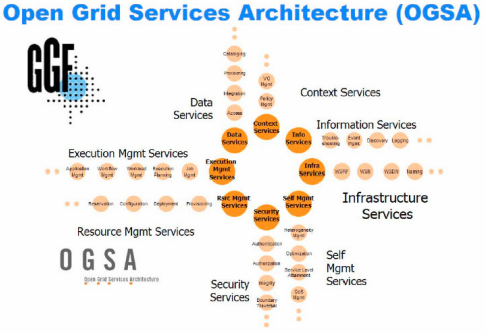
* Copyright 2005 IMEX Research & IMEX International

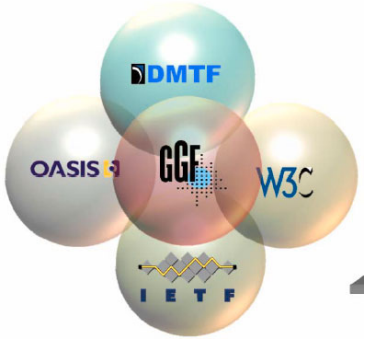
©2000-2005 IMEX Research

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* — alphaWorks
  - 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* — Apache
  - Implementation of WS-RF / Notfication in next full release of Websphere — *Integrated with Websphere programming models & development tools* — Websphere
  - Cooperation between Web Services Standards bodies — *MOU on cooperation around Web Service Management* — OASIS, DMTF, W3C, GGF, 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**

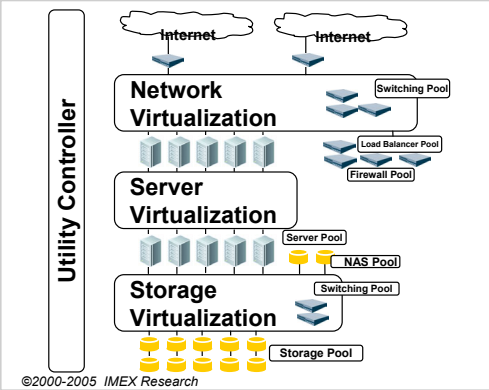
©2000-2005 IMEX Research

IMEX

Utility Computing/Web Services

Autonomics

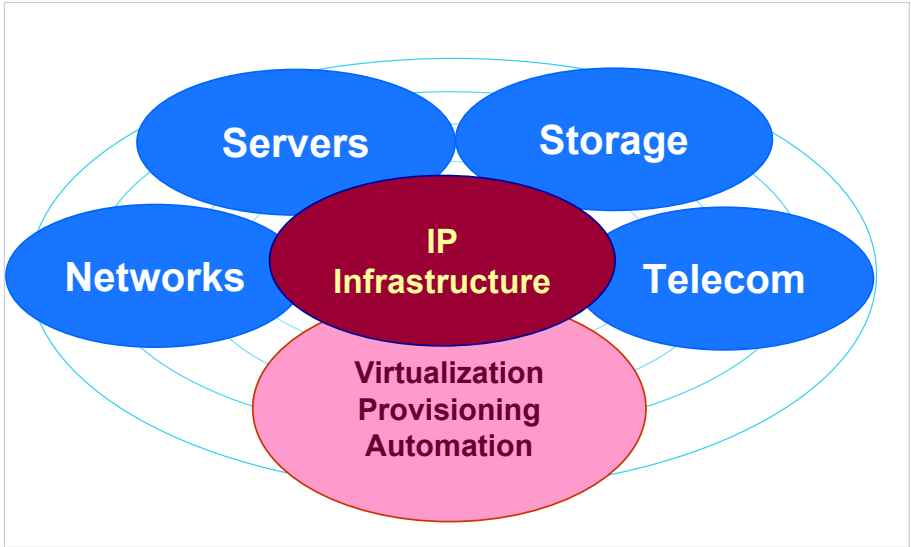
- Self Healing
- Self Optimizing
- Universal Identity
- Single System Image
- ....



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

IMEX

Future - IP Based Infrastructure



©2000-2005 IMEX Research

IMEX

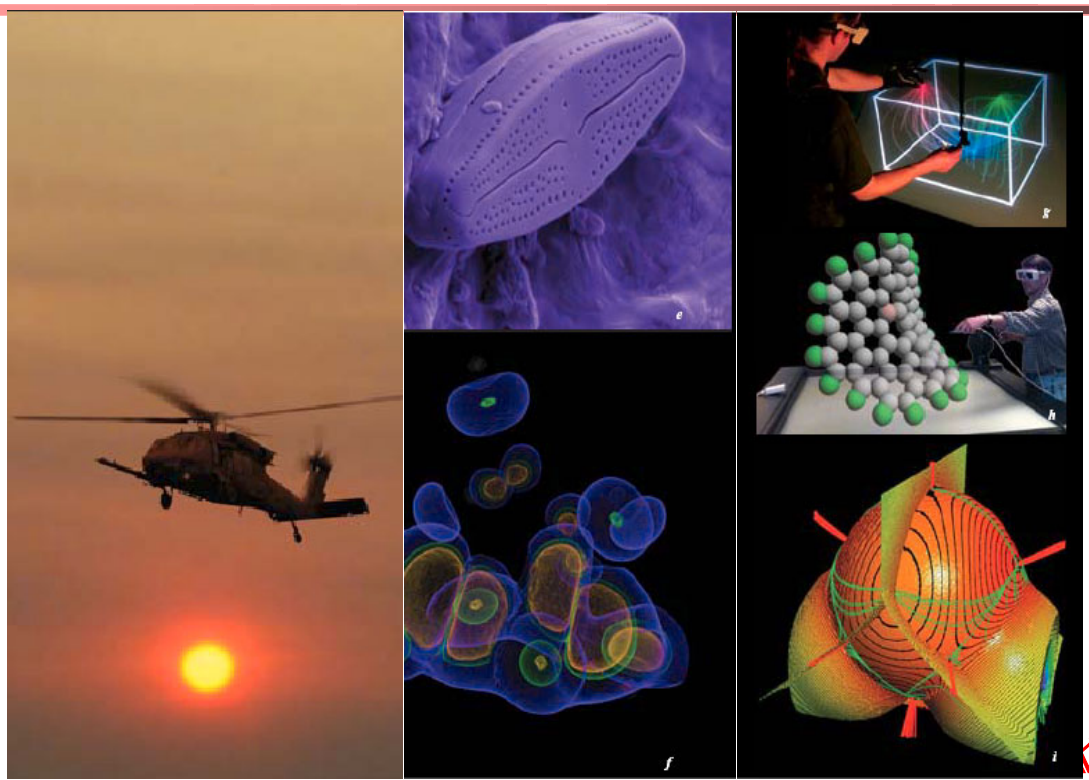
HPC – From Academia to Wall St to Hollywood

High Performance Computing	Commercial Visualization	Bioinformatics	Decision-Support Systems	Entertainment Audio/Video OnDemand
100+ Teraflops	Rendering (Texture & Polygons)		Data rate & capacity	
Throughput = 100 GB/s	Throughput = 1.2 GB/s		Throughput : DSL/Cable	

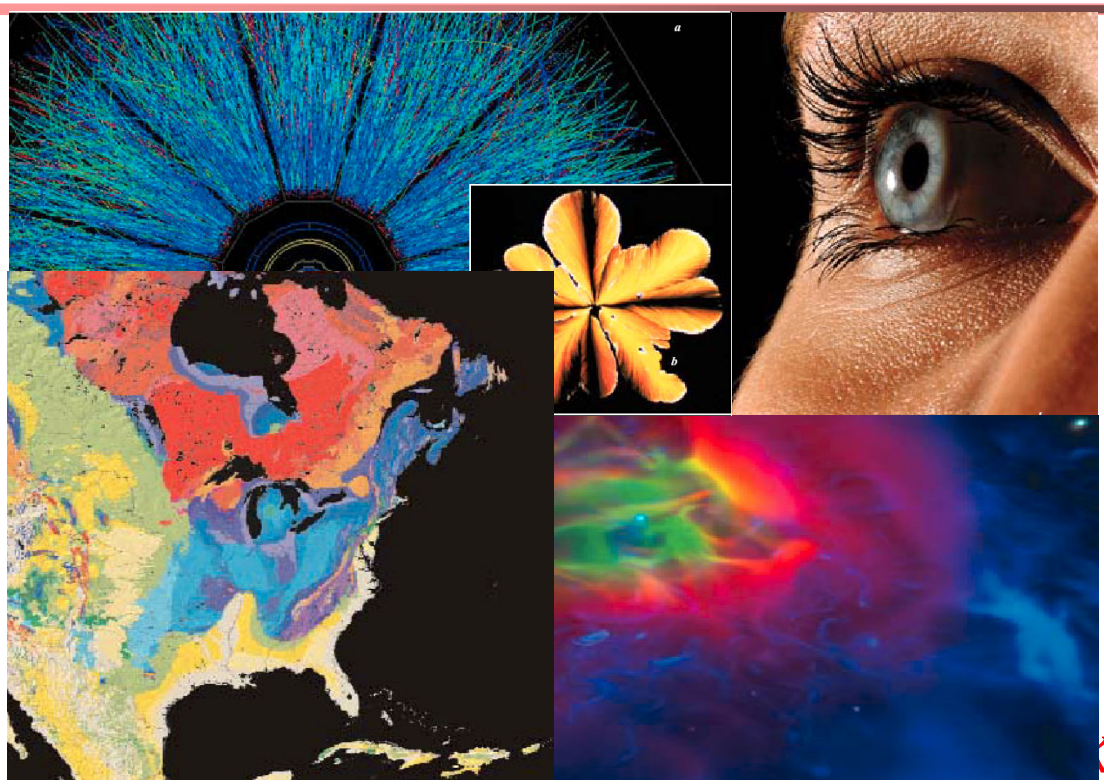
Data: IMEX Research & Panasas

©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
408-268-0800

IMEX
RESEARCH.COM

©2000-2005 IMEX Research