Scali Manage for LC Series

Session 3728

Dan Cox

Manager – ISS HPTC Programs

Hewlett-Packard

Scali Manage for LC Series – Topics

* HP HPTC Portfolio and LC Series Partner Driven Software Solutions

* Scali Manage Overview

* Q&A's

HPTC and new technologies at HP

"High performance technical computing is a strategic focus for HP. It is here that new computing paradigms are created and the applications they enable become early indicators of the general commercial applications that will follow."

Carly Fiorina

Chairman and CEO, HP

Drive technology

Technical users
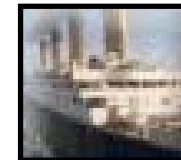
Commercial users

Drive volume
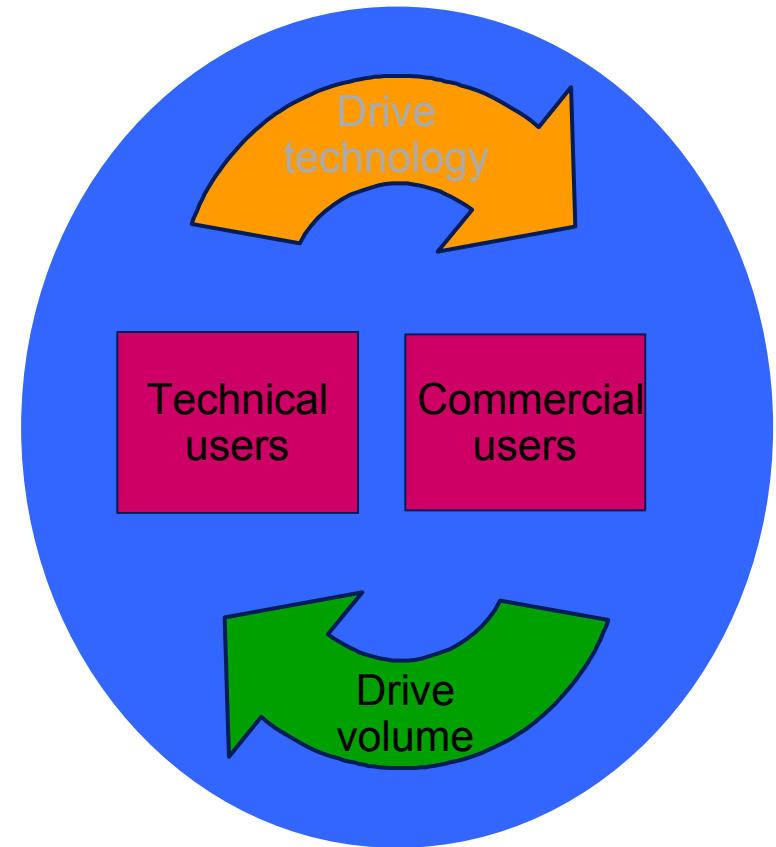
HPC @ HP

Scientific Research
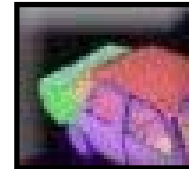
Mechanical Engineering / Virtual Prototyping
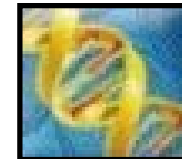
Geo Sciences

High-End Film / Video

HPC problems are characterized by computational, data-intensive, or numerically intensive tasks involving complex computations with large data sets requiring exceptionally fast throughput.
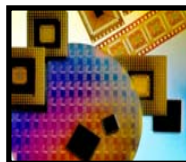
Finance / Securities

Life and Materials Sciences

Electronic Design Automation
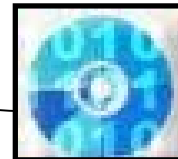
Government Classified / Defense
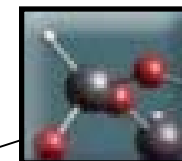
Product Lifecycle Management / Informatics

HP's HPC Platforms

XC6000
HPTC cluster

XC3000
HPTC cluster

LC Series 3000
HPTC cluster

LC Series 1000/2000
HPTC clusters

HP Integrity
Superdome

HP Integrity
rx8620

HP Integrity
rx7620

HP Integrity
rx4640

HP Integrity
rx2600

HP Integrity
rx1600

HP ProLiant
DL585

HP ProLiant
BL20p/BL30p

HP ProLiant
DL360/DL380

HP ProLiant
DL140/DL145
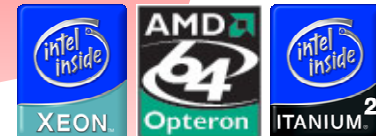
HP's HPC Software Architecture Portfolio

Scalable Cluster Services

HPTC Application and Development/Grid Portfolio

HP Scalable File Share

HP Scalable Visualization

Scalable Cluster Management

Partner software

Partner SW

XC Cluster SW

ClusterPack

Common Operating Environments

Windows

Linux

HP-UX

HPTC Cluster Platforms
ProLiant & Integrity/Multiple Interconnects
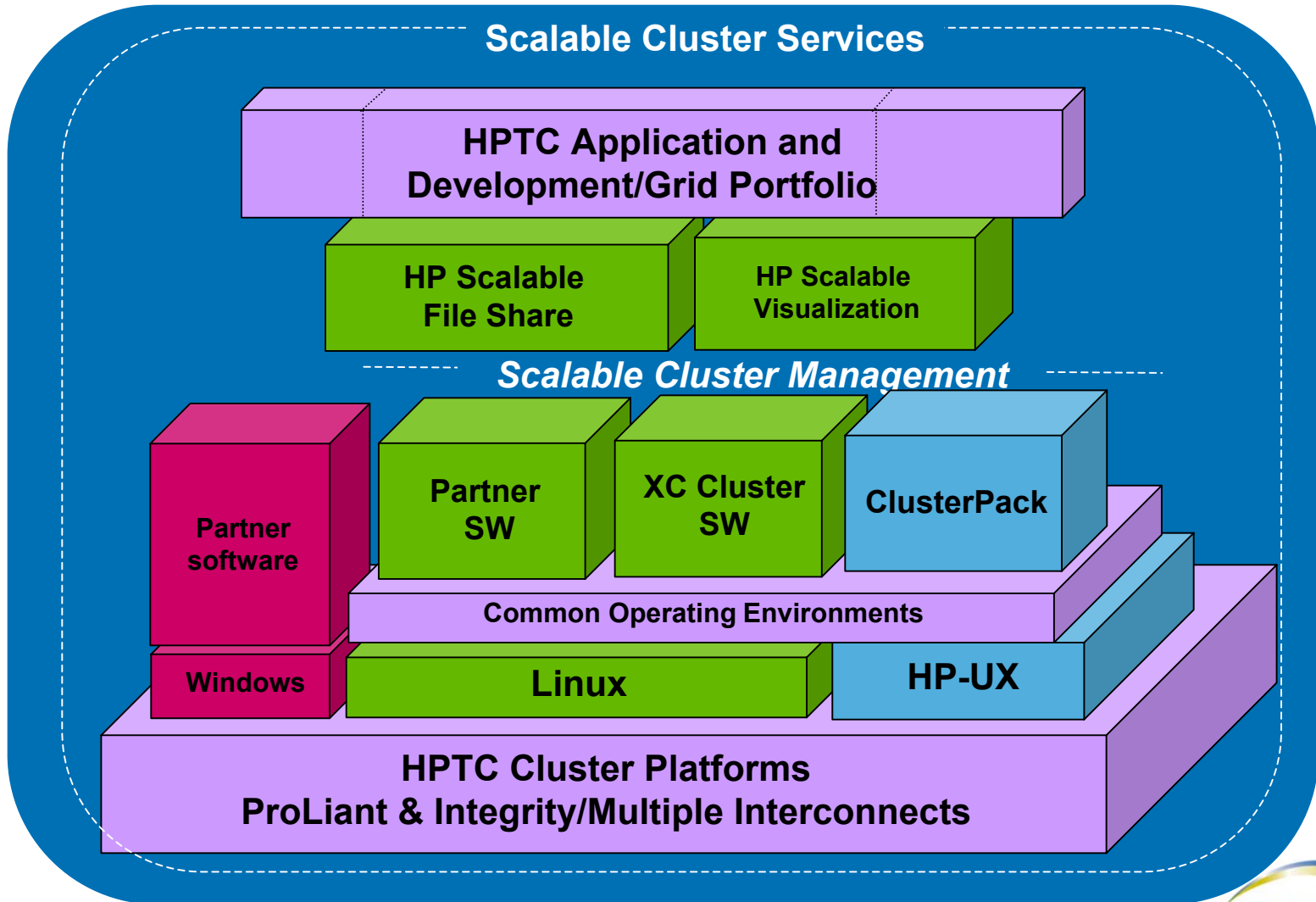
Common products available for multiple OS (OS specific implementation)

Linux

HP-UX

Windows

HP's HPC solution strategy –

- Goal:
  - Deliver complete turn-key integrated solutions that are designed specifically for customer needs

- Challenge:
  - The key is tight linkages with application partners and a process for design, build, delivery, and support
    - Application codes are complex to integrate as hardware components, maybe more so!
    - Software partners and hardware companies have not been taking ownership for this!
  - A viable production cluster solution requires a unique set of skills and products only available through a partnership with leaders in their respective fields and Professional Services trained and aware of HP solutions and alternatives

HP's HPC product delivery

Deliver unprecedented capabilities through pervasive technologies and partnerships

LC Series

Value solutions
Pentium Xeon DP and AMD based

- Customer choice thru partnership
- Certified solutions /options
- Integration by HP or Certified Resellers
- Supported by HP Service

Performance solutions
Itanium 2-based

- Exploit Itanium 2 architectures
- Integrate new I/O architectures
- Advanced tools and utilities for scale
- Contribute to Open Source

Specialty solutions

- Commitment to HPTC excellence
- Be the leader in Linux innovation
- Maintain leadership in our chosen industries
- Take an aggressive stance in industries where we do not lead

Common software architectures

Solution innovation
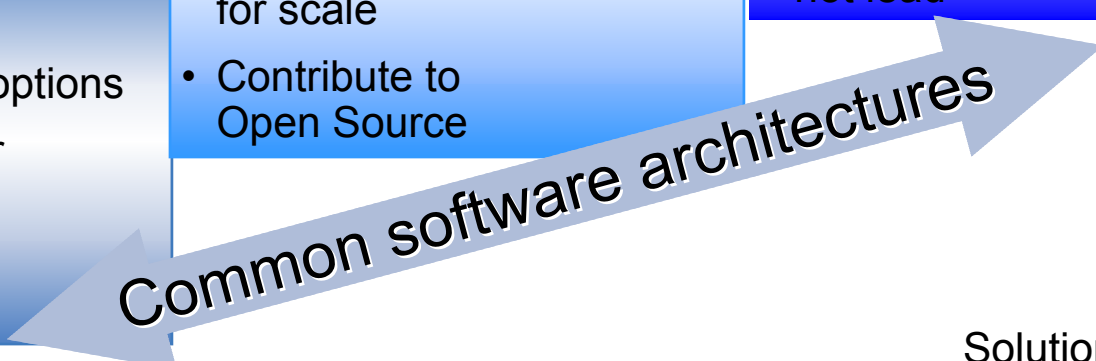
Scalable Linux	**Performance SAN**	**Performance networks / switches**	**Multi-system provisioning**

HP's High Performance Cluster Choices

Components	Linux			HP-UX Clusters
	LC Clusters	XC Clusters	Reference Model	
Optional C&I Services	Optional C&I Services	Optional C&I Services	Optional C&I Services	Optional C&I Services
HP, 3rd Party or Open Source Software	3rd Party or Open Source Software	Integrated cluster & resource mngt RH-compatible, HP supported Linux OS	HP, 3rd Party or Open Source Software	ClusterPack HP-UX 11i V2
Interconnects →	**LC** HP ProLiant servers GigE & Myrinet **LC1000 – DL140** **LC2000 – DL360** **LC3000 – DL145**	**XC** HP Integrity & ProLiant servers QSnet & Myrinet **XC3000 – DL360** **XC6000 – RX2600**	HP Integrity & ProLiant servers QSnet & Myrinet HP Integration	HP Integrity servers GigE, HF2 interconnect Infiniband (Q1/04) HP Integration
Networks →				
Storage →				
Nodes →				

HP Supported Product	**3rd Party Supported**	**C&I Supported**

9

LC Series Overview

- LC for 'loosely coupled'
- Three new orderable 32-node SKU solutions
  - **LC1000 Series** — ProLiant DL380 Control node and 32 ProLiant DL140 Compute nodes
  - **LC2000 Series** — ProLiant DL380 Control node and 32 ProLiant DL360 Compute nodes
  - **LC3000 Series** — ProLiant DL380 Control node and 32 ProLiant DL145 Compute nodes
- Common Packaging for ease of ordering, manufacturing and support
  - Out of Band unique due each platform device, IMPI 1.0, 1.5, and iLO

– Optional software from HP and ISV partners
- Linux and Windows 2003 HPC
- Checkpoint Restart
- HPTC Cluster and GRID Mgmt
- Development ToolKits
- Job Management

– Optional Storage Bundles for NFS and GFS Storage support
- Red hat Sistina
- PolyServe Matrix Server

– Factory integrated and tested in Houston and Erskine

– Complete suite of service and services offerings

HP LC Series Family Overview
LC Series Design and Configuration Guide

Reference Models for flexibility:

- ✓ Any server model (2.8, 3.06, 3.2…)
- ✓ Any compute node count up to 128
- ✓ More than one control node
- ✓ Storage options
- ✓ Change switch types/size
- ✓ Modified cabling
- ✓ Customized software installs
- ✓ Component location changes

Reference Models require:

- ✓ Purchase of the Configuration Resource Kit
- ✓ Factory presets of servers and switches
- ✓ Cabling consistent with the LC cluster specification
- ✓ Use of components certified for use in LC clusters
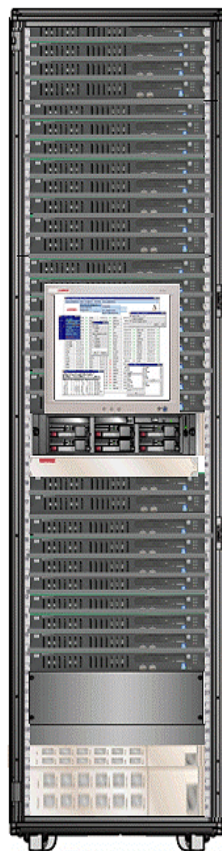
Reference Model Benefits:

- ❖ Speed flexibility
- ❖ 8 to128 node reference design
- ❖ Supports HA Control Node
- ❖ NFS to 48TB GFS options
- ❖ Flexibility and choice
- ❖ Meet specific site need
- ❖ Add infrastructure products
- ❖ Meet specific site need

Reference Model benefits:

- ❖ Fully documented configuration for ease of growth or service
- ❖ Geared to application performance
- ❖ Ease of growth for scale out and serviceability
- ❖ Ease of serviceability and scale out

Over 180 Tested Configuration Options

LC Series Partner Driven Software Portfolio

Applications	Bioscience, Engineering, GeoScience, etc
GRID Mgmt	Axceleon , United Devices
HPTC	Scali
Cluster Reliability	Meiosys– Checkpoint Restart
Job Management	Altair
Development	Engineered Intelligence
Data Mgmt	PolyServe SNFS, Red Hat Sistina GFS
Operating System	Red Hat / SUSE Linux, Windows

Certified Partner Based Solutions - Linux

- Multiple Linux cluster management choices allow customers to find the 'best fit' based on
  - experience with Linux and open source
  - on-site capabilities
  - ISV application focus
  - ISV support for target hardware
- Service programs in place, or in roll-out, for these partners
- Available as integrated solutions from each region

Certified Partner Based Solutions - Windows

- Microsoft support is beginning thru Microsoft and limited partner approaches
- Multiple application choices can be adapted to Windows 2003 Server High Performance Computing Edition
  - Special Windows 2003 HPC pricing for Compute Nodes
- Partner solutions based on end user needs that HP is partnering with
  - Utilities
  - Bio Life and Materials Science
  - Financial Services Modeling
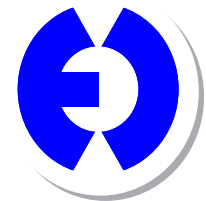
" HP is the unchallenged leader in delivering HPC solutions and provides the most competitive breadth of Linux and Windows cluster solutions offered. LC Series clusters are available to meet just about any customers' needs. They are configured, tested, and shipped ready to run out of the box. No special ordering process is required, regardless of the node architecture."
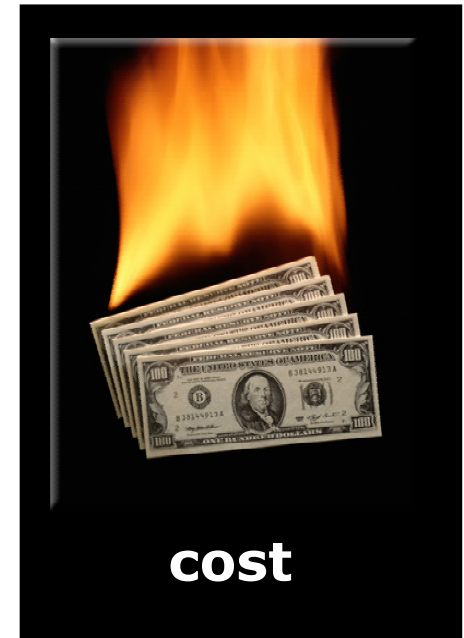
Robert de Sautel

Principal, Harvard Research Group

Harvard Research Group, Inc.
P.O. Box 297
Harvard, MA 01451 - 0297 USA
 Tel. (978) 263-3399

Scali Manage Overview

Business Challenges of Clustering
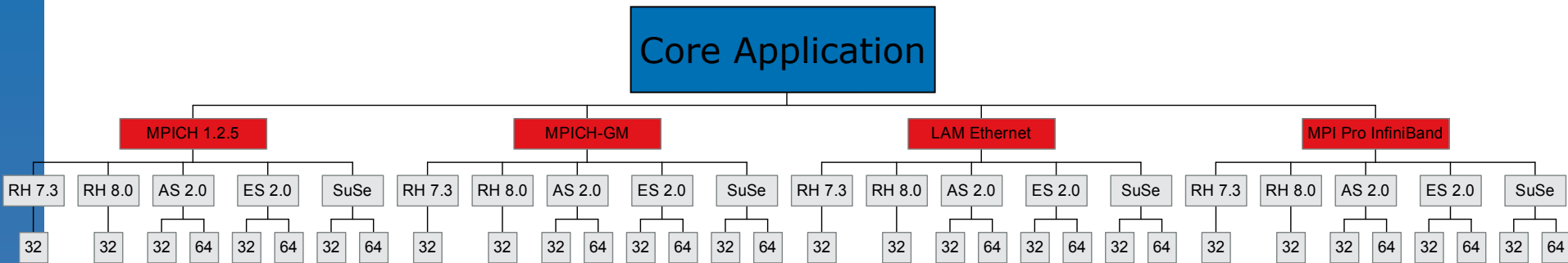
complexity

management

cost

Complexity of Clustering

Growing Number of Application Version

- 1 Application
- 4 MPI Versions
- 3 Interconnects
- 5 OS Versions
- 2 Architectures

32 Application Versions


```
                           ┌──────────────────┐
                           │ Core Application │
                           └──────────────────┘

    MPICH 1.2.5            MPICH-GM              LAM Ethernet          MPI Pro InfiniBand

RH 7.3  RH 8.0  AS 2.0  ES 2.0  SuSe    RH 7.3  RH 8.0  AS 2.0  ES 2.0  SuSe    RH 7.3  RH 8.0  AS 2.0  ES 2.0  SuSe    RH 7.3  RH 8.0  AS 2.0  ES 2.0  SuSe

 32    32    32  64  32  64  32  64  32    32    32  64  32  64  32  64  32    32    32  64  32  64  32  64  32    32    32  64  32  64  32  64
```

ISSUES: High cost => Complex => Inflexible

Complexity of Clustering Software

Application		Interconnect		MPI		Compilers & Libraries		Management Applications

| Application	Interconnect	MPI	Compilers & Libraries	Management Applications
		MPICH 1.2.4 Ethernet		Score
	Myrinet	MPICH 1.2.5 GM		
			Absoft	AlinkA
	Fast Ethernet	MPICH 1.2.5		
	Gigabit Ethernet		Portland	Oscar
		MPI Pro-Ethernet		
	SCI		GNU	CSM
		Proprietary OEM MPI		
	Infiniband		Intel	XCAT
		MPI Pro-Infiniband		
	Quadrics		NAG	
		MPI Pro-Myrinet		Platform
		MPICH for SCI		CMU
		LAM		

Dept Clustering Objectives
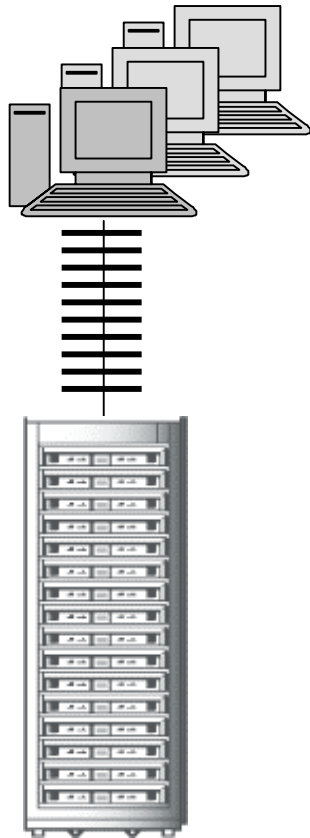
They Want:

- ✕ Focus on ISV application and results
- ✕ Optimized price performance at dept level
- ✕ Localized control of resources
- ✕ Rapid time to productivity
- ✕ Operational system without hassle
- ✕ Easy to use functionality
- ✕ Professional support & accountability

They Don't Want:

- ✔ System admin responsibility
- ✔ Loss of productivity for key personnel
- ✔ Cost and burden of operations
- ✔ Logistic hassle of node replacement, warranties etc.
- ✔ Not driven by corporate standards

Corporate Clustering Objectives
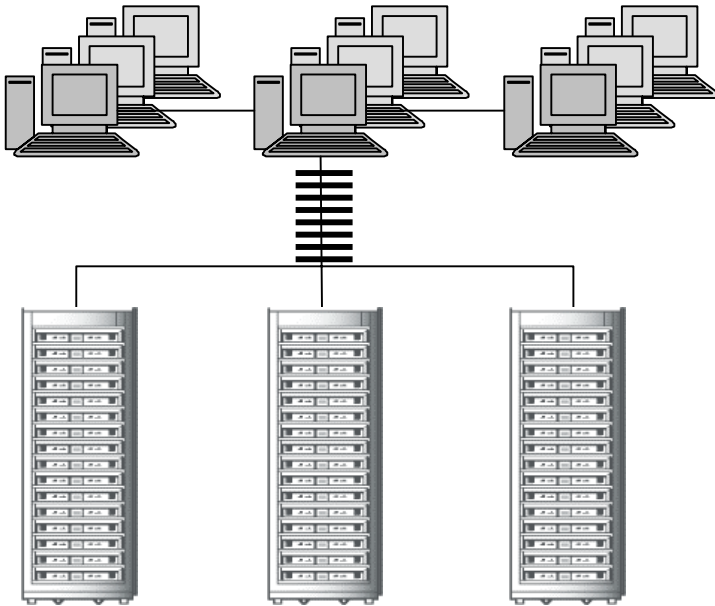
Increased efficiency and reduced cost of ownership

They Want:

- ✕ Standardized MPI and management
- ✕ Centralized IT and security
- ✕ Reduced cost of operations and infrastructure (HVAC)
- ✕ Efficient system administration
- ✕ High cluster availability and utilization
- ✕ Flexibility to run heterogeneous clusters
- ✕ Localized departmental efficiency

They Don't Want:

- ✓ Local system administration
- ✓ Proprietary solutions
- ✓ Ad hoc implementations
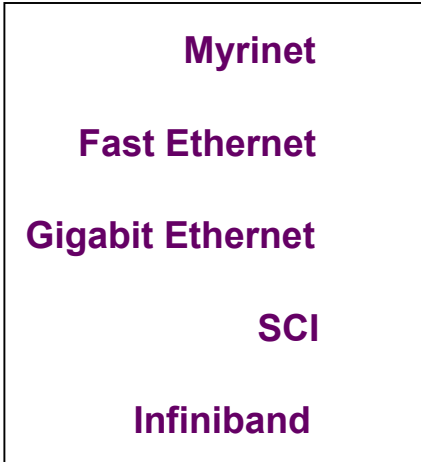- ✓ Reduced risk

Scali's Total Solution

Application	Interconnect	MPI	Compilers & Libraries	Management Applications

Myrinet

Fast Ethernet

Gigabit Ethernet

SCI

Infiniband

Scali MPI Connect ⟩ **Scali Manage**

Software Alternatives

Option 1: Open Source

- Free
- Many and growing number of "flavors"
- Platform and interconnect specific
- High cost of upstart and management
- Undefined product lifecycle
- Lack of support

Option 2: Proprietary

- Free or low cost
- Supported
- Locked to platforms
- Interconnect specific
- More efficient cost of upstart and mgmt
- Step-child product
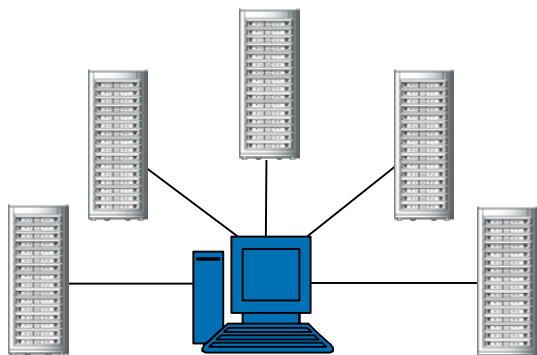- Not core competency

Option 3: Scali

- Professional software
- Single point of support
- Platform and interconnect independent
- Rapid time to start
- Lower cost of upstart and mgmt
- Core competency
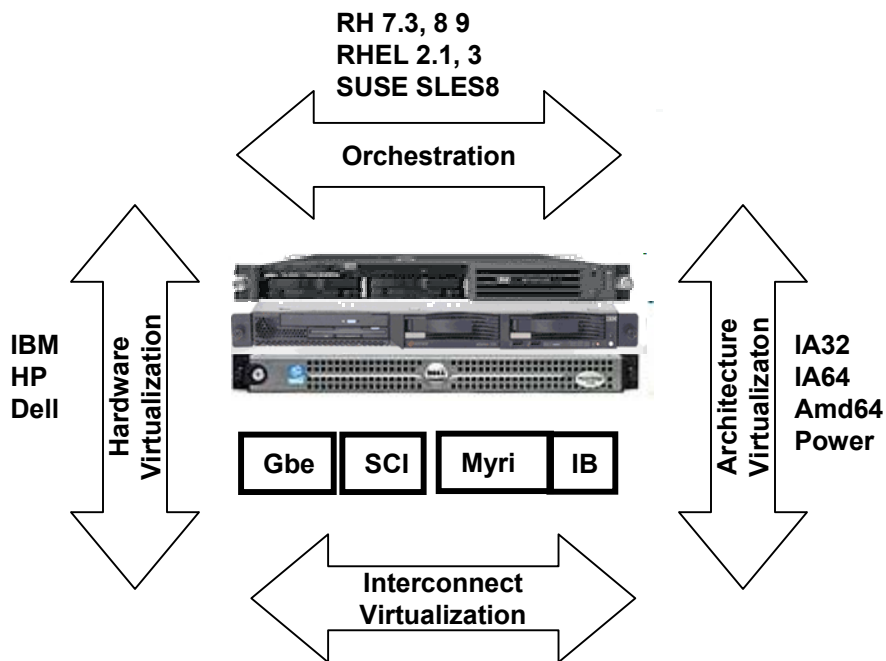- Detailed product roadmap and strategy

Scali Manage / Connect
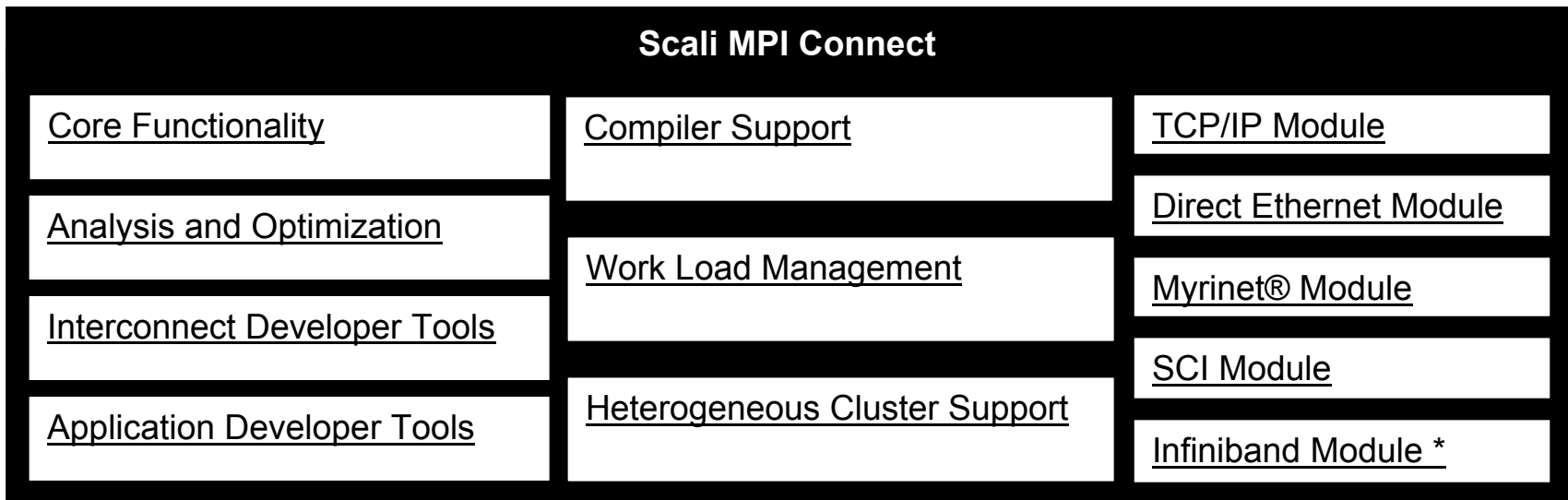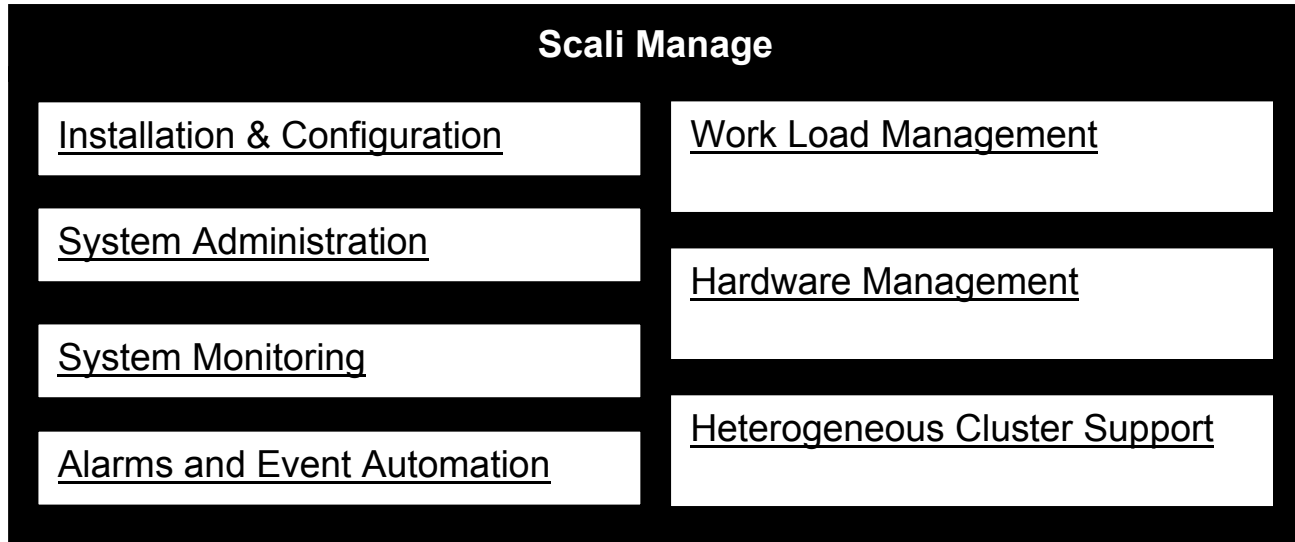
Ease of Installation and Management
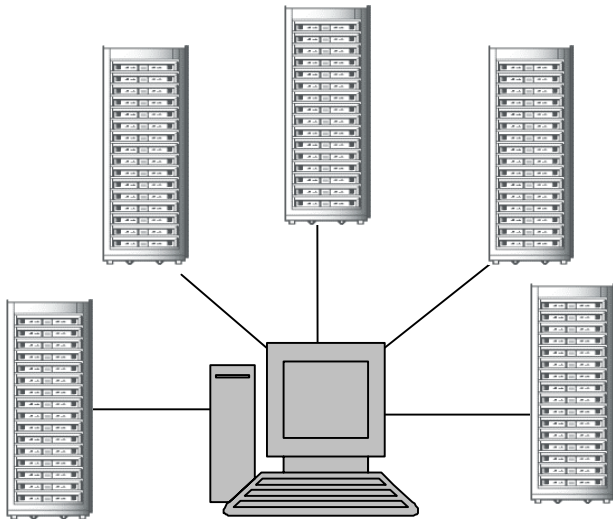
Virtualization of Hardware and OS Layer

Heterogeneous Cluster Support

Alarms & Event Automation

System Monitoring

System Administration

OS & SW Deployment & Configuration

Hardware Management

Hardware Configuration

RH 7.3, 8 9
RHEL 2.1, 3
SUSE SLES8

Orchestration

IBM
HP
Dell

Hardware Virtualization

Architecture Virtualizaton

IA32
IA64
Amd64
Power

| Gbe | SCI | Myri | IB |

Interconnect Virtualization

Scali Product Functionality

Scali Manage

Installation & Configuration	Work Load Management
System Administration	Hardware Management
System Monitoring	Heterogeneous Cluster Support
Alarms and Event Automation	

Scali MPI Connect

| Core Functionality	Compiler Support	TCP/IP Module
Analysis and Optimization		Direct Ethernet Module
Interconnect Developer Tools	Work Load Management	Myrinet® Module
Application Developer Tools	Heterogeneous Cluster Support	SCI Module
		Infiniband Module *

# Scali Manage Highlights


Scali Manage

- Support for leading platforms

- Rapid and hassle-free installation

- Advanced system administration functionality

- Robust monitoring and alarms

- Out-of-band access and administration

- Integrated support for leading Embedded Management Processors

- Integration hooks for leading 3rd party applications
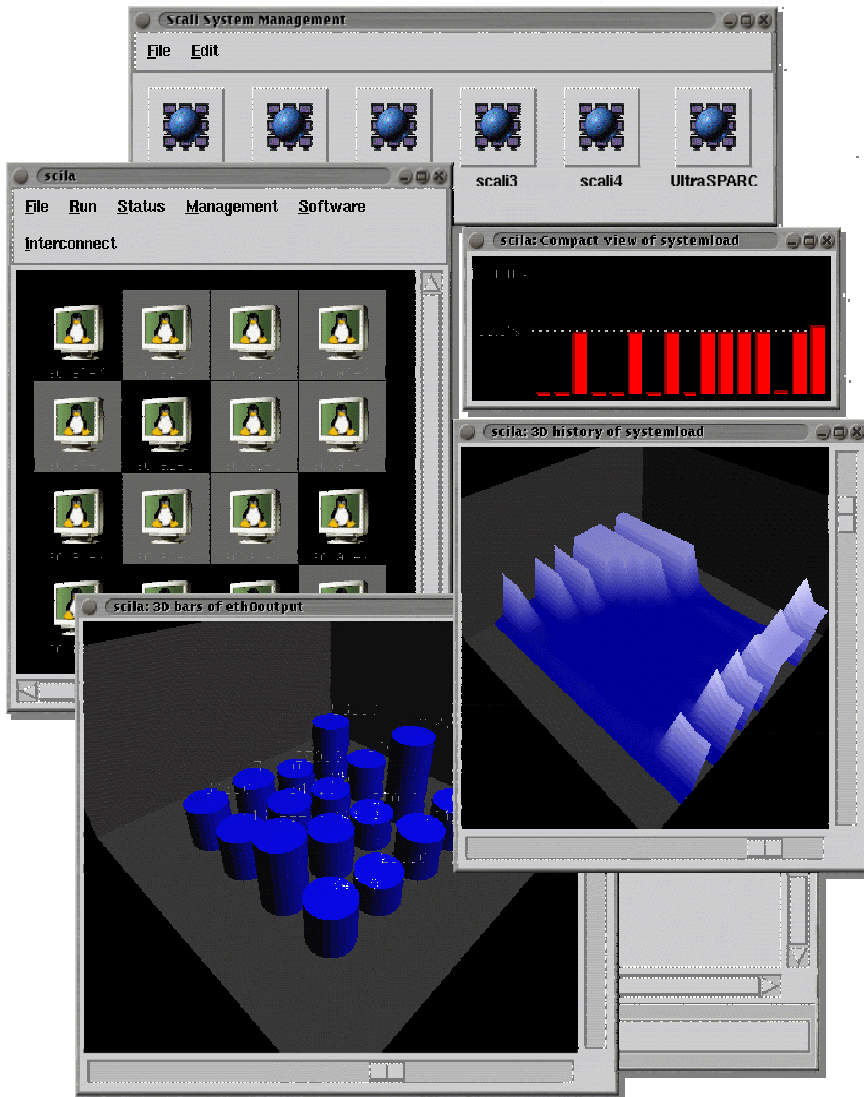
Scali Manage

Multiple Cluster Management
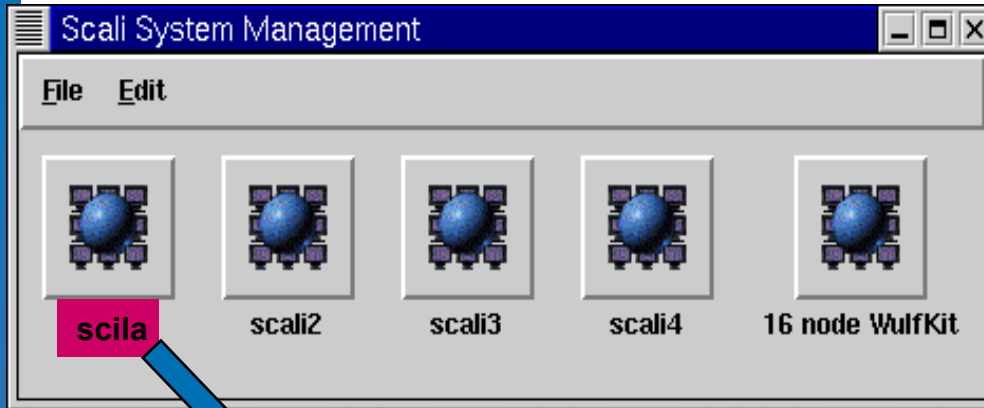
Common Login per Cluster
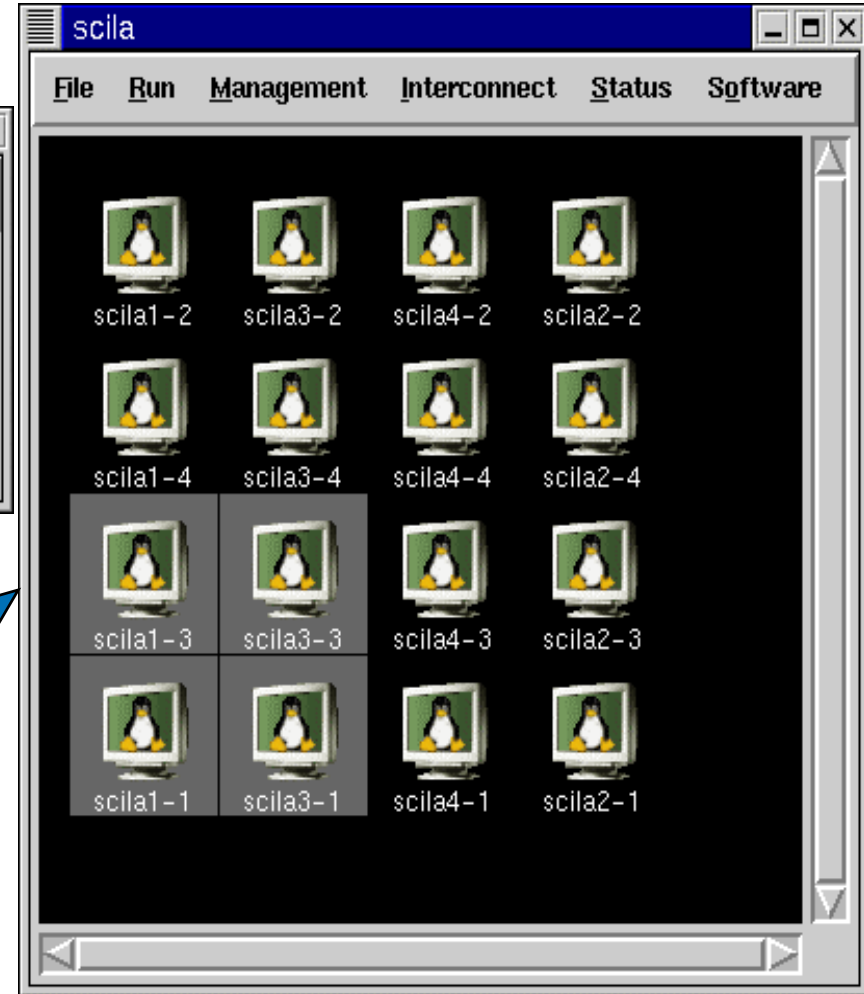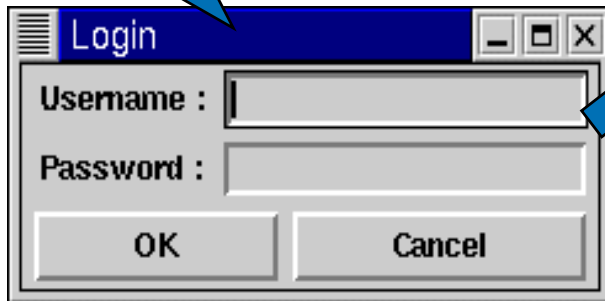
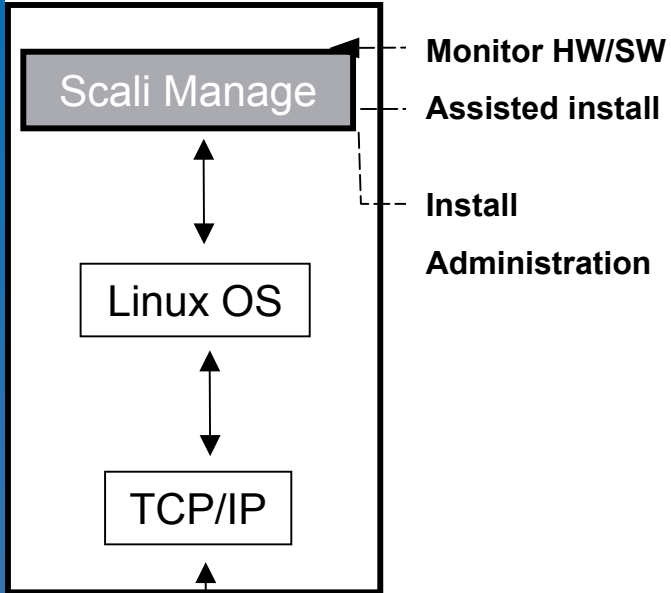**Individual or Grouped Node Management**

System Monitoring

- **Resource Monitoring**
  - CPU
  - Memory
  - Disk
- **Hardware Monitoring**
  - Temperature
  - Fan Speed
  - Voltage
- **Operator Alarms on selected Parameters at Specified Tresholds**

Product Infrastructure

Front-end

Nodes

Scali Manage

Monitor HW/SW

Assisted install

Install

Administration

Parallel Applications

Linux OS

Scali MPI Connect

Linux OS

TCP/IP

TCP/IP | DET | Myrinet | SCI | Infiniband

Ethernet switch

Myrinet switch

Infiniband switch
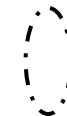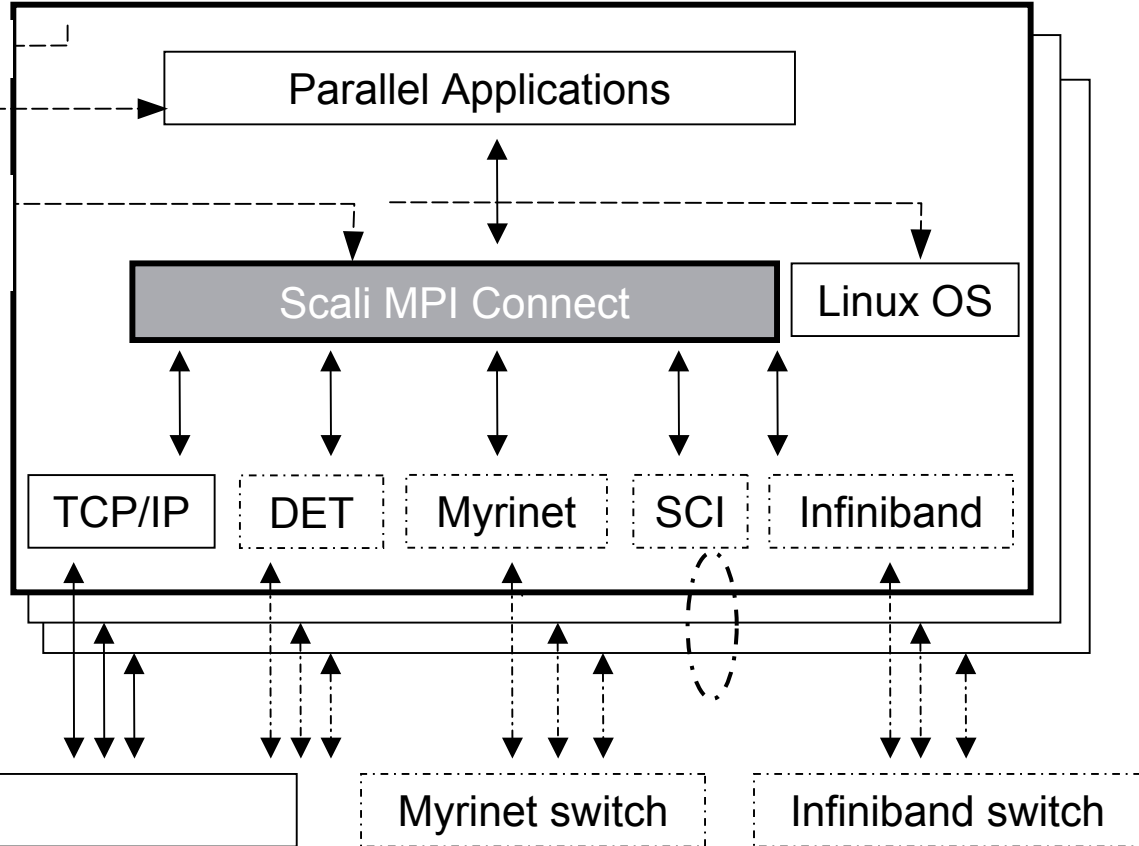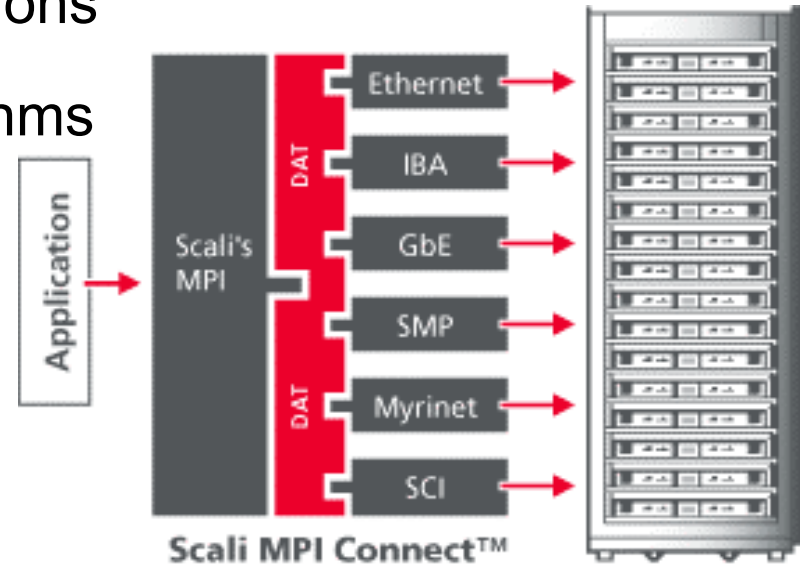
= switch independent

= optional

Scali Connect Highlights

- Interconnect independent architecture

- Dynamic interconnect selection at runtime

- Interconnect failover functionality

- Support for heterogeneous configurations

- Runtime selection of collective algorithms

- Multi thread safe implementation

Scali MPI Connect™

Scali Connect Independence

- Single MPI with integrated support for leading interconnects
  - Multiple interconnects supported in the same library

- Selection of interconnects at run-time without relinking/compilation of ISV applications

- Interconnects selection is prioritized

- Nodes do not have to use the same interconnect

- Additional interconnects (such as Quadrics) can be added

Interconnects	Drivers
▪ Gigabit Ethernet	▪ Direct Address Transfer (DAT)
▪ InfiniBand	▪ Direct Ethernet Transport (DET)
▪ Myrinet	- MAC-devices
▪ SCI	- channel bonding
	- coexistence with legacy protocols

Scali Connect Architecture

Scali Manage™ and Scali MPI Connect ™ enable companies to standardize on a single cross-enterprise clustering solution
=> reducing complexity and TCO
=> increasing value

Runtime selection of interconnect w/ failover

Dynamic link

Application

Scali's MPI

Ethernet

IBA

GbE

SMP

Myrinet

SCI

DAT Interface
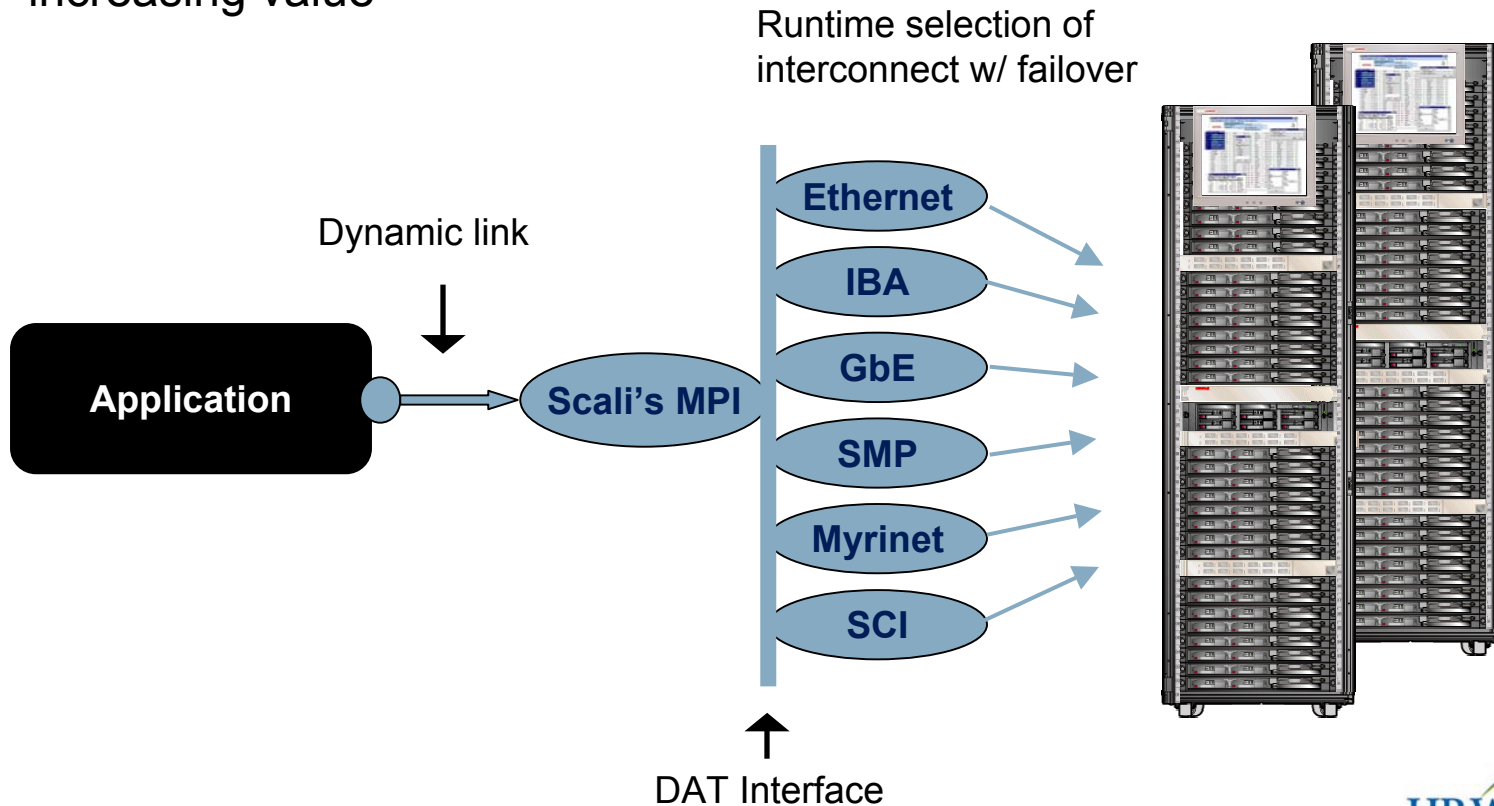
HP and Scali –University of Houston

Advanced Computing Research Laboratory

Challenge: Required system with high-bandwidth, low latency performance with increased ROI

Purpose:
- design software for computation and data grids
- research air quality and seismic modeling and genetic medication development

Cluster Configuration:

- 19 HP zx6000 workstations with dual 900 MHz Itanium 2 processors

- 1 HP rx5670 server with quad 1 GHz Itanium 2 processors

- Scali Manage and Scali MPI Connect for SCI

- 385 Mbytes per second with less than 4 microseconds latency

Record Breaking MPI Performance

385 MBytes per Second on HP

Itanium 2-Based Linux Cluster

(PCI 64-bit 66 MHZ)

"The proven performance of Scali's MPI together with their flexible cluster management software has presented us with highly favourable price performance ratios"

Dr. Lennart Johnsson,

Director ACRL, TLC2

University of Houston

HP, Scali, and IPS - Fraunhofer Bremen

- <u>Purpose</u>: Molecular modeling

- <u>Applications</u>: Turbomol, Material Studio and Gaussion

- <u>HP Hardware</u>: 31 dual 2.8 GHz HP DL360, 1 dual 2.8 GHz HP DL380 with GbE interconnects

- <u>Goals</u>:
  - 1) rapid time to production
  - 2) minimal day-to-day administration and
  - 3) single source of MPI, management and accountability

- <u>Issues</u>: Fraunhofer and IPS had unsuccessfully used 3+ months to install the cluster with an alternative mgmt system

- <u>Result</u>: With Scali Manage, cluster was live in less than 4 hours

"Based on previous experience, IPS estimates that an alternative solution would have taken closer to five days to install as opposed to the less than four hours it took using Scali Manage,"

Mr. Dziuba

CTO at IPS

Fraunhofer Bremen project in Germany

"In our evaluation process, we found that there were few software companies that could provide an integrated cluster management and MPI solution in addition to meeting our performance, professional services and support requirements,"

Dr. Schiffels

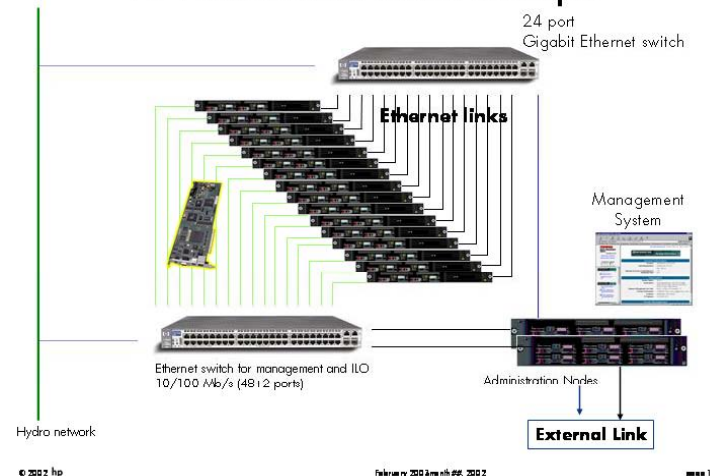Molecular Modelling Group

Fraunhofer IFAM

HP and Scali - Norsk Hydro

- <u>Purpose</u>: Geophysical processing and reservoir simulation

- <u>Applications</u>: Eclipse, Paradigm, CGG ++

- <u>Business Requirements</u>:
  - reduced TCO
  - quick time to production
  - immediate increase in capacity
  - Scalability to support
    future processing capacity
  - standardized solution from single vendor

- <u>Solution:</u>
  - 96 ProLiant DL360
  - 4 ProLiant DL380
  - Scali Manage
  - Scali MPI Connect for Direct Ethernet

HP Proliant Scali Cluster example

24 port
Gigabit Ethernet switch

Ethernet links

Management
System

Ethernet switch for management and ILO
10/100 Mb/s (48+2 ports)

Administration Nodes

Hydro network

External Link

© 2002 hp

February 2003 month ##, 2002

page 1

Please note that Scali and HP have not been given permission to reference Hydro externally

Gartner Group Recommendations

"Reducing Total Cost of Ownership of Servers"

Scali Enables

- Leverage more smaller servers pooled together (cluster/grid) -> increases server utilization

- Deploy centralized monitoring

- Automate configuration

- Leverage integrated hardware monitoring

- Maintain "generic" server purchasing to avoid vendor lock-in

- For OEM -> alliance with software vendor is key

Gartner Report dated October 2, 2003

Scali Manage - Summary
Reduces TCO and Increases Value

Scali Manage
**Scali MPI Connect**
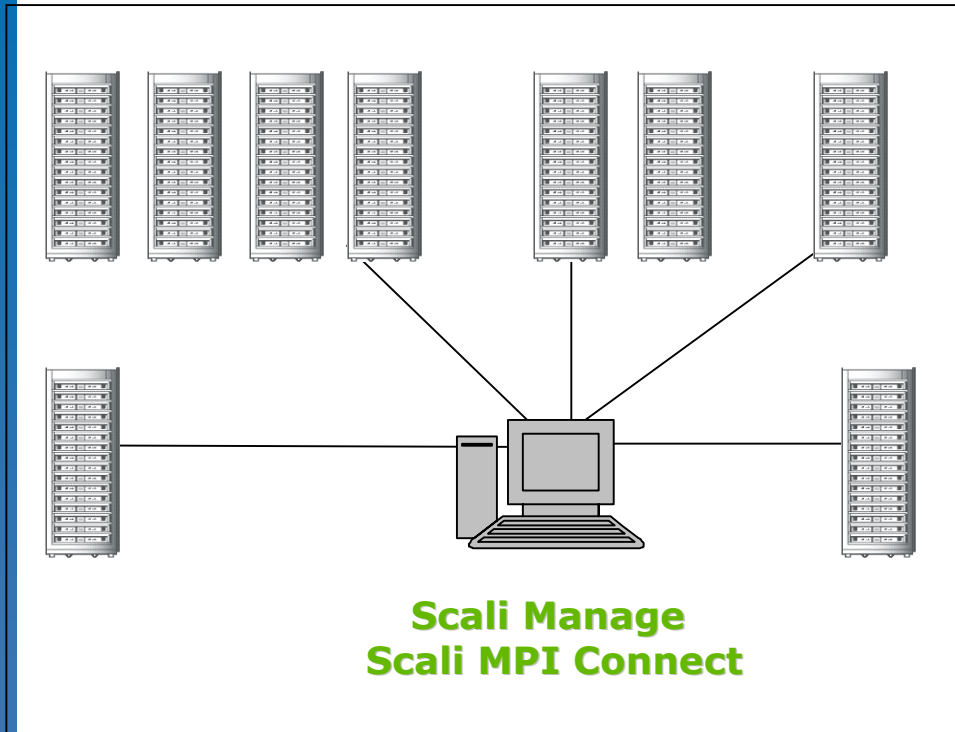
Benefits

- Standardized solution
- Single MPI and mgmt system
- Synergies across the org.
- Highly efficient administration
- Ongoing maint and support
- Higher availability and utilization
- Benefit from economies of scale

Bottom Line

- Lower TCO
- Increased value of clustering

Value Proposition

- Scali is integrated and certified for HP technology
  - HP iLO for integrated power & console management
  - Automatic integration and installation of HP InsightManager for hardware monitoring
- Partner proven and factory installed software for LC-Series and ClusterBlocks
- Joint testing and certification of HP clusters
- Formal relationship for support, services and training
- Ongoing joint development for optimized HP cluster solutions
- Proven customer references: Williams F1, Norsk Hydro, UofH, Fraunhofer, etc.
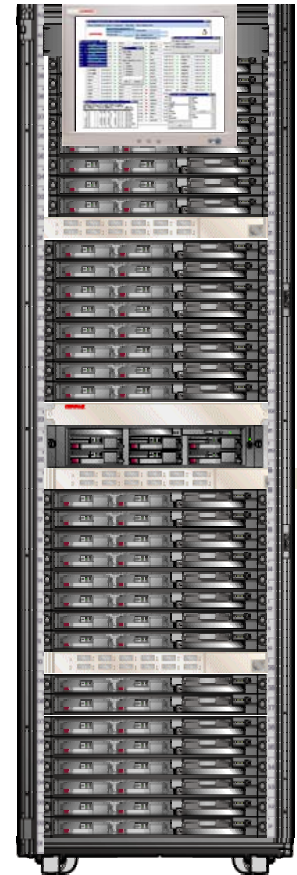
Q&A

HP delivering choice in HPC
Scale-up, Scale-out, Scale-simply

ProLiant Servers with 64-bit extensions

- Price/performance leadership with 32/64-bit co-existence
- Highest clock speed, peak performance
- Extensive 32-bit, and emerging 64-bit ecosystems
- Scale-out for simple, highly parallel workloads (2p nodes)
- Linux & Windows

- Price/performance leadership with 32/64-bit co-existence
- 32-bit throughput performance leadership
- Highest bandwidth for sustained performance
- Extensive 32-bit, and emerging 64-bit ecosystems
- Scale-out for moderate workloads (2p/4p nodes)
- Linux & Windows

Integrity Servers

- Highest performance 64-bit processor core for sustained performance
- Highest SMP scalability (to 128p)
- HP-UX for mission-critical technical computing
- Extensive 64-bit ecosystem (and 32/64-bit on HP-UX)
- Scale-up and scale-out for complex workloads
- HP-UX, Linux & Windows

Co-produced by: