Microsoft Compute Cluster Server Partner Training

Jay Ferron CISSP, MCDBA, MCSE, MCT
Windows Server® HPC

Microsoft® Windows® Compute Cluster Server 2003

HPC goes mainstream

We’re Going to Talk About …

- What defines ‘supercomputing’?
- Current market trends
- Customer challenges
- Microsoft plans for this space
- Product offering overview

Microsoft Windows Compute Cluster Server 2003

Introductions

- Name
- Company
- Role (Sales, Technical, Management)
- Experience in clusters
- One question you want answered today

Microsoft Windows Compute Cluster Server 2003

What Defines ‘Supercomputing’

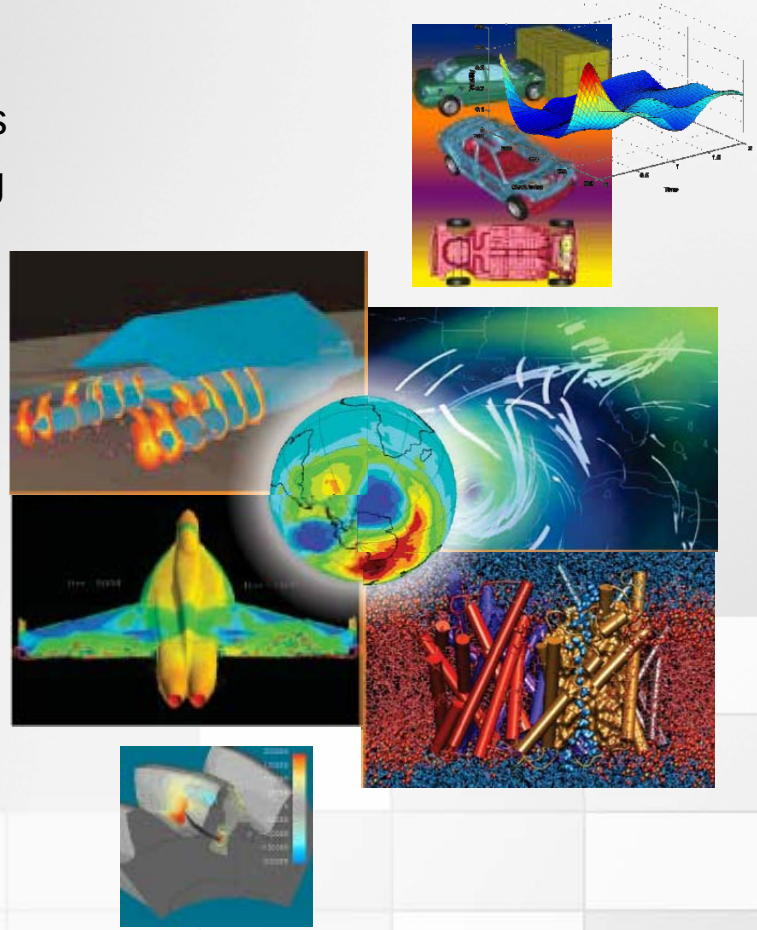
- Also called…
  - High Performance Computing (HPC)
  - Technical Computing

- Cutting edge problems in science, engineering and business require capabilities beyond those of standard desktops and servers

Microsoft Windows Compute Cluster Server 2003

High Performance Computing

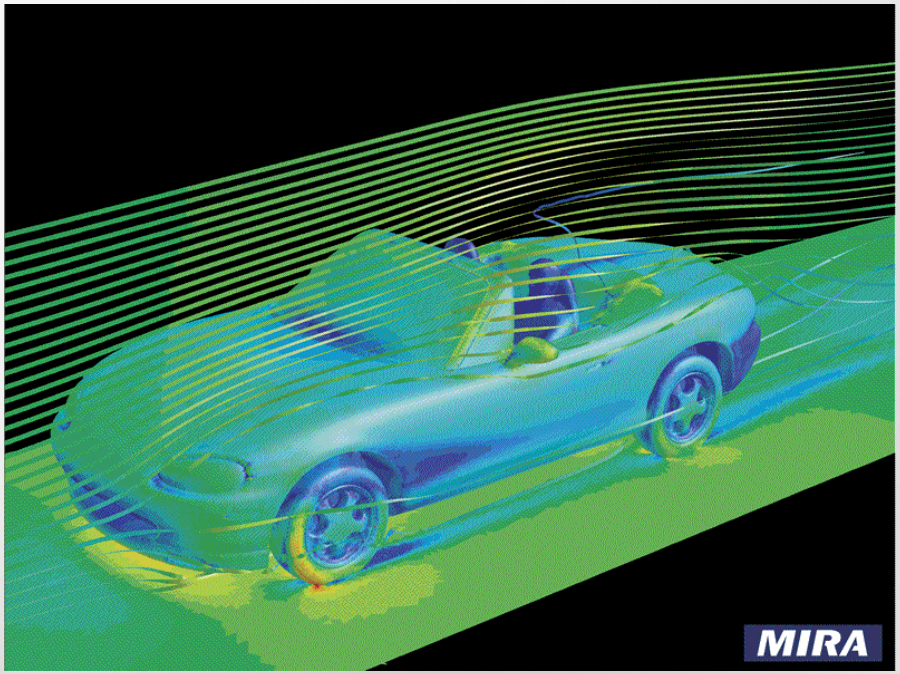
- Cutting edge problems in science, engineering and business always require capabilities beyond those of standalone computers
- Market pressures demand accelerated innovation cycle, overall cost reduction and thorough outcome modeling
  - Aircraft design utilizing composite materials
  - Vehicle fuel efficiency and safety improvements
  - Simulations of enzyme catalysis, protein folding
  - Reservoir Simulation & Seismic Processing
  - Targeted material and drug design
  - Simulation of nanoscale electronic devices
  - Financial portfolio risk modeling
  - Digital content creation and enhancement
  - Supply chain modeling and optimization

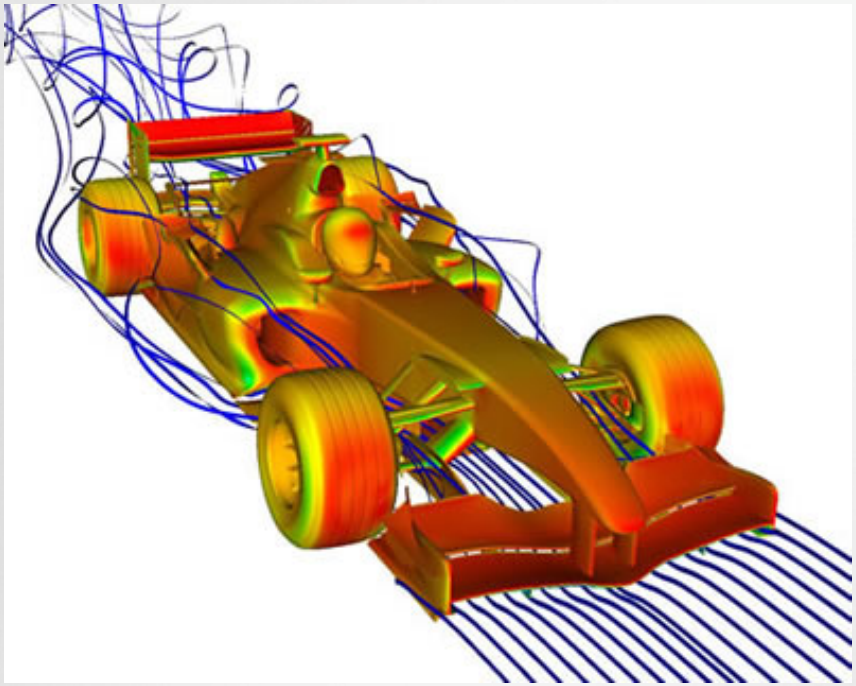


Volume economics of industry standard hardware and commercial software applications are rapidly bringing HPC capabilities to broader number of users

Microsoft Windows Compute Cluster Server 2003

Improving Vehicle Aerodynamics



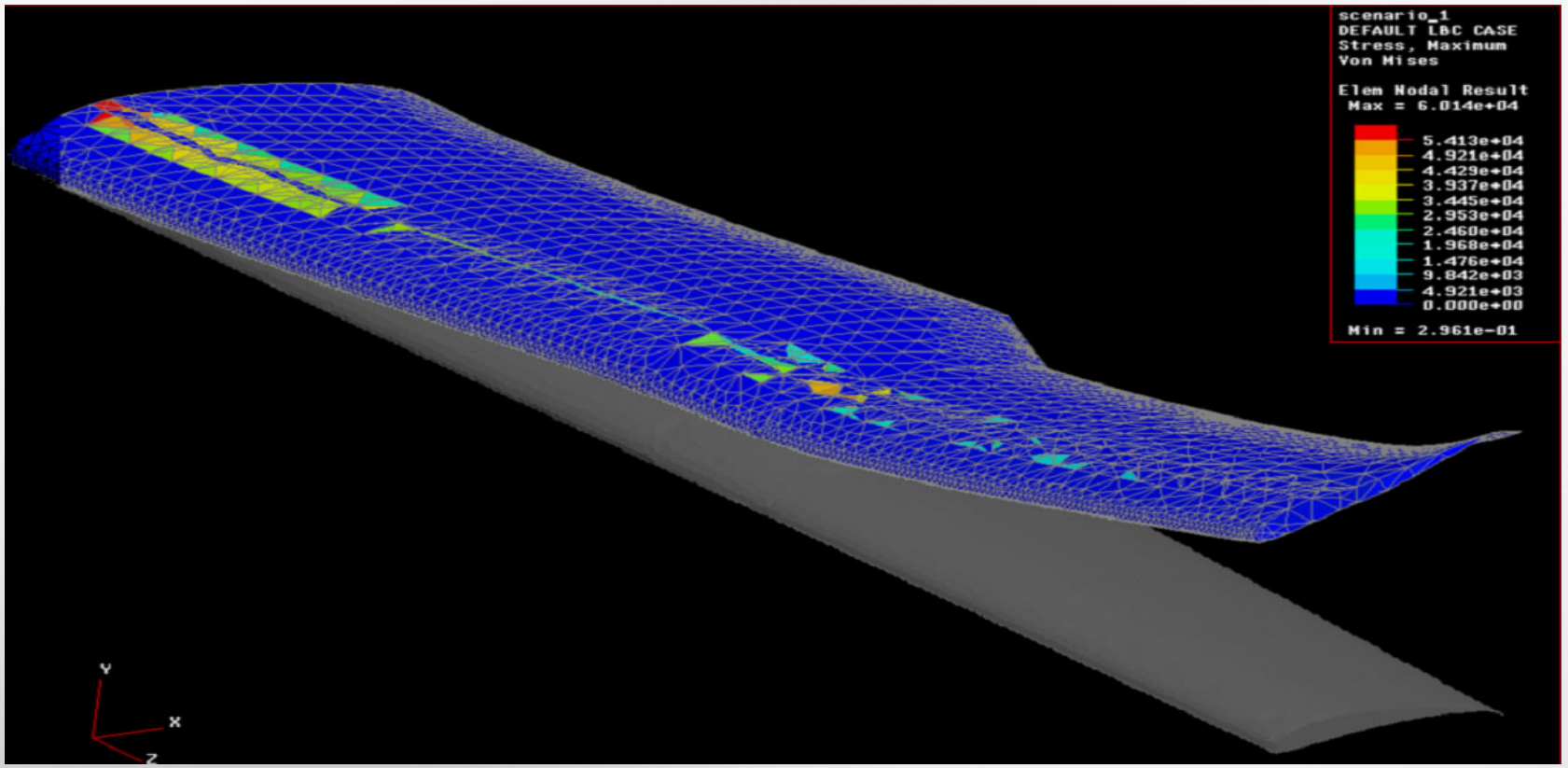


- Computational fluid dynamics (CFD) packages used to model vehicle aerodynamics

Microsoft Windows Compute Cluster Server 2003

Source: www.fluent.com/ about/news/pr/pr48.html ; www.fluent.com

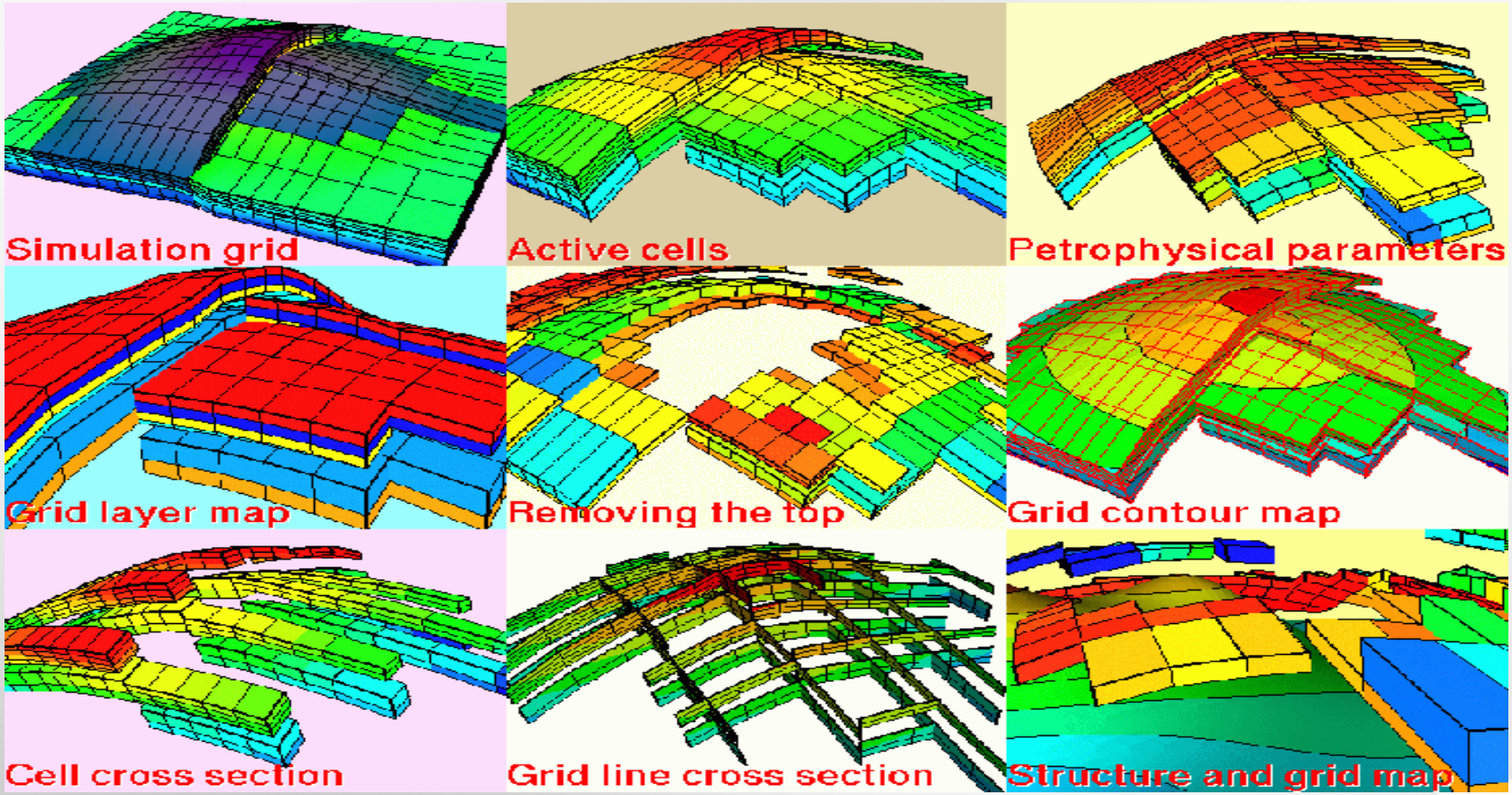
Building Safer Structural Elements



- Finite element analysis (FEA) packages model element stress-profiles under various conditions

Microsoft Windows Compute Cluster Server 2003

More Efficient Oil Extraction

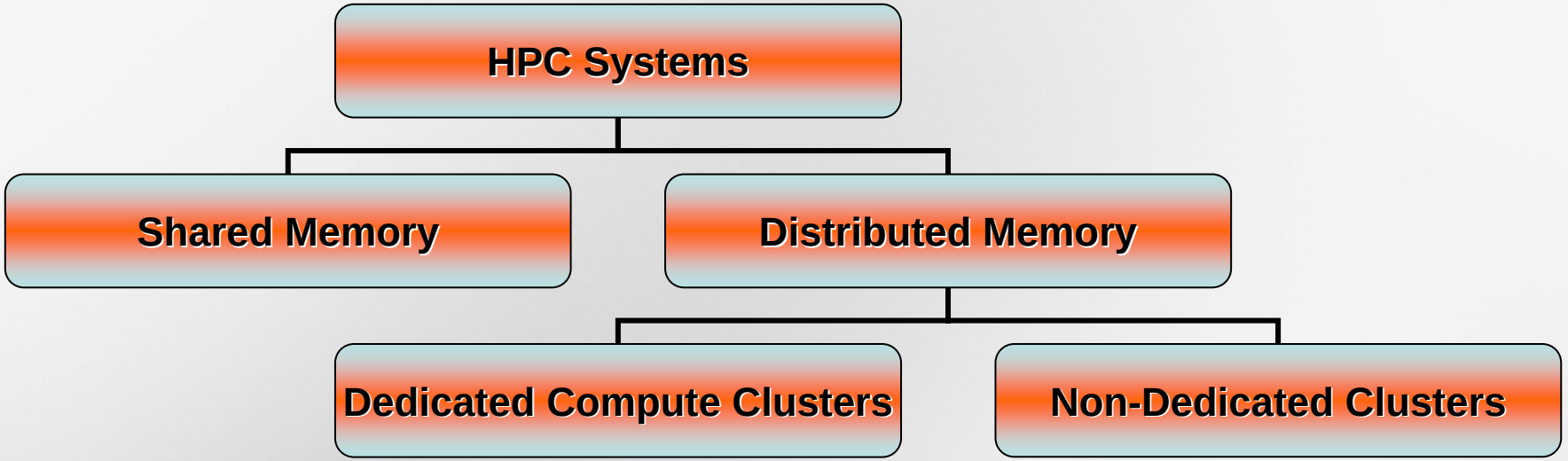


- Reservoir simulation packages model oil/gas reservoir rock and fluid properties

Microsoft Windows Compute Cluster Server 2003

Source: http://www.geocap.no/products/a4.eclipse.shtml

Some Clarification on Terminology …

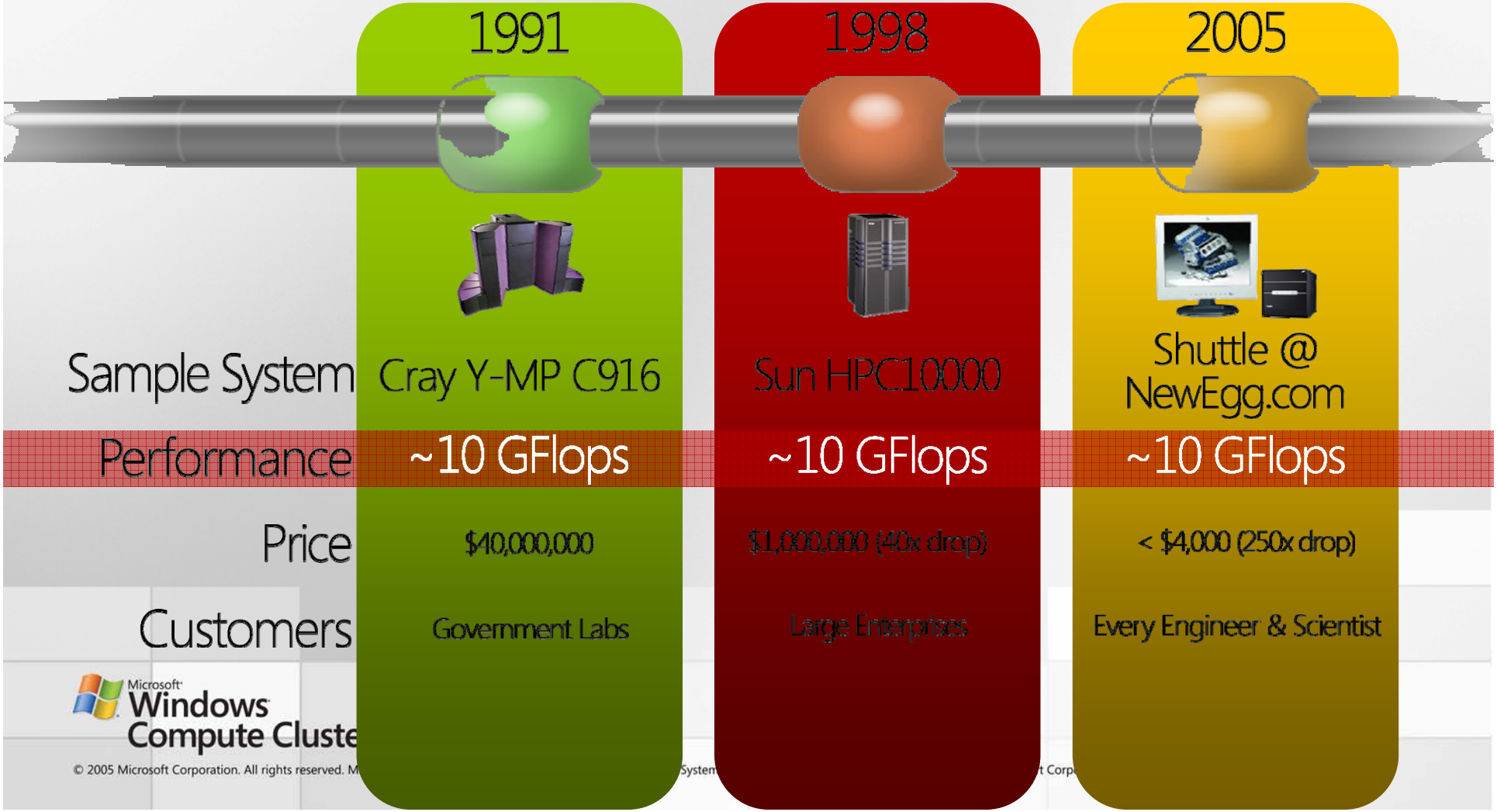


- Collections of computers, networked together and acting as a single large machine
- Each ‘node’ is dedicated to the compute cluster

- Collections of computers, networked together and acting as a single large machine
- Nodes have ‘day-jobs’ and can only provide cycles when unused

Microsoft Windows Compute Cluster Server 2003

Entry Price Point for HPC Has Dropped



Windows Compute Cluster Server 2003

Mission: Deliver the easiest to deploy and most cost effective solution for solving scaled-out business, engineering and scientific computational problems.

Windows Server 2003, Compute Cluster Edition

+ Compute Cluster Pack =

Microsoft Windows Compute Cluster Server 2003

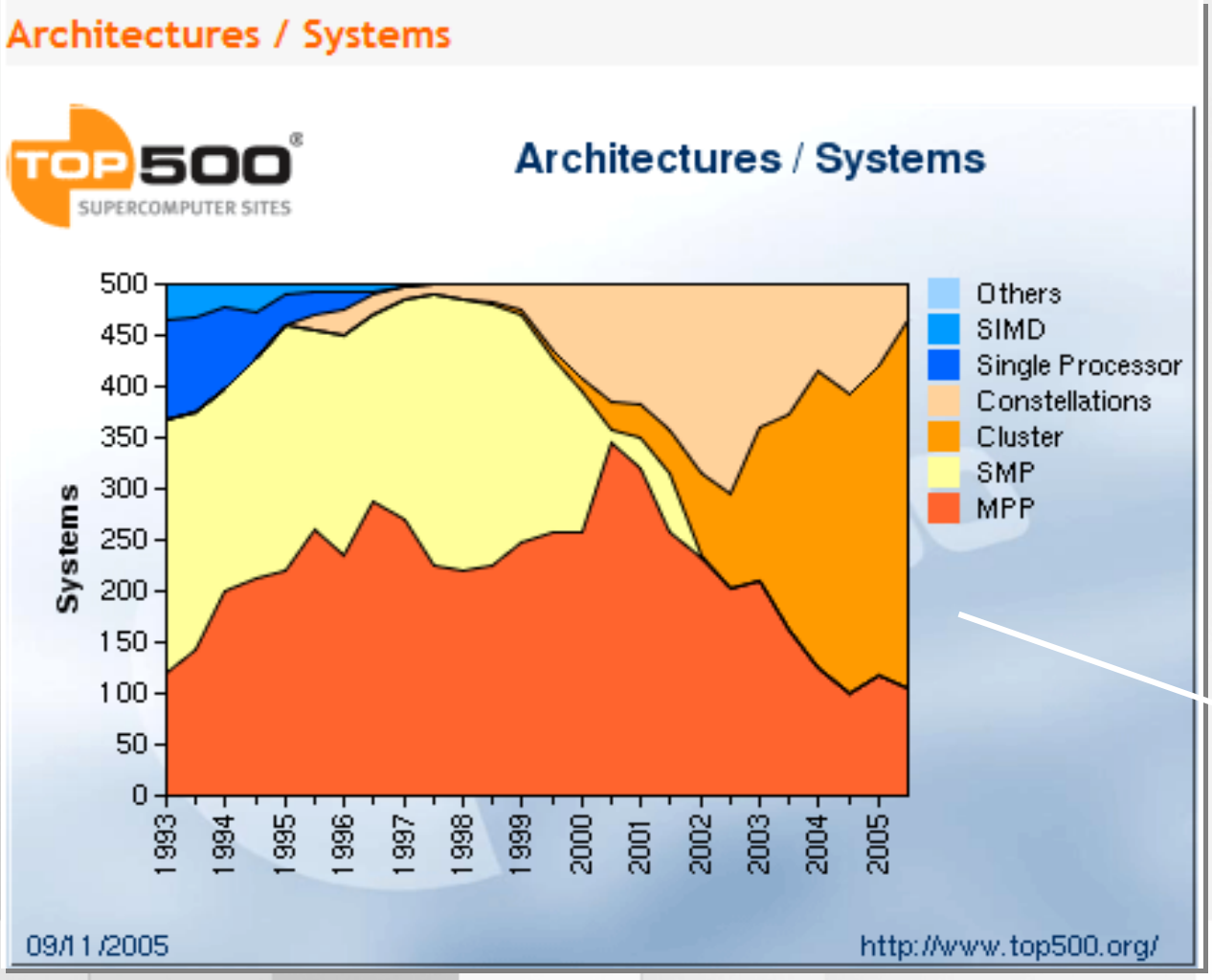
- Support for high performance hardware (x64bit architecture)
- RDMA support for high performance interconnects (Gigabit Ethernet, Infiniband, Myrinet, and others)

- Support for Industry Standards MPI2
- Integrated Job Scheduler
- Cluster Resource Management Tools

- Integrated Solution out-of-the-box
- Leverages investment in Windows administration and tools
- Makes cluster operation easy and secure as a single system

Microsoft Windows Compute Cluster Server 2003

Distributed Memory Becoming the Norm

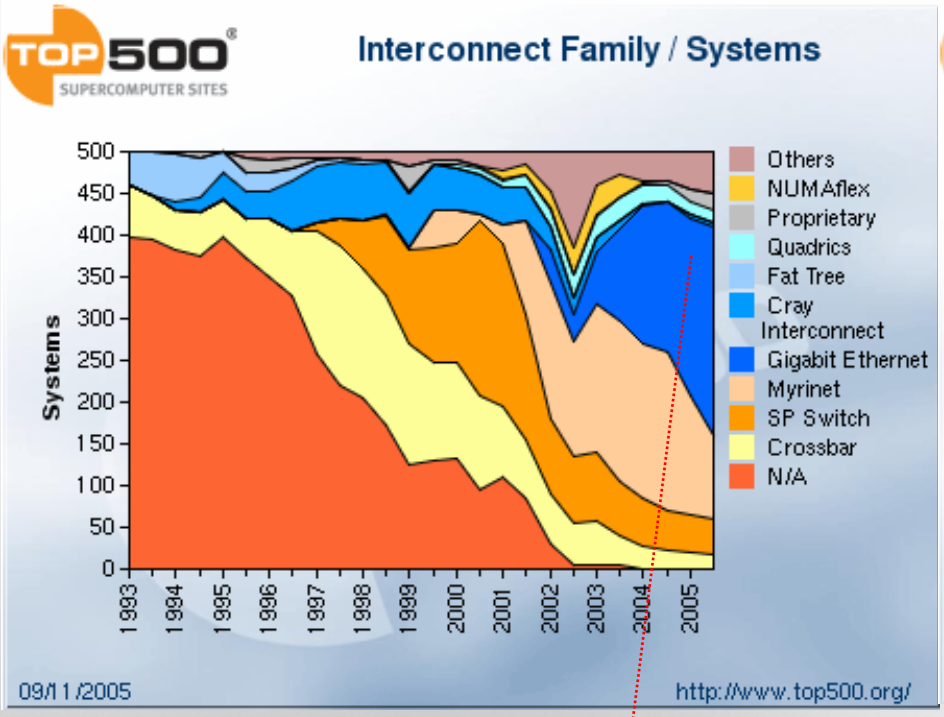


Clusters over 70%

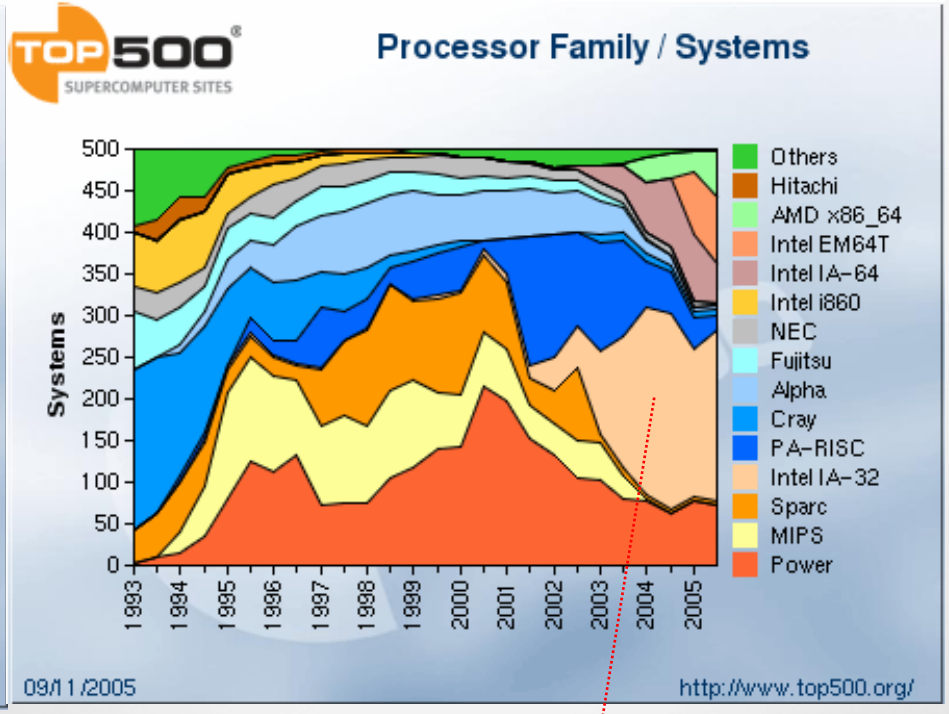
Microsoft Windows Compute Cluster Server 2003

COTS Components Gaining Ground

Interconnect Family / Systems



Processor Family / Systems

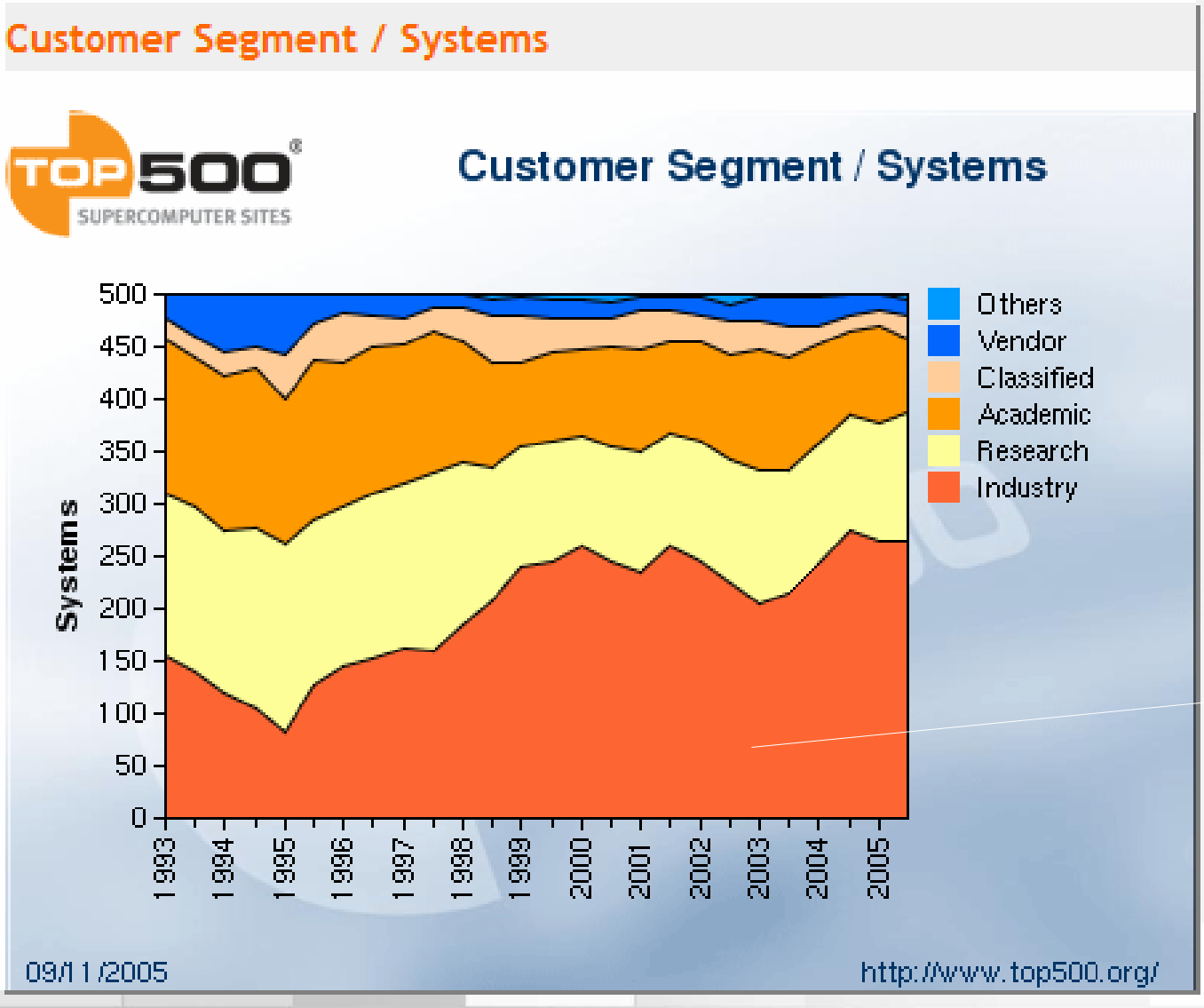


GigE is gaining (50% of systems)

x86 is leading (Pentium 41%,EM64T 16%, Opteron 11%)

Microsoft Windows Compute Cluster Server 2003

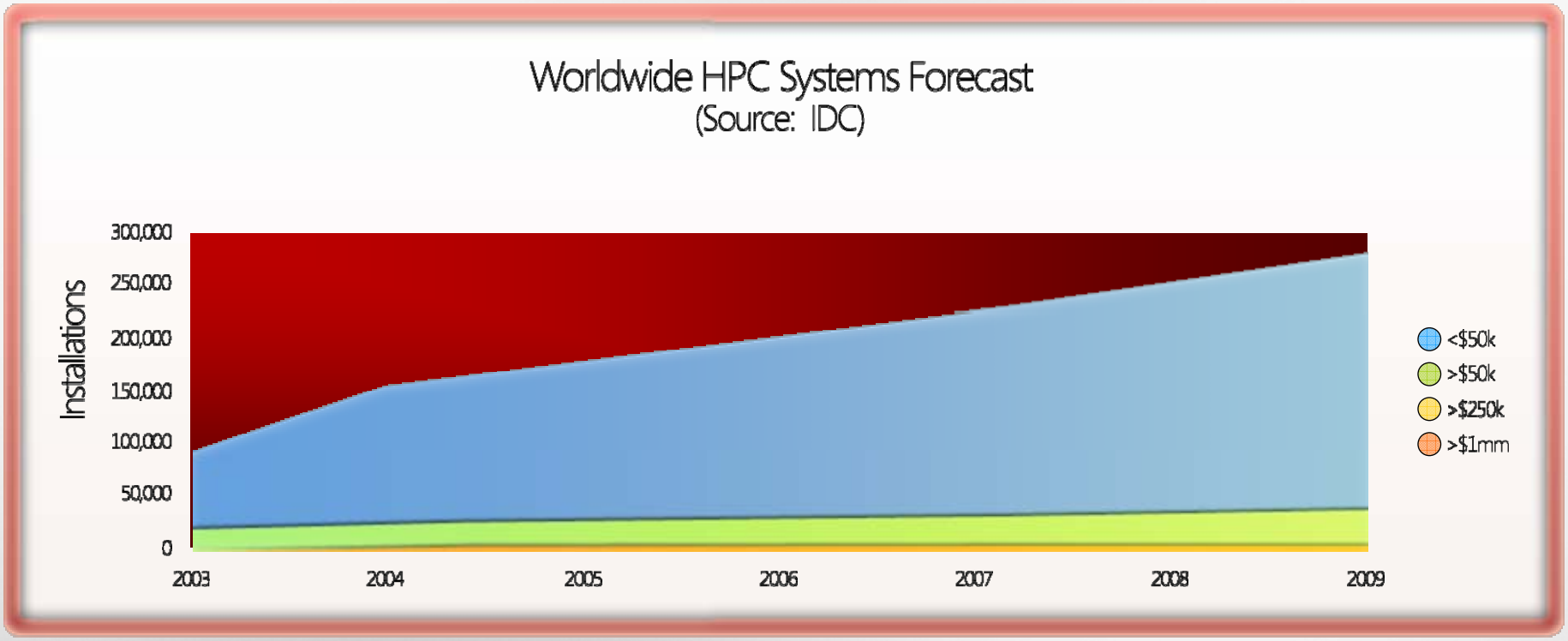
Commercial Adoption Is on the Rise



Industry usage rising

Microsoft Windows Compute Cluster Server 2003

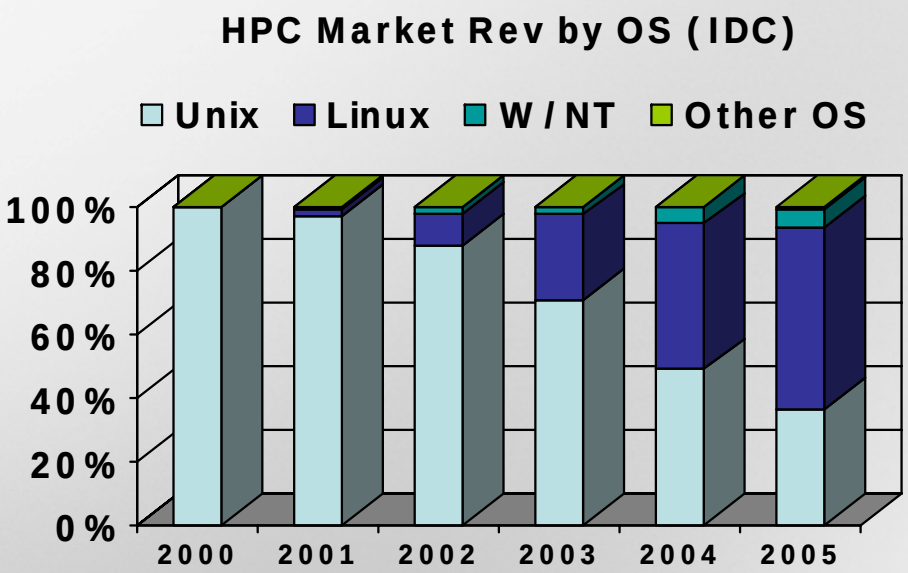
Market Growth Driven by the Low End



- HPC system growth: 70% in 2004; 25% in 2005 (IDC)
- Growth being driven by <$250k segment (clusters)
- x86 server clusters growing faster than market (15%-20% for HPC clusters, 11.4% for x86 overall)

Microsoft Windows Compute Cluster Server 2003

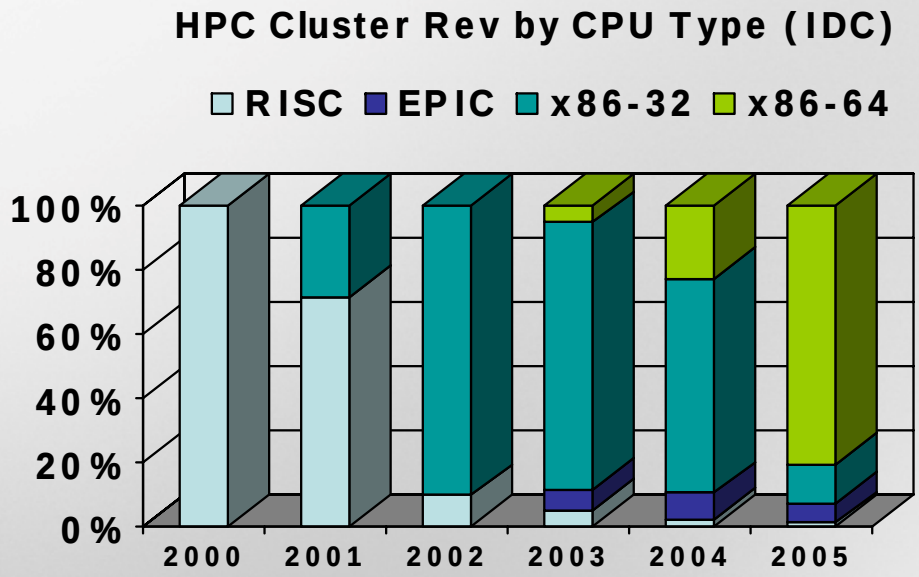
Linux Owns the HPC Cluster Market



- Linux dominates the HPC cluster market
  - Over half of all HPC; Over 90% of all HPC clusters
- HPC is a critical workload for Linux
  - ~35% of all Linux server shipments go into HPC

Microsoft Windows Compute Cluster Server 2003

HPC Is a Market of Early Adopters



- Early adoption is the norm
  - Always interested in the next faster/cheaper/better technology
- Industry shifts in last 5 years alone
  - Unix to Linux
  - RISC to x86-32 to x86-64
  - SMP to clusters
- Disruptive changes are frequent and sweep through the industry quickly

Microsoft Windows Compute Cluster Server 2003

Challenges in HPC Today

End Users

".. want to do more science & less IT"

"..apps not keeping up with ideas"

App availability & integration

Simpler job submission & monitoring

IT Pros

"…takes a village to raise a cluster"

Easier deployment and setup

Simplified mgmt environment

Developers

"...parallel programming is difficult"

"..not enough tools"

Integrated & complete development environment

Platform standards

Microsoft Windows Compute Cluster Server 2003

Microsoft's Vision for HPC

- Empower scientists and engineers to interactively employ massive computational resources to solve complex technical problems

- Make distributed computing resources productive to develop for and cost-effective to operate

Microsoft Windows Compute Cluster Server 2003

Microsoft Strategy

- **Provide a complete Microsoft platform solution for HPC**
  - Enable rapid migration of ISVs to a standard, stable, supported platform
  - Differentiate by providing turnkey clusters that securely integrate into Windows infrastructure
  - Integrate clusters into end user experience through application/desktop
- **Focus on departmental customers in targeted verticals**
  - Build deep expertise for Manufacturing, Geosciences, Life Sciences and Federal
  - Leverage Excel Service to penetrate Financial services
  - Engage customers with SWAT team to drive proof of concept, benchmarks, and joint OEM/ISV engagements
  - Focus on Department & Workgroup clusters
- **Grow and leverage the HPC eco-system**
  - OEM partners to sell pre-configured HPC clusters
  - Secure top HPC cluster ISVs for each targeted verticals
  - Evangelize new application development to our platform
- **Establish credibility in the HPC community**
  - Engage Open Source projects & leverage technologies in HPC stack
  - Create world-wide Innovation Centers around top HPC researchers and practitioners

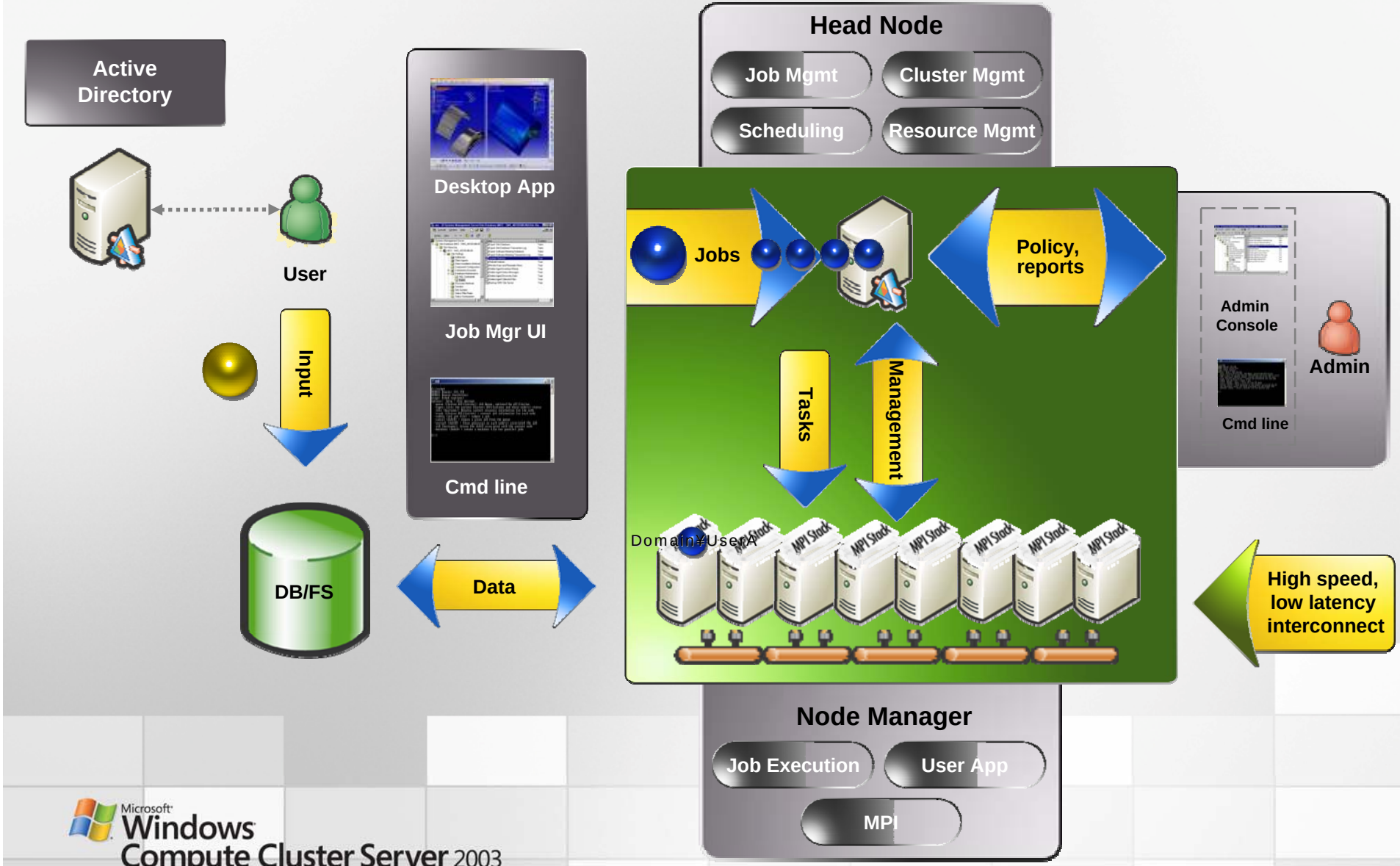
Microsoft Windows Compute Cluster Server 2003

Compute Cluster Server 2003 (CCS)

- Complete, integrated platform for computational clustering
- Core Platform
  - Windows Server 2003 Compute Cluster Edition
  - x64 only
- Cluster Tools
  - Microsoft Compute Cluster Pack
  - Built-in Cluster Mgmt, Job Scheduler & MPI
  - Interconnect support leveraging Winsock Direct
- Availability
  - Bits at http://www.microsoft.com/hpc
  - RTM in June

Microsoft Windows Compute Cluster Server 2003

Windows Compute Cluster Server 2003



Microsoft Windows Compute Cluster Server 2003

The CCS Value Proposition

- Faster time-to-insight
  - …through simplified cluster deployment, job submission and status monitoring
- Integration with Windows infrastructure
  - … allowing customers to use existing technology and skill-sets
- Integrated development environment
  - … allowing developers to write parallel applications from within the Visual Studio IDE

Microsoft Windows Compute Cluster Server 2003

CCS Packaging & Licensing

Packaging

CD 1: Windows Server 2003: Compute Cluster Edition

- *Identical* to Windows Server 2003 x64 operating system
- Priced lower but CCE PIDs restrict usage to computational use only
- Purchase full Windows Server operating system for general purpose server usage

CD 2: Microsoft Compute Cluster Pack

- Additional components for parallel computing
- Job Scheduler, Management Tools, MPI, SDK
- Setup installation options: Head Node, Compute Node, Client machine, SDK

Purchasing

- Three SKUs on pricelists
  - CD 1: Windows Server 2003 CCE
  - CD 2: Microsoft Compute Cluster Pack
  - CCS: Combines CD 1 and CD 2
- Marketing focus will be on combined offering only (Windows Compute Cluster Server 2003)

Channels

- OEM, System Builder, Embedded
- Volume Licensing
- MSDN Academic Access
- Academic OEM and VL

Microsoft Windows Compute Cluster Server 2003

CCS Key Features

Easier node deployment and administration
- Task-based configuration for head and compute nodes
- UI and command line-based node management
- Monitoring with Performance Monitor (Perfmon), Microsoft Operations Manager (MOM), Server Performance Advisor (SPA) & 3rd-party tools

Extensible job scheduler
- 3rd-party extensibility at job submission and/or job assignment
- Submit jobs from command line, UI, or directly from applications
- Simple job management, similar to print queue management

Secure MPI
- User credentials secured in job scheduler and compute nodes
- Microsoft provided stack reduces application/MPI incompatibility issues

Integrated Development Environment
- OpenMP Support in Visual Studio, Standard Edition
- Parallel Debugger in Visual Studio, Professional Edition

Microsoft Windows Compute Cluster Server 2003

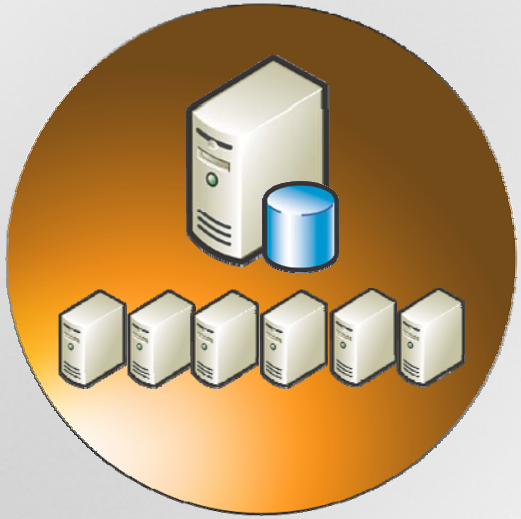
Leveraging Existing Windows Infrastructure

Active Directory

- Integration with IT infrastructure
- Resource management
- Group policies

Windows Security

- Kerberos authentication
- Secure job execution
- Secure MPI



Microsoft Windows Server 2003 Compute Cluster Edition

Microsoft Enterprise Management Tools

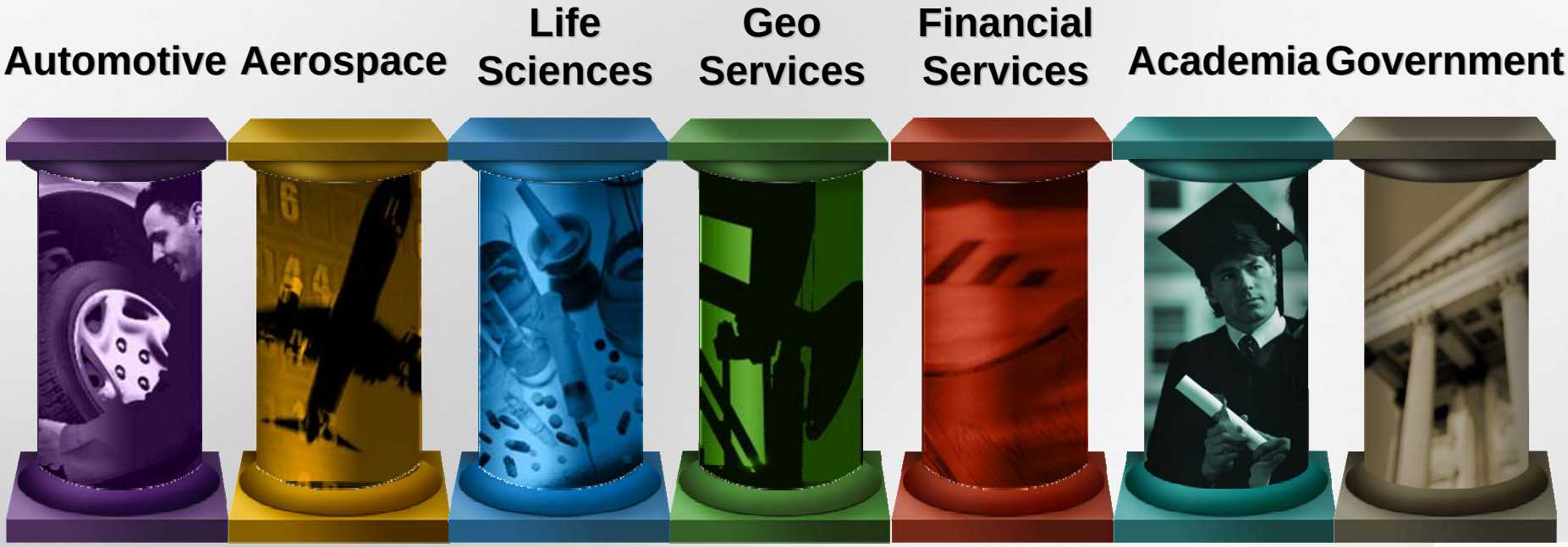
- Operations manager
- Windows Update services
- Systems Management Server

Compute Cluster Built-in Tools

- Job scheduler
- Admin console
- Performance monitor
- Command line interface
- Remote Installation services

Microsoft Windows Compute Cluster Server 2003

Customer Focus



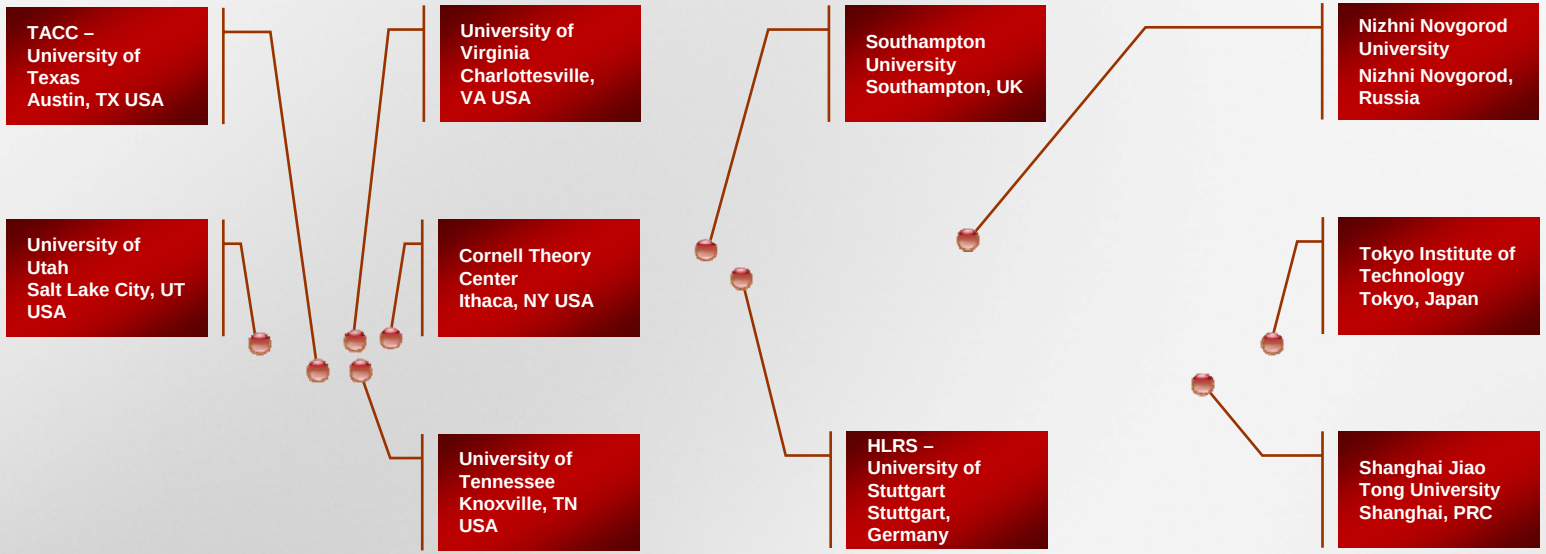
Microsoft Windows Compute Cluster Server 2003

CCS Positioning

- In v1, CCS is targeted at *dedicated compute clusters*
- Expect the first customers to be Windows shops with a need for HPC
- Expect the most traction in departmental and workgroup clusters
  - Systems under 64 nodes make up over 95% of cluster installations
  - As HPC continues to penetrate the mainstream, these are expected to be the highest-growth segments
- Does NOT mean that CCS cannot handle large clusters
  - Bill Gates keynote demo at Supercomputing'05 used 512 CPUs running the CCS Beta 2; Top 500 number expected in June will be the largest CCS cluster to date (~900 processors)
  - Packaging allows customers flexibility to replace components of the stack

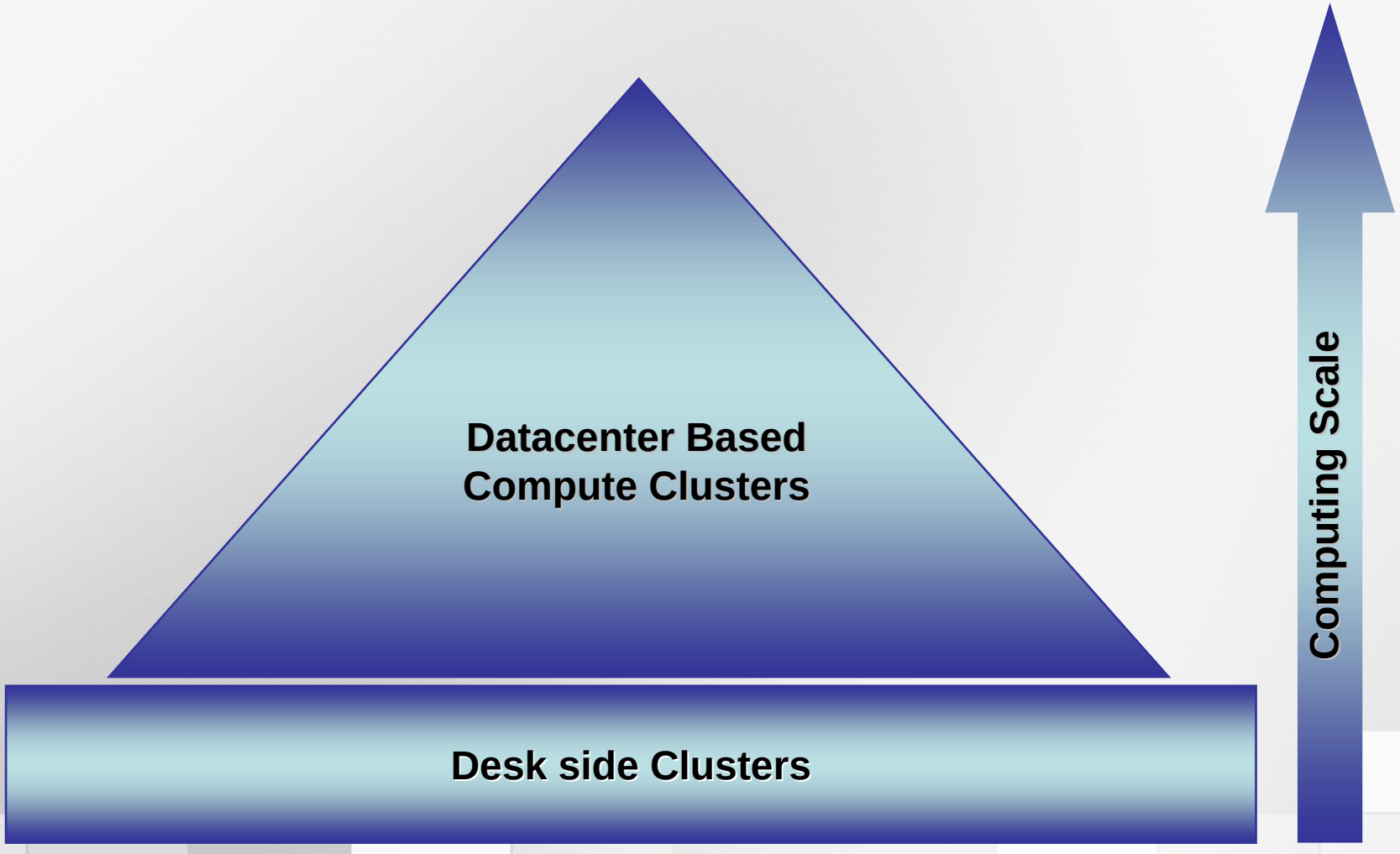
Microsoft Windows Compute Cluster Server 2003

HPC Institutes



Microsoft Windows Compute Cluster Server 2003

Bringing HPC mainstream



Microsoft Windows Compute Cluster Server 2003

Desk-Side Supercomputing

- End users that are caught between the desktop and the datacenter
  - Need access to 4-8 nodes for *most* of their processing
  - Might have access to larger HPC systems but want ‘in-lab’ clusters for development purposes
  - Allows user to refine algorithms and code before deployment on larger departmental clusters
- User profile – Engineers and scientists solving “what if” problems in:
  - Life sciences
  - Computer aided engineering
  - Finance
  - Oil and gas/geophysics

Microsoft Windows Compute Cluster Server 2003

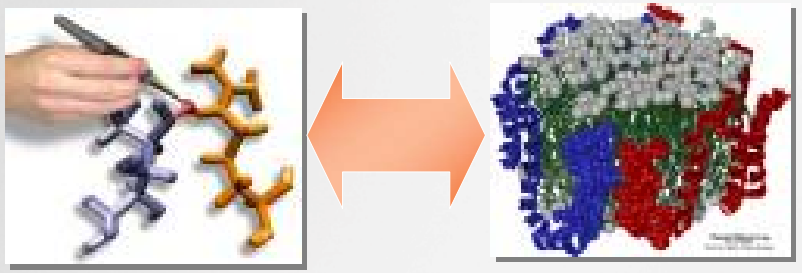
What does a desk-side cluster look like?

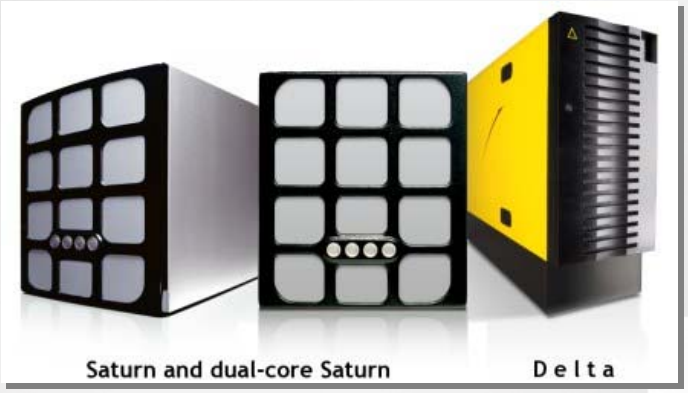
- True high performance parallel computers
- Small enough for the office and lab
- Powered from a standard wall outlet
- Cool running and quiet
- Priced under single-user purchase limits
- As easy to use as a PC
- Comes with OS pre-installed (and potentially app as well)

Microsoft Windows Compute Cluster Server 2003

Desk-side Scenario: Molecular Modeling

- Personal 4 node cluster
  - Per node: Dual core, dual proc x64, 2GB RAM
  - Dual NIC (Ethernet + GigE or IB)
  - 1 head node, 3 compute nodes
  - 110v pluggable, low noise, desk-side tower
- Front end workstation used to build, view and interact with the model in 3D
- Coarse-grained simulations run against desk-side cluster
- Larger models handed off to datacenter-based resources





www.rocketcalc.com

http://www.tyan.com/products/html/clusterservers.html

Microsoft Windows Compute Cluster Server 2003

Summary

- Parallel computing is here to stay
- Windows Compute Cluster Server 2003 is Microsoft's first step towards bringing HPC into the computing mainstream
- CCS will serve existing scenarios…
  - Datacenter-based compute clustering
- …and drive momentum around new ones
  - Desk-side supercomputing

Microsoft Windows Compute Cluster Server 2003

Call To Action

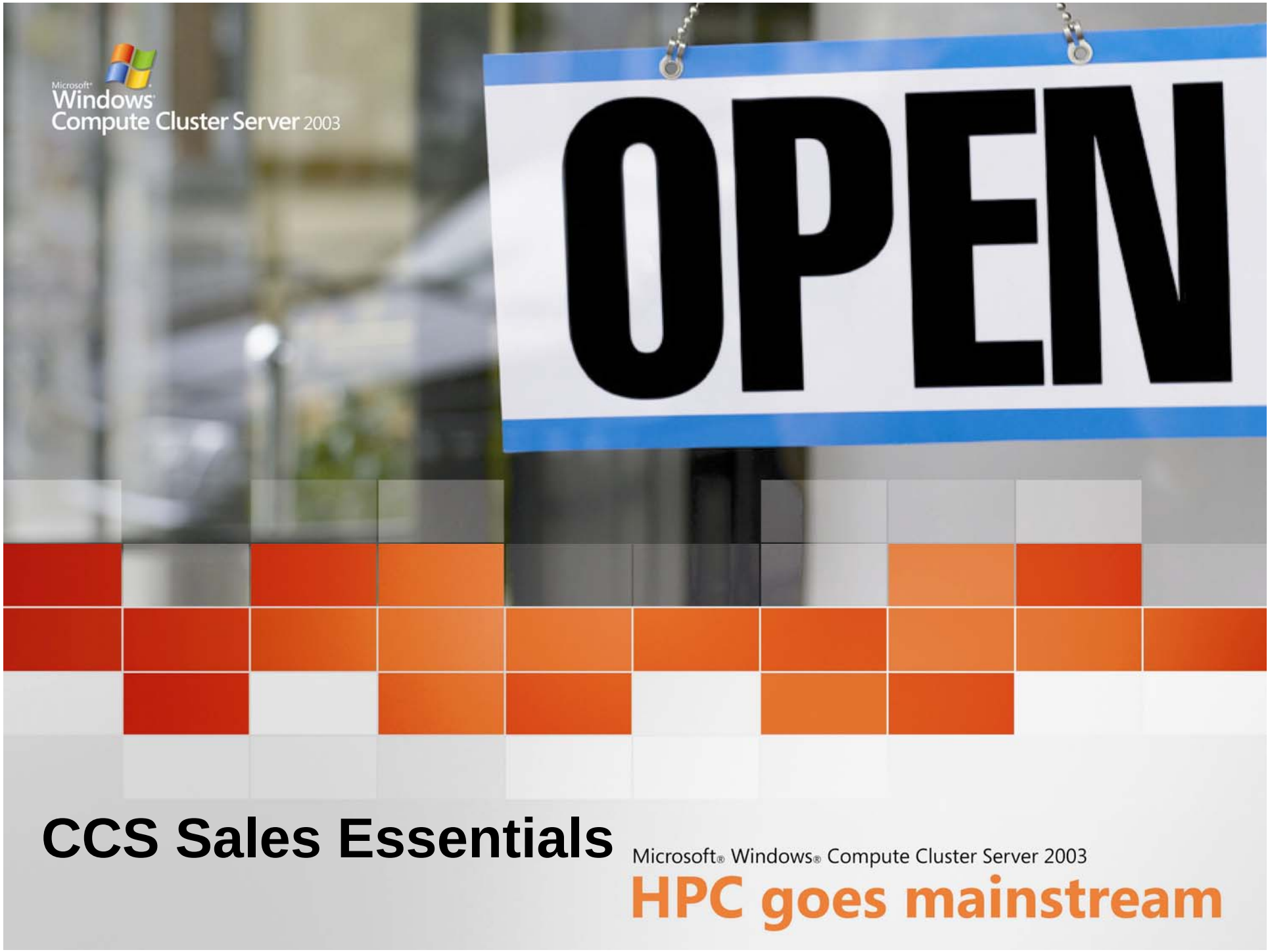
- Download Compute Cluster Server 2003
  - www.microsoft.com/hpc
- Build, test and prepare to ship products on CCS
- Raise awareness of HPC among mainstream consumers

Microsoft Windows Compute Cluster Server 2003

Resources

- Microsoft HPC Web site
  - http://www.microsoft.com/hpc/
- CCS Community Site
  - http://www.windowshpc.net
- Windows Server x64 information
  - http://www.microsoft.com/x64/
- Get the Facts Web site
  - http://www.microsoft.com/getthefacts

Microsoft Windows Compute Cluster Server 2003



CCS Sales Essentials

Microsoft® Windows® Compute Cluster Server 2003

HPC goes mainstream

Overview

Objective

- Provide guidance on:
  - Positioning and selling CCS
  - Opportunity identification

What this presentation covers:

- What does a CCS customer look like?
- What are the key scenarios they would use CCS for?
- What's the customer value proposition?
- How does CCS measure up against the competition?
- Objections and how to handle them?

What this deck is not

- Product overview of CCS
  - Refer to http://www.microsoft.com/hpc for additional content
- Linux Compete Content
  - Refer to http://www.microsoft.com/getthefacts

Recap: Takeaways from Previous Session

- HPC is …..
  - A class of problems that require more computing power than can be delivered through standard desktops and servers
  - HPC systems are broadly divided into two categories
    - Shared memory systems – SMP, NUMA
    - Distributed memory systems – Compute Clusters
- Dedicated compute clusters are…
  - A collection of industry-standard servers that can be ‘clustered’ together to serve as a single, large, parallel-processing machine
- HPC cluster shipments are growing fast
  - 10% of WW x86 shipments now in HPC; outpacing overall x86 market ramp
- Microsoft’s first step into HPC
  - Windows Compute Cluster Server 2003
  - Complete, integrated platform for dedicated computational clustering

Microsoft Windows Compute Cluster Server 2003

Recap: Customer Challenges in HPC Today

Hard to use

- System setup, deployment & maintenance is complex
- End-users doing too much IT
  - Poor integration from user-desktop to back-end cluster
  - Job submission requires scripting, programming
- Systems management is a challenge in heterogeneous environments
  - Organizations want a single platform that can accommodate all workloads

Not enough apps

- ISV eco-system still waking up to HPC
- Reliance on open source codes customized for local use

Difficult to write code for distributed platforms

- Parallel programming is a new (& difficult) concept
- Tool support for parallel code development is lagging
- Lack of standards makes portability difficult (OS, MPI, supporting libraries)

"Make high-end computing easier and more productive to use. Emphasis should be placed on time to solution, the major metric of value to high-end computing users… A common software environment for scientific computation encompassing desktop to high-end systems will enhance productivity gains by promoting ease of use and manageability of systems."

High-End Computing Revitalization Task Force, 2004
(Office of Science and Technology Policy, Executive Office of the President)

Recap: CCS Packaging & Licensing

Packaging

- CD 1: Windows Server 2003: Compute Cluster Edition
  - Identical to Windows Server 2003 x64 operating system
  - Priced lower but CCE PIDs restrict usage to computational use only
- CD 2: Microsoft Compute Cluster Pack
  - Additional components for parallel computing
  - Job Scheduler, Management Tools, MPI, SDK

Purchasing

- Three SKUs on pricelists
  - CD 1: Windows Server 2003 CCE
  - CD 2: Microsoft Compute Cluster Pack
  - CCS: Combines CD 1 and CD 2
- Marketing focus will be on combined offering only (Windows Compute Cluster Server 2003)

Channels

- OEM, System Builder, Embedded
- Volume Licensing
- MSDN Academic Access
- Academic OEM and VL

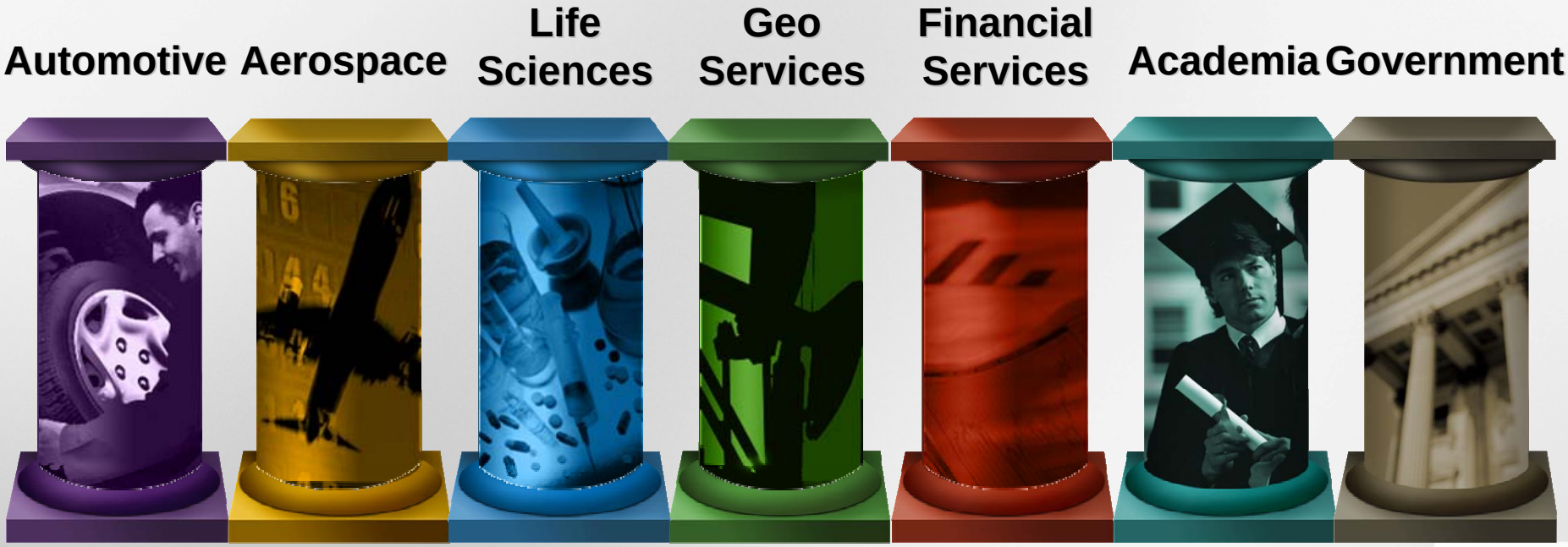
Microsoft Windows Compute Cluster Server 2003

CCS Market Strategy

- Build a complete platform
  - Operating System, Message Passing Interface (MPI), Job Scheduler, Cluster Management
- Develop the eco-system
  - Hardware platform and interconnect support
  - Application support from key players
- Raise awareness
  - Evangelize parallel computing
- Partner with the community
  - HPC Institutes for thought leadership
  - Development of cross-platform standards
  - Leverage (and give back) to the open-source community

Microsoft Windows Compute Cluster Server 2003

Customer Focus



Microsoft Windows Compute Cluster Server 2003

Application Scenarios: Manufacturing/Engineering

Application Domains

- Finite Element Analysis, Computational Fluid Dynamics, Crash Simulation

Scenarios

- Aerospace:
  - Bird-strike design, temperature fluctuations, acceleration loads
  - Simulations to test the viability of composite materials in airframe design
  - Flow simulations to reduce drag coefficient
- Automotive
  - Power-train and engine analyses to reduce emissions
  - Aerodynamic studies to raise fuel efficiency & performance
  - Vehicle Crash simulations

Product Fit

- CCS ideally suited for departmental & workgroup clusters commonly used in this sector

Application Scenarios: Geo Sciences

Application Domains

- Reservoir Simulation, Seismic Processing

Scenarios

- Reservoir Engineers modeling the most efficient way to pump oil out of a reservoir by introducing water or natural gas into the reservoir under pressure
- Seismic Processing typically uses very large clusters (1000-2000 nodes) with proprietary or 3rd party schedulers and management packs

Product Fit

- CCS for reservoir simulation
- For larger seismic processing clusters, customers might want only CD 1
- Cross sell: SQL, MOM, SMS

Microsoft Windows Compute Cluster Server 2003

Application Scenarios: Financial Services

Domains

- Financial Analytics

Scenarios

- Monte Carlo simulation, Algorithmic trading, Model portfolio work
- Default risk analysis, Diversification analysis
- Sophisticated portfolio performance measurement
- Fraud detection and money-laundering alerts
- Econometrics

Product Fit

- CCS for workgroup and smaller personal (4-8 node) clusters
- CCE for larger clusters and/or if customer already has scheduler and other cluster management tools
- Cross-sell: Excel Server, Excel 12 (multi-threaded), SQL, MOM, SMS

Microsoft Windows Compute Cluster Server 2003

Application Scenarios: Life Sciences

Application Domains

- Genomics, Proteomics, Drug Design

Scenarios

- Computational Chemistry
- Chemical Informatics
- Bio-Informatics
- Computational Biology

Product Fit

- CCS ideally suited for personal and workgroup cluster sizes in this segment

Application Scenarios: Government & Academia

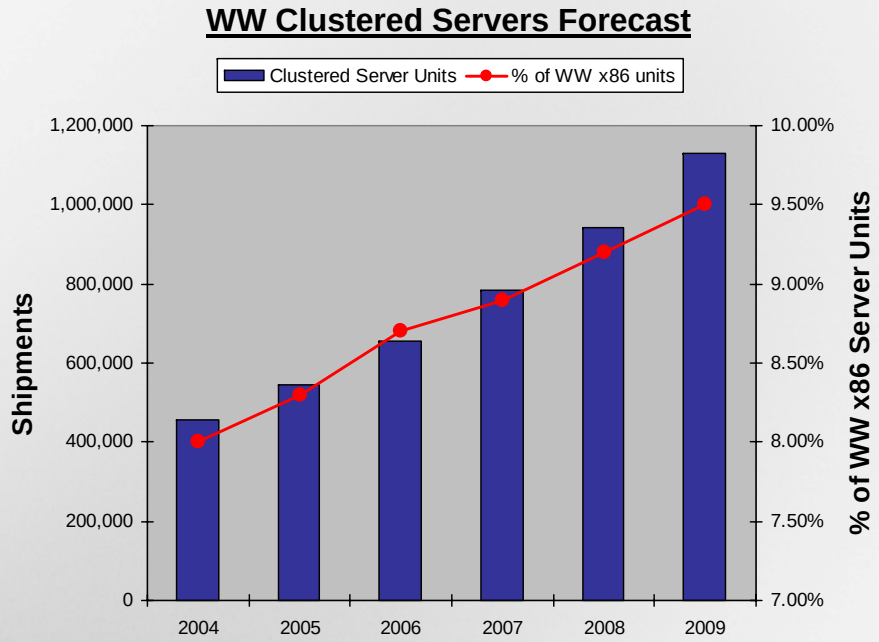
Characteristics

- Strong influencers of commercial decisions
- Will often leverage other industry applications
- .. Or create open-source codes that grow into widespread use
- Often funded by commercial entities

Sample targets

- Public Sector
  - Center for Disease Control - Epidemic evolution modeling
  - Department of Energy - Seismic processing
  - Department of Defense - Aircraft wing design/fluid dynamics
  - Department of Homeland Security - Disaster impact and response modeling
  - Nationally funded research labs; Public/Private Consortiums
- Academia
  - Departments: Physics, Biology, Computer Science, Engineering
  - Open-source codes often see widespread commercial use

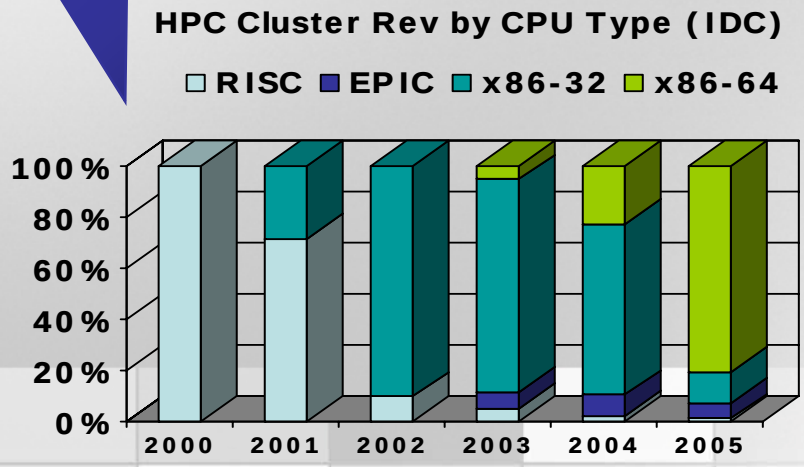
HPC Market Trend

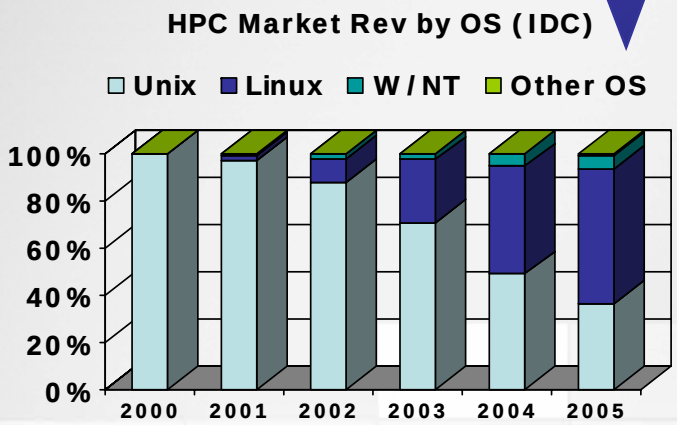


(IDC)

HPC clusters more than 50% revenue, x64 dominates

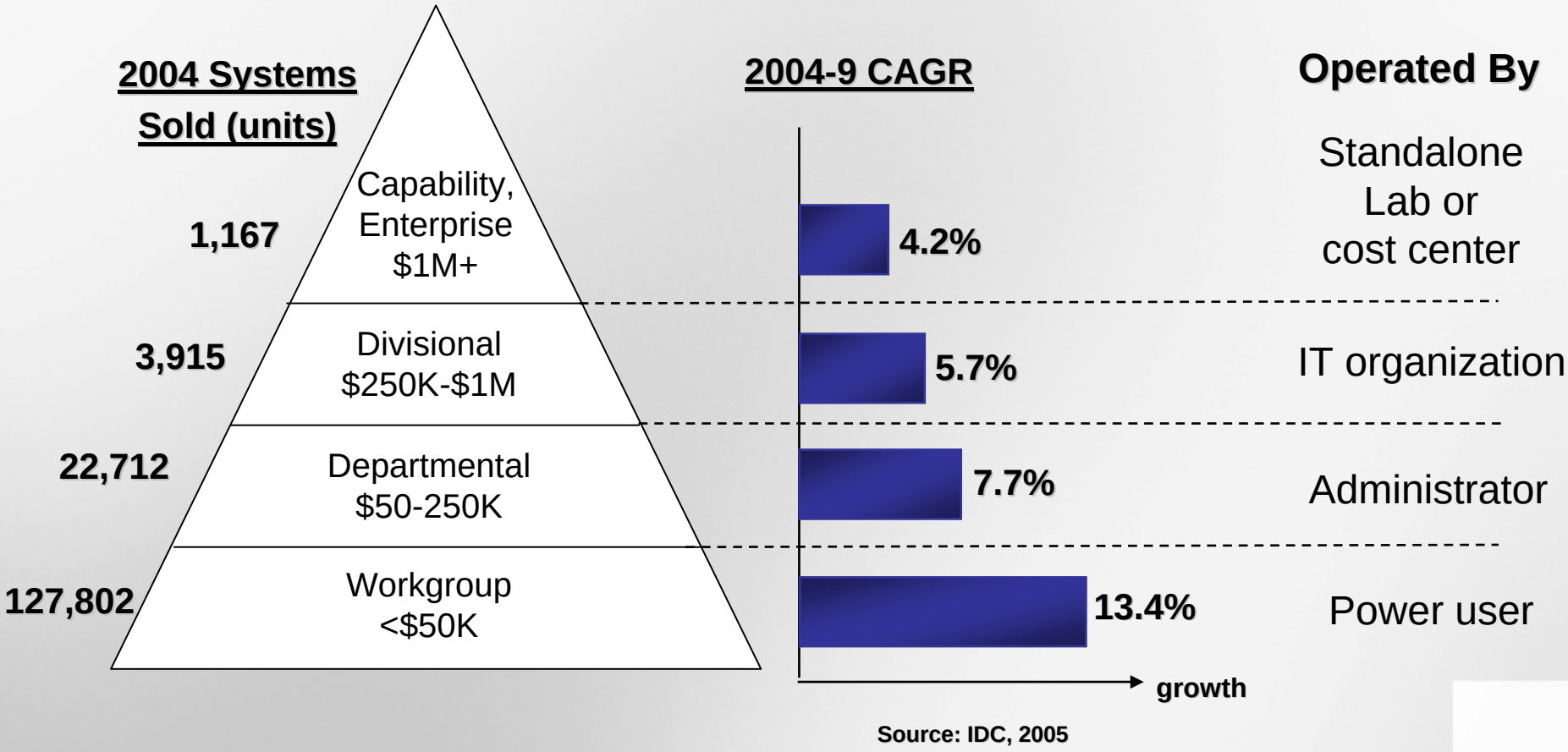
Linux is replacing Unix
Windows market is growing





Microsoft Windows Compute Cluster Server 2003

HPC Market Segments & Growths



Source: IDC, 2005

<$250K – 97% of systems, 52% of revenue.

In 2004 clusters grew 96% to 37% by revenue

Microsoft Windows Compute Cluster Server 2003

Opportunity

- Departmental & Workgroup segment is the under-represented majority
  - "Takes too long to setup"
  - "Hard to maintain and update"
  - "Don't integrate with end-user apps"
- Needs
  - ***Customers*** require simple job submission and compute performance
  - ***Administrators*** require simplified IT environment
  - ***Developers*** require standards-based environment and advanced tools

Microsoft Windows Compute Cluster Server 2003

Source – WW Server Forecast IDC, 2005, MS Analysis

Competitive Offerings

- IDC estimates that Linux runs on over 90% of x86-based HPC systems
- Competitive Linux/OSS solutions are a collection of technologies used by commercial vendors
  - Many different stacks/solutions out there with questionable inter-operability
- Commercial solutions for cluster computing fall into two categories
  - Component-based: Various stack components bundled together or sold separately
  - Integrated: Entire stack built, purchased and deployed as one solution
  - Integrated solutions are easier to deploy; easier to maintain consistency in versions across different layers in the stack
- Key areas of differentiation across solutions
  - Processor architectures supported
  - Interconnect support
  - Ease of cluster deployment
  - Cluster administration
  - MPI implementation
  - Job scheduling capability
  - Supporting tools – math libraries, performance tuning, debugging etc

Pricing Comparison

- Why it's hard to do a straight-up comparison
  - Lots of different software stacks none of which are identical
  - Services component is significant for OEM/SI and is part of the overall price package
- Customer perception will be that Linux is 'free'
  - End-users often unaware of support costs
  - Tendency to fixate on up-front costs not total TCO
- CCS value/price should compare very favorably with Linux
  - This assumes supported Linux stacks with equivalent functionality
- Will find more traction in commercial environments than in Academia
  - CCS academic pricing is lower than commercial pricing
  - .. But academics not as sensitive to support costs and TCO

CCS vs. Competitive Offerings (Head to Head)

	Compute Cluster Server 2003	Linux-Based Solutions
Cost	Startup costs equivalent or slight premium Lower cost of ownership over system lifetime due to 1-time pricing	Lower initial software price for some customers, some base software free Subscription-based pricing spreads out cost over multiple years
End-User experience	Potential for better integration to the (Windows-based) desktop and existing tools (Word, Excel, Windows apps)	Users often unfamiliar with Unix¥Linux must remote-in to Linux head node to submit jobs
IT Integration	Better integration to existing Windows-based infrastructure common to most commercial environments	Have to carve out 'Linux islands' in commercial environments. The opposite is true for Unix-friendly environments typical to academic & research labs
Deployment & Maintenance	Integrated stack from single-vendor simplifies deployment and ongoing maintenance	Integrated offerings are rare. Support for the entire stack can be purchased but typically includes multiple vendors
Support	More consistent vendor support	Multi-vendor support issues may be mitigated by hiring experienced sys administrators
IT skill sets	Cluster deployment requires specialized skills Larger pool of Windows IT Pros and devs to hire from	Cluster deployment requires specialized skills Larger pool of HPC-proficient IT Pros and developers
Security	CCS jobs run with the AD context set for the user submitting the job	Many installations seem to run all jobs on the cluster under a single user-account with no differentiated permissions.
Dev Environment	Superior & well-integrated development environment	• Dev tool environment is more primitive • Better support for parallel code development & HPC (debugging, tuning, numerical libraries)
Storage	No parallel file system implementation in v1	Many options for Parallel File Systems
Scalability	Perception that "CCS is only good for small clusters" will continue multiple in-production proof points to counter. Top 500 number available by June	Lots of Top 500 & customer proof-points to showcase scale-out capability
Ecosystem	Application eco-system is small but includes major players; Strong ISV interest in standardizing on Windows platform Microsoft acknowledged as having the power to drive a standard platform	Larger pool of available apps – commercial and open-source ISVs frustrated by lack of standard s/w stack

Scenario #1: Windows Shop running HPC Clusters on Linux

Perspective	Why CCS?
Business User	**Tighter desktop integration allows users to do more science and less IT** • Easy to integrate from desktop apps to remote cluster. Job creation/submission can be done from within the app and CCP includes a GUI-based client tool for job monitoring on remote clusters. • AD integration provides single sign-on capability.
IT Professional	**Better TCO** • Leverage existing Windows admin skill-sets to manage HPC systems • Lower costs over system lifetime than on Linux **Integration to Windows IT Infrastructure** • CCS integrates to AD; eliminating multiple user logins and ensuring that job access permissions map to AD policies/profiles • CCS nodes are Windows Server machines that can be managed using existing infrastructure such as MOM, SMS etc
Developer	**Comprehensive and familiar development environment** • Can use the integrated Visual Studio IDE • VS 2005 includes a parallel debugger • Intel compilers and dev tools for HPC being ported to Windows

Microsoft Windows Compute Cluster Server 2003

Scenario #2: Windows Shop not using HPC clusters

Why HPC?

- Is business user productivity constrained by performance limits of workstation-based environments?
  - Bioinformatics: Matlab runs on single servers that take up to 40 hours
  - Financial Analytics: Brokers unable to manipulate large Excel models even on high-end workstations
- Seeking competitive advantage
  - Companies in target verticals already using HPC to speed up the innovation cycle and reduce costs

Why HP Clusters?

Flexible scale-out	• Clusters can be expanded as needed, using non-proprietary components, and have much more headroom than SMP machines
Clusters have market momentum	• SMPs are giving up HPC market share to clusters. ISV support for SMP will continue to decline as this trend continues • The eco-system for available apps on compute clusters now includes industry-leading solutions for Auto, Aero, Oil & Gas etc
Leverage Windows skills	• CCS is based on Windows Server 2003 and allows IT Pros to leverage existing skill sets for server deployment and management
Lower cost	• Clusters deliver more performance for the dollar than SMP systems

Microsoft Windows Compute Cluster Server 2003

Scenario #3: Unix¥Linux shop running HPC on Linux

Perspective	Why CCS?
Business User	**Tighter desktop integration allows users to do more science and less IT** • Easy to integrate from desktop apps to remote cluster. Job creation/submission can be done from within the app and CCP includes a GUI-based client tool for job monitoring on remote clusters. • AD integration provides single sign-on capability.
IT Professional	**Better Security Model** • CCS integrates with Active Directory, allowing admins to set access policies once (within AD). Jobs are then submitted and run under the security context set for the submitting user **Integrated software stack with single-vendor support** • The CCS stack is a complete platform for cluster deployment. All stack components have been tested for inter-operability and a single version number applies to the entire stack.
Developer	**Comprehensive and familiar development environment** • The integrated Visual Studio IDE provides a complete development environment • VS 2005 includes a parallel debugger • Intel compilers and dev tools for HPC being ported to Windows

Assumption: Using Windows on the desktop

Microsoft Windows Compute Cluster Server 2003

Customer Audiences

Departments:

Business – Engineering, R&D

IT – Corporate IT, Engineering IT

Titles/Roles:

Business Users

- Researchers, Scientists, Statisticians
- Engineers - Mechanical, Civil, Chemical
- Biologists, Bio-Informaticians
- Geophysicists, Reservoir Engineers
- Traders, Brokers, ‘Quants’

IT Professionals in Corp IT and ‘Engineering/Research IT Application developers

If You Could Only Ask 3 Questions …

- Does the customer use any applications that run computationally intensive calculations or simulations?
  - E.g: portfolio modeling, oil reservoir simulation, fluid dynamics modeling, or genome modeling)?
- What are the greatest challenges regarding these ISV and/or custom applications?
- What's the impact to the organization of those challenges?
  - Product Quality?
  - Time to Market?
  - Risk?
  - Financial?

Microsoft Windows Compute Cluster Server 2003

Key Sales Takeaways

- We don’t want to focus on the top 500, we want the bottom 500,000.
- Sell to accounts with AD infrastructure and Windows staff
- Sell where we have partner relationships
  - Sell into the focused verticals
  - Sell solutions using our managed ISVs
  - Leverage our partner OEMs (Dell, HP, IBM)
- Engage your Incubation Sales SSP early – mailto:hpcsales@microsoft.com

Microsoft Windows Compute Cluster Server 2003

Qualification Questions – Process

- **Describe the "end-to-end" end user process/experience and primary challenges related to using HPC applications:**
  - Gather/aggregate data from various sources
  - Pre-process data
  - Submit computational jobs
  - Validating job status
  - Retrieve processed data
  - Post-process data
  - Manipulate/query data to generate "insight"
- **Describe the high level developer process/experience and primary challenges related to bringing a new (ISV or custom) HPC application on-line:**
  - Develop/debug
  - Test
  - Deploy
  - Update
- **Describe the high level IT administrator process/experience and primary challenges related to bringing a new HPC application on-line:**
  - Evaluate solution alternatives and purchase hardware and (platform) software
  - Configure desktop and/or server hardware
  - Configure infrastructure integration (to data sources, between servers, security, etc.)
  - Patch/maintain desktop and/or server hardware (on-going maintenance)

Qualification Questions – Business

What are the customer's top 3 challenges and/or project success factors?

Addressed in early process questions:

- Ease of use
- Ease of development of (ISV and/or custom) HPC applications
- Ease of administration

Other possibilities:

- Data accuracy
- Performance
- System/data security
- System reliability/stability (9's of uptime)
- Ability to leverage existing Microsoft/IT infrastructure and resources (such as AD, Visual Studio, Excel)
- Ability to leverage industry standards
- End of life hardware/operating system or hardware lease expiration

Microsoft Windows Compute Cluster Server 2003

Qualification Questions – IT (Infrastructure)

- Applications
  - Name and primary functions of the ISV and/or custom applications (w/ versions)
  - Programming language (such as Fortran or C)
  - Job scheduler
  - Message Passing Interface (MPI)
  - 3rd party libraries (such as FlexLM)
  - Management tools (such as HP OpenView, CA Unicenter, IBM Tivoli)
  - BI tools (such as SAS, Cognos, Business Objects, Hyperion)
- Desktop Productivity
  - Are they currently using Excel as a desktop productivity tool, esp. to parse data before, during, or after calculations?
  - Are they currently using a backend database, esp. to aggregate, store, and query pre- and/or post processing data?
- Directory Services/Security
  - Are they currently using AD for directory services?
- Development Tools
  - Are they currently leveraging .NET based technologies such as Visual Studio

Qualification Questions – IT (Other)

- Why are they looking to deploy compute clusters vs. a dedicated workstation or server?
- Does the customer have existing benchmarks (external or internal) that need to be met?
- What size data sets are being computed?
  - Terabytes
  - Petabytes
- # of Users
  - End-Users (using ISV/customer applications; submitting compute jobs)
  - System Administrators
  - Developers
- SIs
  - Does the ISV provide infrastructure consulting and implementation?
  - Does the customer typically work with a 3rd party preferred SI?
- Are they performing calculations/simulations using in-house or outsourced resources?
- If outsourced, what were your reasons for doing so?
  - High up front capital costs
  - Low utilization

Profile of a “Good” Target Account

- Organization is in one of our target industries
  - Automotive, Aerospace, Oil and Gas, Life Sciences, Financial Services, Public Sector
- Heavy reliance upon Microsoft technologies
  - Active Directory, MOM, Visual Studio, Excel, etc
- Significant internal Microsoft expertise/support
  - Developers, System Administrators, etc
- Running applications that are on our “top 20” target ISV list
  - Note that these apps are used outside our target verticals as well
- Running applications on either …
  - a single workstation/server and experiencing performance issues
  - or on a compute cluster with associated end-user or IT pain
- Currently purchasing 64 bit server hardware platforms
- Early adopter profile

Profile of a “Less than Optimal” Target Account

- Running applications that are not on our Tier 1 list
  - Applications from smaller ISVs; custom code
- Running applications on a large (1000+node)/centralized Linux/UNIX cluster
- Minimal reliance upon Microsoft technologies
  - high level of investment/commitment to Linux; Linux as a religion
  - minimal internal Microsoft expertise
- Not interested in purchasing 64-bit server hardware platforms
- Laggard profile (rigid)

Sales Strategy

- Large cluster installations likely to already include scheduling and cluster management infrastructure
  - Aim to sell CCE (CD 1) as platform OS
- Identify other, smaller scale HPC opportunities in the account
  - Are they interested in deploying desk-side clusters?

Microsoft Windows Compute Cluster Server 2003

Objection Handling

Objection	Questions To Ask	Things to Highlight
Why would I want to pay for a Microsoft solution when I can get Linux for free?	• How much are you currently paying per node for software support for your Linux cluster?	For a stack with equivalent functionality and support, CCS will price very competitively, especially over the system lifetime since CCS is a 1-time payment not a subscription
CCS won't scale to large clusters. I won't have headroom for later	• How many nodes does your application support?	•CCS has been shown to scale to ~900 processors. Will publish a Top 500 number in June and add customer evidence as we enter the market
Windows is slow. CCS won't deliver the performance I can get on Linux.	• What's the 'wall-clock' time window within which you need to complete processing on your current model?	• CCS Benchmarks should be competitive with Linux. • Tier 1 ISV app benchmarks to be published closer to app GA dates • Customer feedback has been that ease of deployment, manageability & TCO take priority over pure benchmarks. It's cheaper to add nodes to make up for minor performance shortfalls
My application vendor does not support Windows	• What's the application? What's the problem domain (FEA, CFD etc)	• Does one of our partner ISVs fall provide an equivalent solution? • If the ISV is not on our list, should engage to see if porting is of interest. Plans for programmatic support for porting in FY07 are in discussion

Objection Handling

Objection	Questions To Ask	Issues to Highlight
Can I trust a v1 product from Microsoft		• Customer evidence • Top 500
Are applications available for CCS?	•Which applications are you running?	•MSFT invested heavily to enter the market with 20 industry leading ISV apps. that will be supported within 3 months of the June RTM •Additional ISV apps. are being ported to Windows CCS on a regular basis
We've already invested heavily in Linux and Linux based applications? Your stuff isn't any faster. Why should I switch?		This is a broader Unix¥Linux migration debate. Get The Facts

Microsoft Windows Compute Cluster Server 2003

Next Steps…

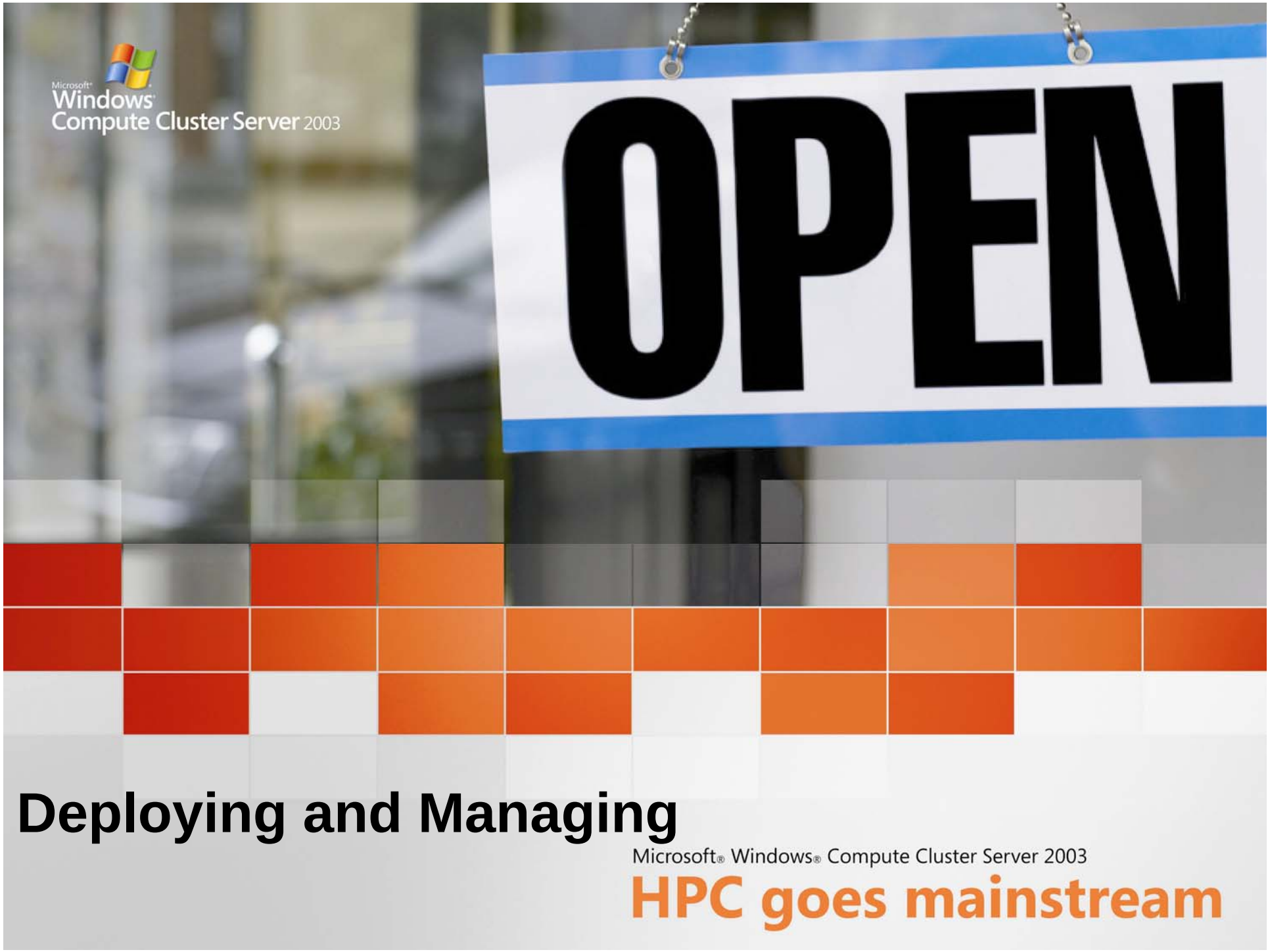
- ‘Hunting Guide’
  - 1-page cheat sheet with target-scenarios and qualification questions by vertical
  - Currently under construction
  - Input and other Feedback on content very welcome
- Call to action
  - Learn the product (http://www.microsoft.com/hpc)
  - Start talking to your customers about HPC

Microsoft Windows Compute Cluster Server 2003

For More Information

- Microsoft HPC Web Site:
  - http://www.microsoft.com/windowsserver2003/ccs/default.mspx
- Microsoft HPC FAQ:
  - http://www.microsoft.com/windowsserver2003/ccs/faq.mspx
- Microsoft Windows Compute Cluster Server 2003 Community Site
  - http://www.windowshpc.net
- Windows Server x64 information
  - http://www.microsoft.com/x64/
- Get the Facts Web site
  - http://www.microsoft.com/getthefacts

Microsoft Windows Compute Cluster Server 2003



Deploying and Managing

Microsoft® Windows® Compute Cluster Server 2003

HPC goes mainstream

Agenda

- Compute cluster overview
- Compute cluster deployment
  - Understand software and hardware requirements
  - Active Directory considerations
  - Head node setup
  - Configure cluster topology
  - Enable/disable FW on cluster public network
  - Install and Configure RIS (optional)
  - Add compute nodes (manual or automated)
  - ADS vs. RIS
- Compute cluster administration
  - UI overview
  - CLI overview
  - Add cluster admins and users
  - Run command on multiple nodes
  - Launch remote desktop connection
  - Open system monitor
  - Open event viewer
  - Script Center for Compute Cluster scripting

Microsoft Windows Compute Cluster Server 2003

What is Compute Cluster Server 2003?

Microsoft Windows Compute Cluster Server 2003

Windows Compute Cluster Server 2003

Mission: Deliver the easiest to deploy and most cost effective solution for solving scaled-out business, engineering and scientific computational problems.

Windows Server 2003, Compute Cluster Edition

+ Compute Cluster Pack =

Microsoft Windows Compute Cluster Server 2003

- Support for high performance hardware (x64bit architecture)
- RDMA support for high performance interconnects (Gigabit Ethernet, Infiniband, Myrinet, and others)

- Support for Industry Standards MPI2
- Integrated Job Scheduler
- Cluster Resource Management Tools

- Integrated Solution out-of-the-box
- Leverages investment in Windows administration and tools
- Makes cluster operation easy and secure as a single system

Microsoft Windows Compute Cluster Server 2003

CCS V1 Key Features

Node Deployment and Administration

- Simple deployment and administration experience
- Task-based configuration for head and compute nodes
- UI and command line-based node/job management
- Monitoring with PerfMon, MOM & 3rd party tools

Integration with existing Windows and management infrastructure

- Integrates with Active Directory, Windows security technologies, management, and deployment tools

Extensible job scheduler

- 3rd party extensibility at job submission and/or job assignment
- Submit jobs from command line, UI, or directly from applications
- Simple job management, similar to print queue management

Secure MPI

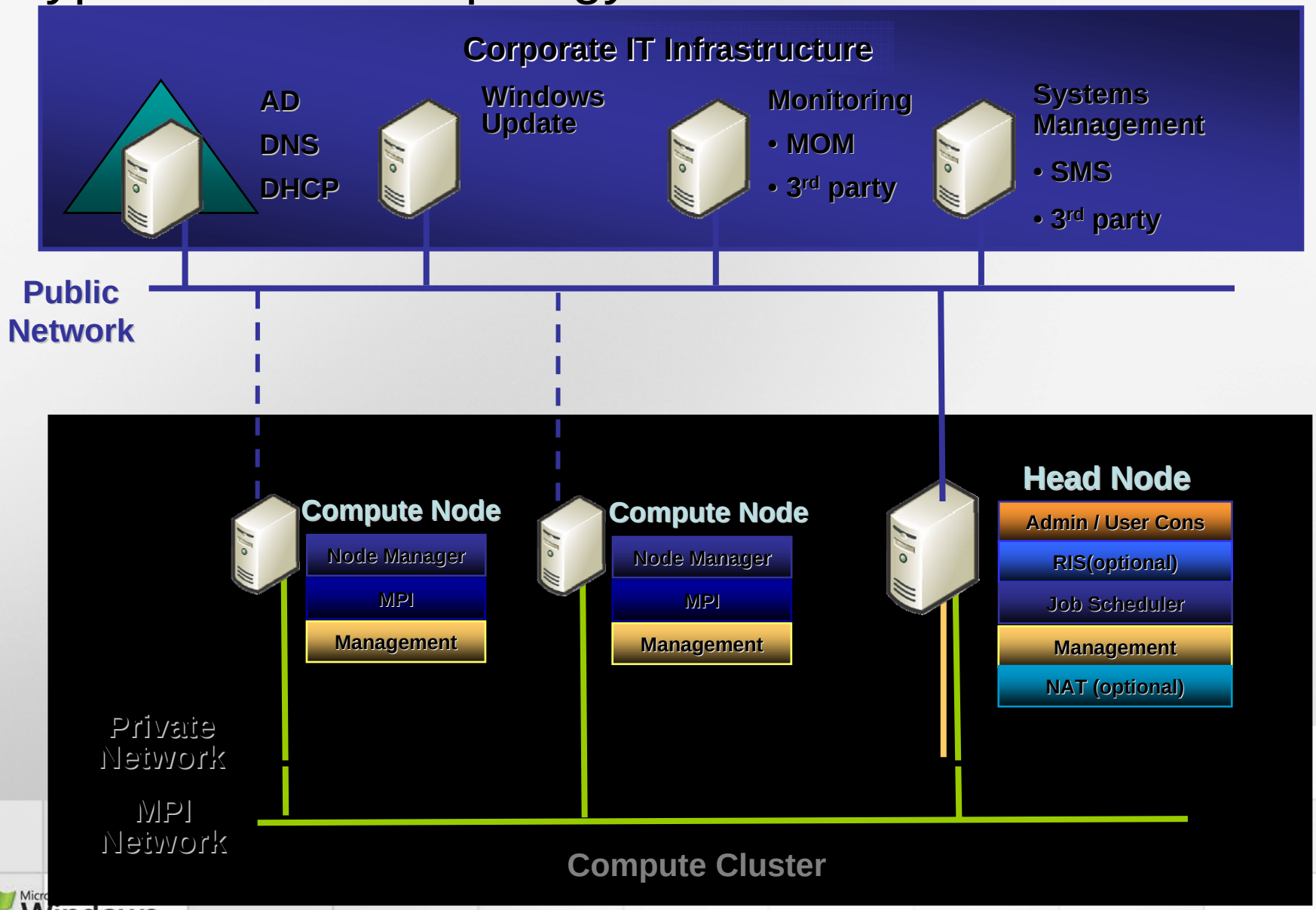
- User credentials secured in job scheduler and compute nodes
- Standardized MPI stack
- Microsoft provided stack reduces application/MPI incompatibility issues

Integrated Development Environment

- Parallel Debugger and OpenMP Support in Visual Studio

Microsoft Windows Compute Cluster Server 2003

Typical Cluster Topology



Windows Compute Cluster Server 2003

Leveraging Existing Windows Infrastructure

Active Directory

- Integration with IT Infrastructure
- Resource Management
- Group Policies

Windows Security

- Kerberos Authentication
- Secure Job Execution
- Secure MPI

Microsoft Enterprise Management Tools

- Operations Manager
- Windows Update Services
- Systems Management Server

Compute Cluster Built-in Tools

- Job Scheduler
- Admin Console
- Performance Monitor
- Command Line Interface
- Remote Installation Services

Microsoft Windows Compute Cluster Server 2003

Compute Cluster Deployment

Deployment Checklist

Step	Action
1.	Review concepts for CCS, including system hardware and software requirements.
2	Determine topology for your compute cluster.
3.	Create your head node. Install supported operating system. Create a second disk volume (D:¥) to store installation images if you planning to use RIS on HN
4.	Join head node to an Active Directory domain (existing) or create a new domain.
5.	Install MMC 3.0 and required Hot fixes (RIS and ICS) on head node
6.	Install Compute Cluster Pack on the head node.
7.	Configure network topology on the head node.
8	Enable or disable the Windows Firewall on public network interfaces of the cluster.
9.	Enable RIS (optional) from the Compute Cluster Pack if you intend to use the Automated method to add compute nodes to your cluster. Then add one or more installation images and configure the images with valid product keys and identifying descriptions.
10.	Add compute nodes to your cluster, using either the Automated or Manual Addition method.
11.	Approve nodes.
12.	Define cluster users and administrators.
13.	Download the Microsoft Compute Cluster Pack SDK, if desired.

Hardware Requirements

Hardware requirements for Windows Compute Cluster Server 2003 are the same as those required by Windows Server 2003, Standard x64 Edition.

Processor	x64 architecture computer with Intel Pentium or Xeon family processors with Intel Extended Memory 64 Technology (EM64T) processor architecture; AMD Opteron family processors; AMD Athlon family processors; compatible processor's).
RAM	512 MB
Multiprocessor Support	Windows Compute Cluster Server 2003 and Windows Server 2003 Standard x64 Edition support up to four processors per server. Windows Server 2003 Enterprise x64 Edition supports up to eight processors per server.
BIOS	Ensure that the server BIOS is the latest offered by the hardware manufacturer.
Minimum Disk Space for Setup	4 GB
Disk Volumes	**Head Nodes:** Configure two volumes (C:¥ and D:¥) on the head node if you plan to add nodes using the Automated Addition method. One volume will be the system partition, and another partition will be used to store Remote Installation Services (RIS) images to support the Automated Addition method. The partition that stores RIS images should be a minimum of 2 GB. It must be formatted NTFS. You should install sufficient disk space on the head node to store any input and output files associated with applications and scripts that will be used by cluster users. RAID 0,1,5 may be used as appropriate, although not required. Compute Nodes: A single disk volume is usually sufficient.
Network Adaptor	All nodes require at least one network adaptor. If a private network is planned, and depending on the network topology selected, the head node requires a minimum of two network adaptors to create a public and private network. Each node may require additional adaptors as appropriate in case of public network access or in support of an MPI network.
PXE Boot Support (BIOS and network adaptor)	If you plan to use the Automated Addition method of adding compute nodes, servers that you plan to add as compute nodes must support PXE in the boot sequence. This is usually configured in the BIOS. In addition, verify that the network adaptor of each compute node that is used on the private network supports PXE boot.

Microsoft Windows Compute Cluster Server 2003

Software Requirements

- **Head Node and Compute Node Operating system**
  - Microsoft® Windows Server™ 2003, Compute Cluster Edition
  - Microsoft® Windows Server™ 2003, Standard x64 Edition
  - Microsoft® Windows Server™ 2003, Enterprise x64 Edition
  - Microsoft® Windows Server™ 2003 R2, Standard x64 Edition
  - Microsoft® Windows Server™ 2003 R2, Enterprise x64 Edition
- **Active Directory**
- **Remote Installation Services (optional)**
- **Required hot fixes and prerequisites on HN**
  - ICS hot fix (http://go.microsoft.com/fwlink/?linkid=55166)
  - RIS hot fix (http://go.microsoft.com/fwlink/?linkid=55167)
  - MMC 3.0 (http://go.microsoft.com/fwlink/?linkid=62400)
- **MSSQL Desktop Edition**, installed as part of the Compute Cluster Server 2003 head node installation process.
- **.NET Framework 2.0**, installed as part of the Compute Cluster Server 2003 head node and compute node installation process.

Microsoft Windows Compute Cluster Server 2003

Active Directory Considerations

CCS v1 requires Active Directory

Scenarios we support:

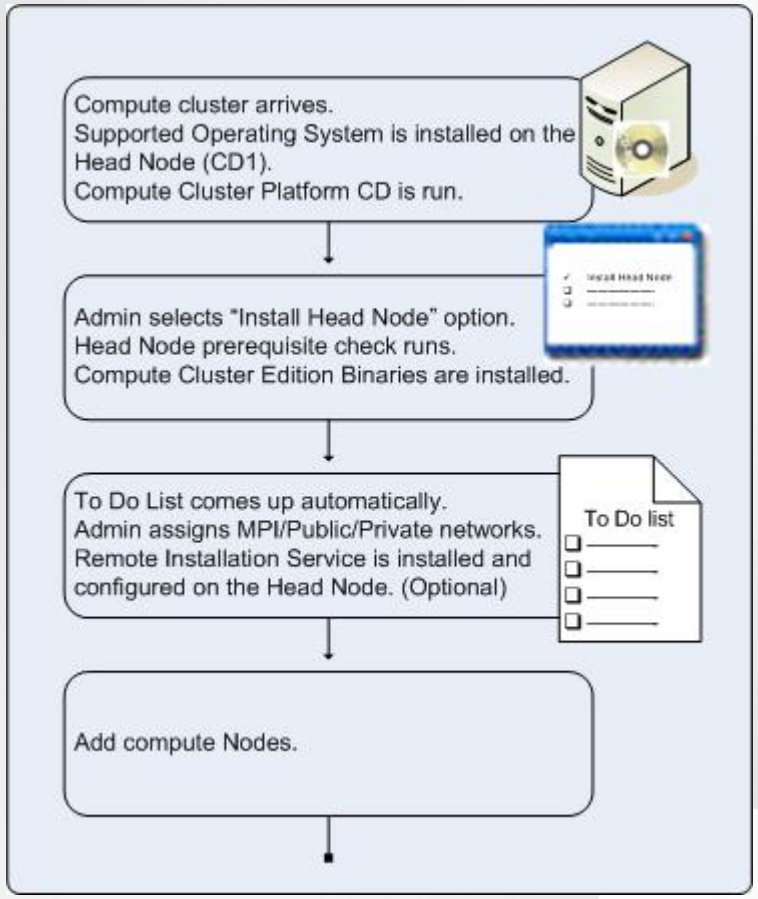
- **All compute nodes and head node are member servers in an existing AD domain**. "Recommended"
- **HN running as DC in a new domain in a new forest**

  We just wanted leave some options for the customers who do not have AD in place. But…
  - Within the forest, trust relationships between the domains are automatically created if you have multiple compute clusters, each in their own domains and forests. Managing trusts and permissions can be quite a hassle (going back to NT 4.0 days).
  - AD Redundancy. Once you have a DC then you need to think about redundancy if your DC goes down. Having a second DC, distributing AD FSMO roles across these DCs, and perhaps having another DNS server which stores the copy of AD integrated DNS zone on second DC are all ideas that you need to think about to support AD redundancy. As you can imagine, these are not really very simple tasks for a scientist and research engineer, which, by the way, is already covered in various AD documentations.

Important: Please have HN and CNs as member servers in existing AD domain unless you have absolutely no other option.

Head Node Setup

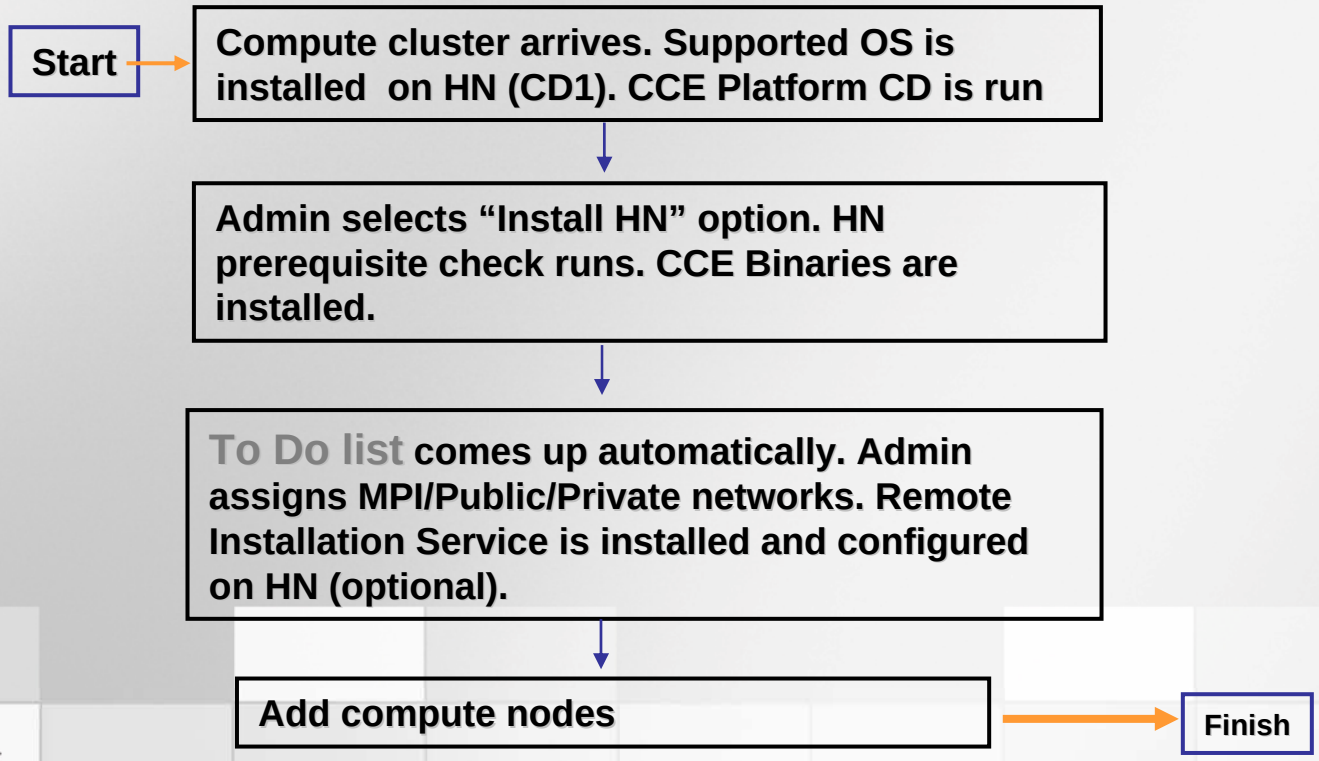
- HN must already have Windows Server 2003, Compute Cluster Edition or a supported OS installed
- HN must be connected to at least one network, depending on your choice of cluster networking topology, and you must be a local administrator on that computer
- HN must be a member of an Active Directory domain (recommended) or a domain controller for the cluster domain (not recommended)



Microsoft Windows Compute Cluster Server 2003

Simplified Head Node Installation

- Head Node installs only on x64 in first release
  - Windows 2003 Compute Cluster Edition
  - Windows 2003 SP1 Standard & Enterprise
  - Windows 2003 R2

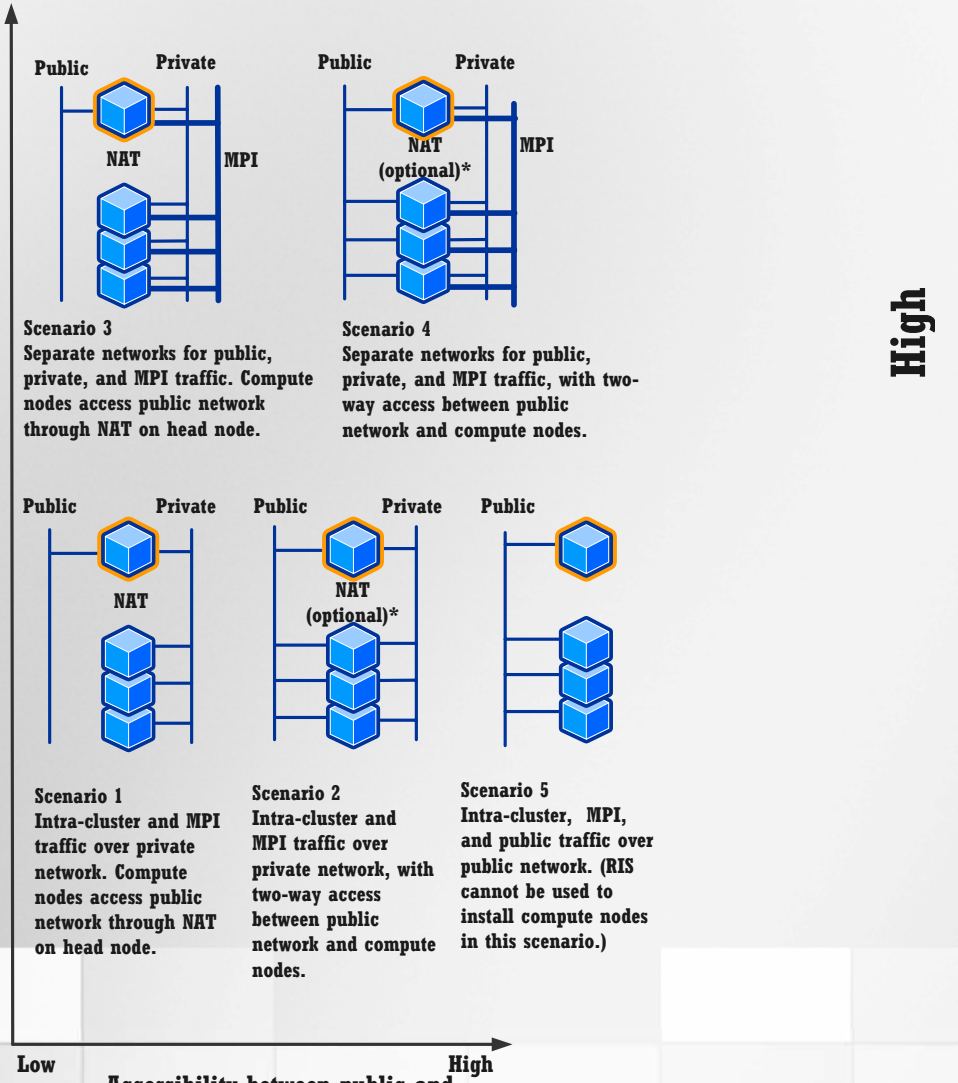


Microsoft Windows Compute Cluster Server 2003

Configure Cluster Topology

Public network	An organizational network connected to the head node and optionally, the cluster compute nodes. The public network is often the business or organizational network most users log onto to perform their work. All intra-cluster management and deployment traffic is carried on the public network unless a private network (and optionally, an MPI network) also connect the cluster nodes.
Private network	A dedicated network that carries intra-cluster communication between nodes. This network, if it exists, carries management, deployment, and MPI traffic if no MPI network exists.
MPI network	A dedicated network, preferably high bandwidth and low latency, that carries parallel MPI application communication between cluster nodes. This network, if it exists, is usually the highest bandwidth network of the three listed here. If the jobs you intend to submit to the cluster do not use MPI libraries, no MPI traffic will be generated and an MPI network is not needed.

Configure Cluster Topology



Microsoft Windows Compute Cluster Server 2003

Enable/Disable FW on Public Network

- **Enable Windows Firewall**
  Enable Windows Firewall on the public interface of the head node and on any public network interfaces that exist on the compute nodes. If your cluster has private or MPI networks, the firewall is disabled by default on those interfaces
- **Disable Windows Firewall**
  Disable Windows Firewall on all head node network interfaces

Install and Configure RIS

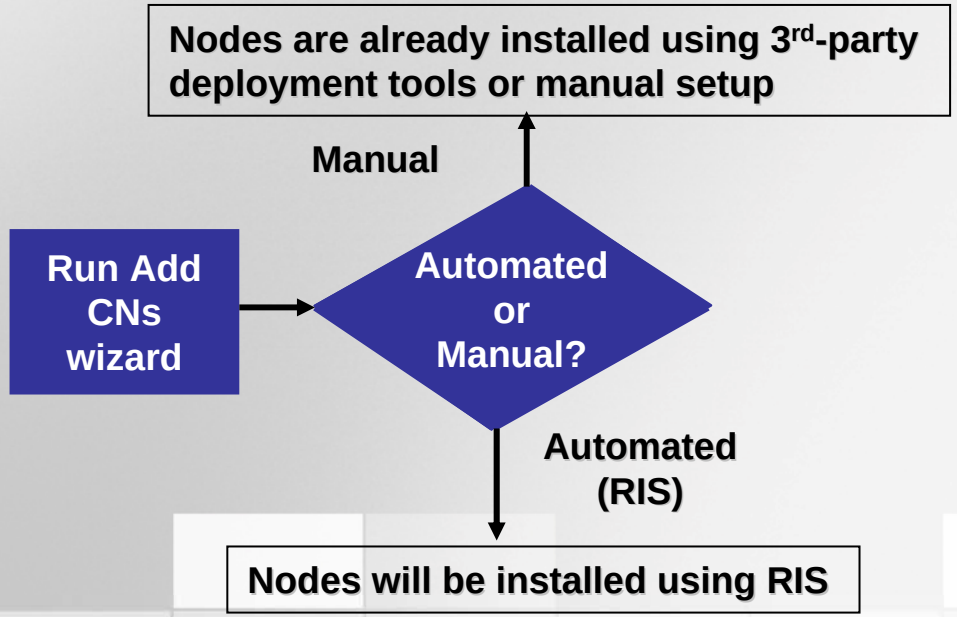
- **Install RIS, prerequisites are :**
  - A private network connects the cluster nodes
  - The prerequisite RIS hot fix is installed on the head node
  - Enable ICS on the HN during setup or configure a DHCP server on the private network
- **Add Image**
- **Assign a license key to image**
- **Watch out: TX mode and plug and play drivers**
  - Many new 64-bit systems use disk drives such as SATA drives or network cards which are not included in the Windows Server 2003 Standard x64 Edition setup
  - If the RIS image that you apply to a computer lacks the required drivers. You must add the drivers to the base image and modify unattended setup files to point to these new drivers. (Check modify RIS section in product documentation)

Add Compute Nodes

- **Automated addition:**
  - RIS must be installed and configured
  - You must have already created and modified your RIS installation images with valid software license keys
  - The cluster network topology must include a private network. If the cluster is installed only on a public network, you cannot use RIS
- **Compute Cluster pack installation method:**
  - Run Compute Cluster Pack Setup
  - Setup presents the option: Join this server to an existing compute cluster as a compute node
  - Specify a cluster Head Node (HN)
- **Manual addition:**
  - Can be used only after a supported OS and the CCP service are installed and running on the prospective node
  - The computer must also be added to the same Active Directory domain as the head node
  - You must be a local administrator on the computer and it must be attached to a network
  - You must know the computer name of any node that you want to add to your cluster.

Flexible Compute Node Installation

- Three ways to add compute nodes to cluster
  - Automated setup through RIS. “Just power on nodes”
  - 3rd party deployment tool support *setup.exe -computenode:<Headnodename>*
  - CD2 compute node manual setup option



Microsoft Windows Compute Cluster Server 2003

ADS vs. RIS

Remote Installation Services

- Built-in in W2K3 x64, requires CCS RIS QFE (KB 907639)
- Automated node deployment relies on RIS in CCS v1
- RIS uses unattended setup base installation (not single file traditional image files), so it is not fast but certainly good enough for small and medium size clusters
- TX mode and plug and play drivers must be added to RIS images
- Does not require full blown DHCP which requires authorization, works with mini DHCP provided with ICS

Automated Deployment Services 1.1

- ADS 1.1 services can not be installed on W2K3 x64
- Once you install ADS on x86 hw you can deploy x64 machines
- Requires full blown DHCP on the network
- Completely image base (Capture golden image and then restore)
- Provides better orchestration but requires ADS knowledge
- HPC team is planning to write a whitepaper on “Using ADS 1.1 to deploy CCS v1”

Compute Cluster Deployment

Microsoft® Windows® Compute Cluster Server 2003

HPC goes mainstream

Compute Cluster Administration

Microsoft Windows Compute Cluster Server 2003

Simple Deployment and Management

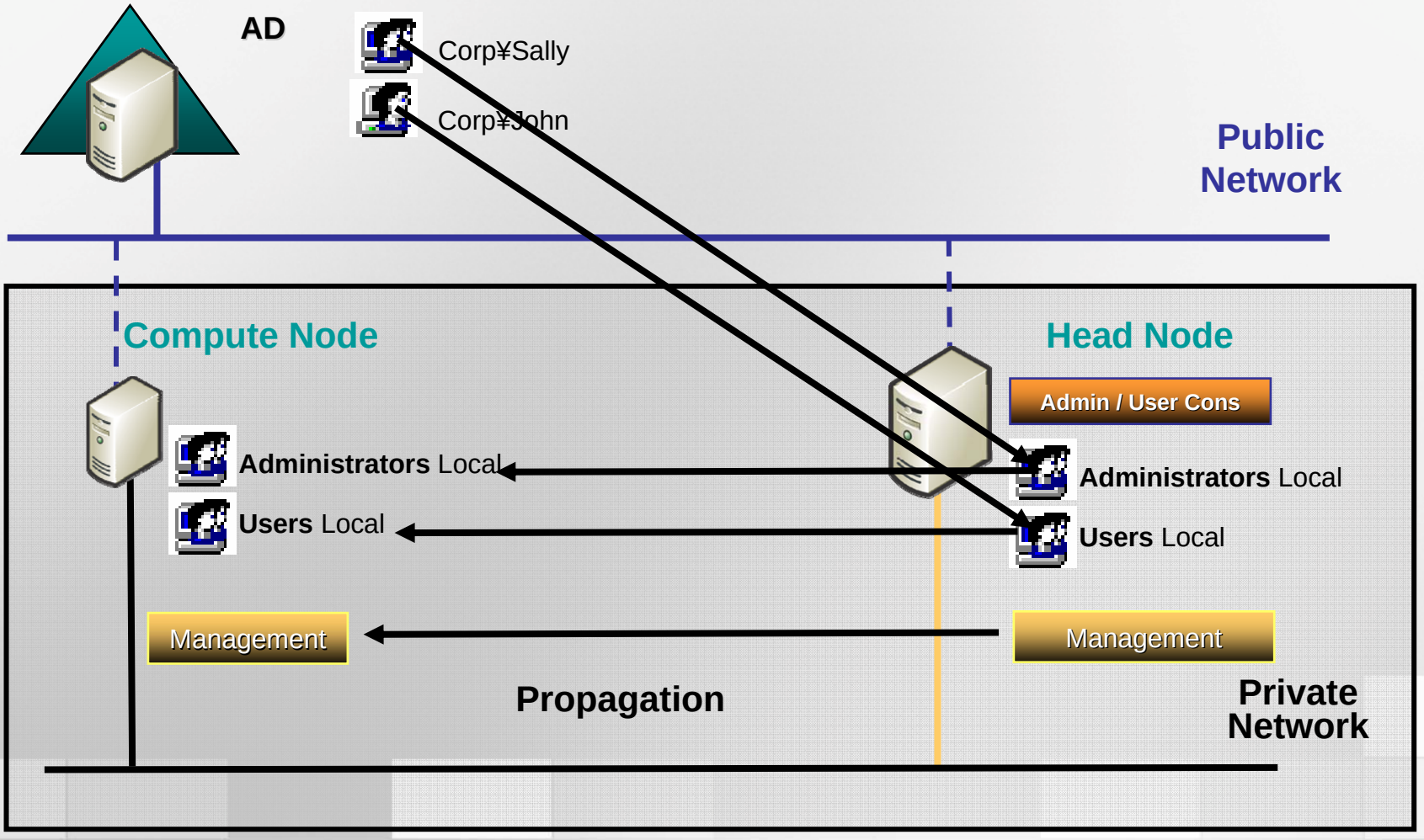
- Windows Compute Cluster Server 2003 speeds time to insight by simplifying cluster deployment
  - Prescriptive, simplified cluster setup and administration
  - Automated deployment and configuration
  - Integration with existing Windows and management infrastructure
    - Reduces setup time
    - Requires less specialized skill set

Microsoft Windows Compute Cluster Server 2003

Key Management Features

- Integration with existing infrastructure
  - Leverage Active Directory and Windows security
  - No special permissions required for installation of cluster
  - Integrate with Microsoft management technologies (MOM, SMS…etc)
- Model based management
  - Use System Definition Model (SDM)
- Simplified administration and node deployment
  - Task based configuration for head node and adding nodes
  - Multiple ways to add compute nodes
  - Node management through UI and command line
  - Monitoring with Perfmon, MOM, or 3rd-party management tools

Role-Based Cluster Management

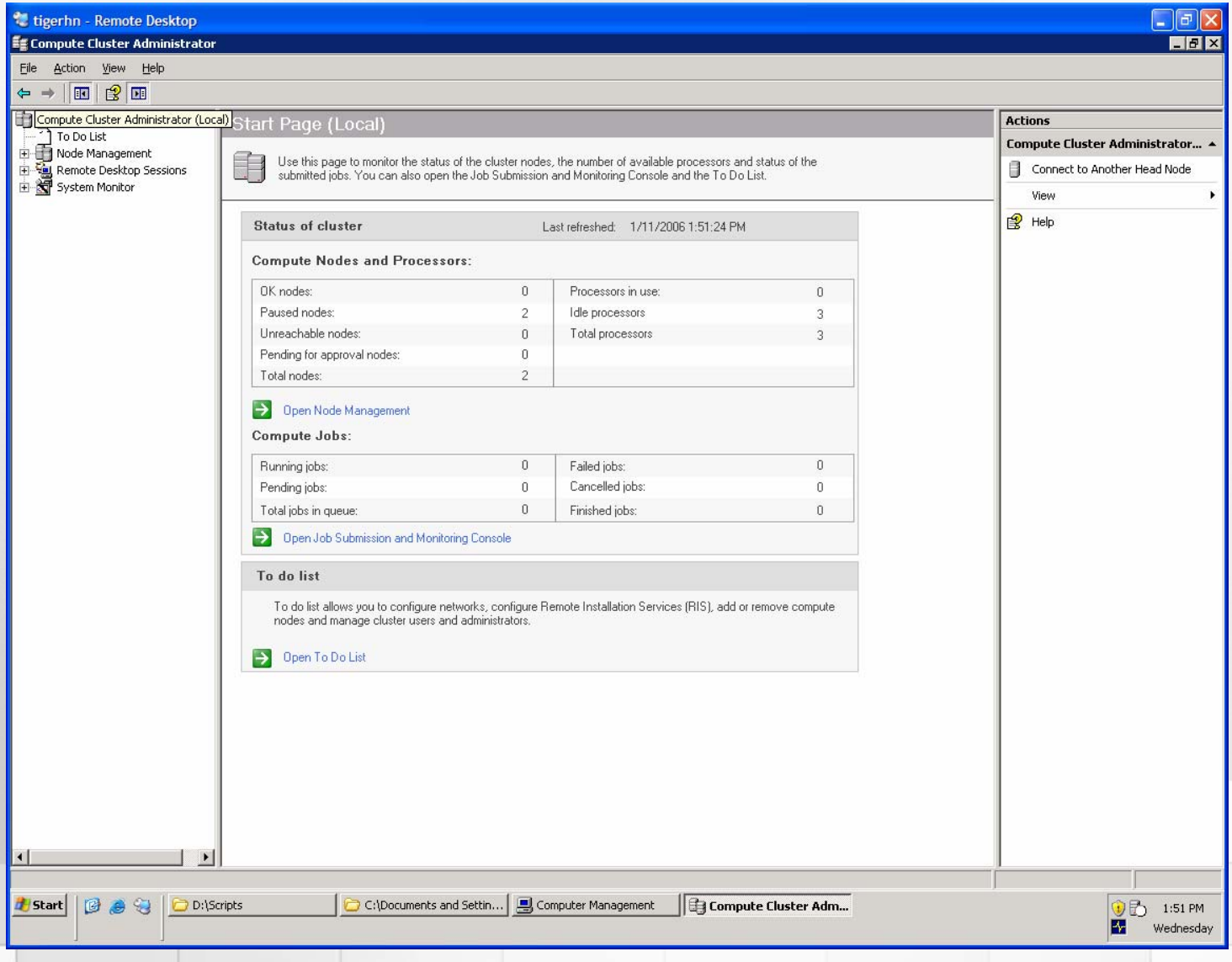


Microsoft Windows Compute Cluster Server 2003

New Admin Console

- Not building a new systems management paradigm
- Leveraging MMC for simple management (MMC 3.0)
- MMC used for cluster administration
- Has 5 branches
  - Start Page
  - To Do List
  - Node Management
  - Remote Desktop
  - Performance

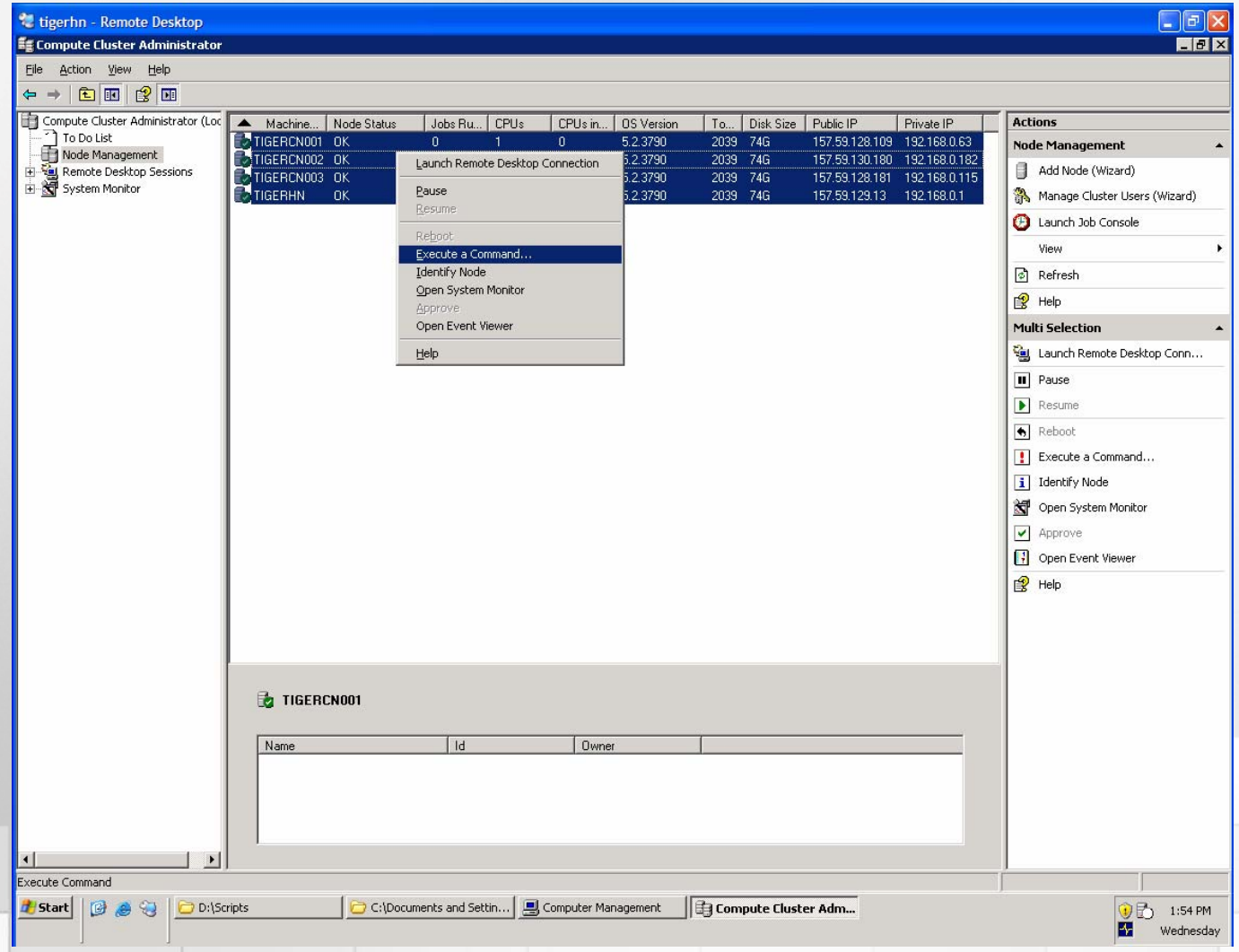
Admin Console (Start page)



Microsoft Windows Compute Cluster Server 2003

Admin Console (Node Management)

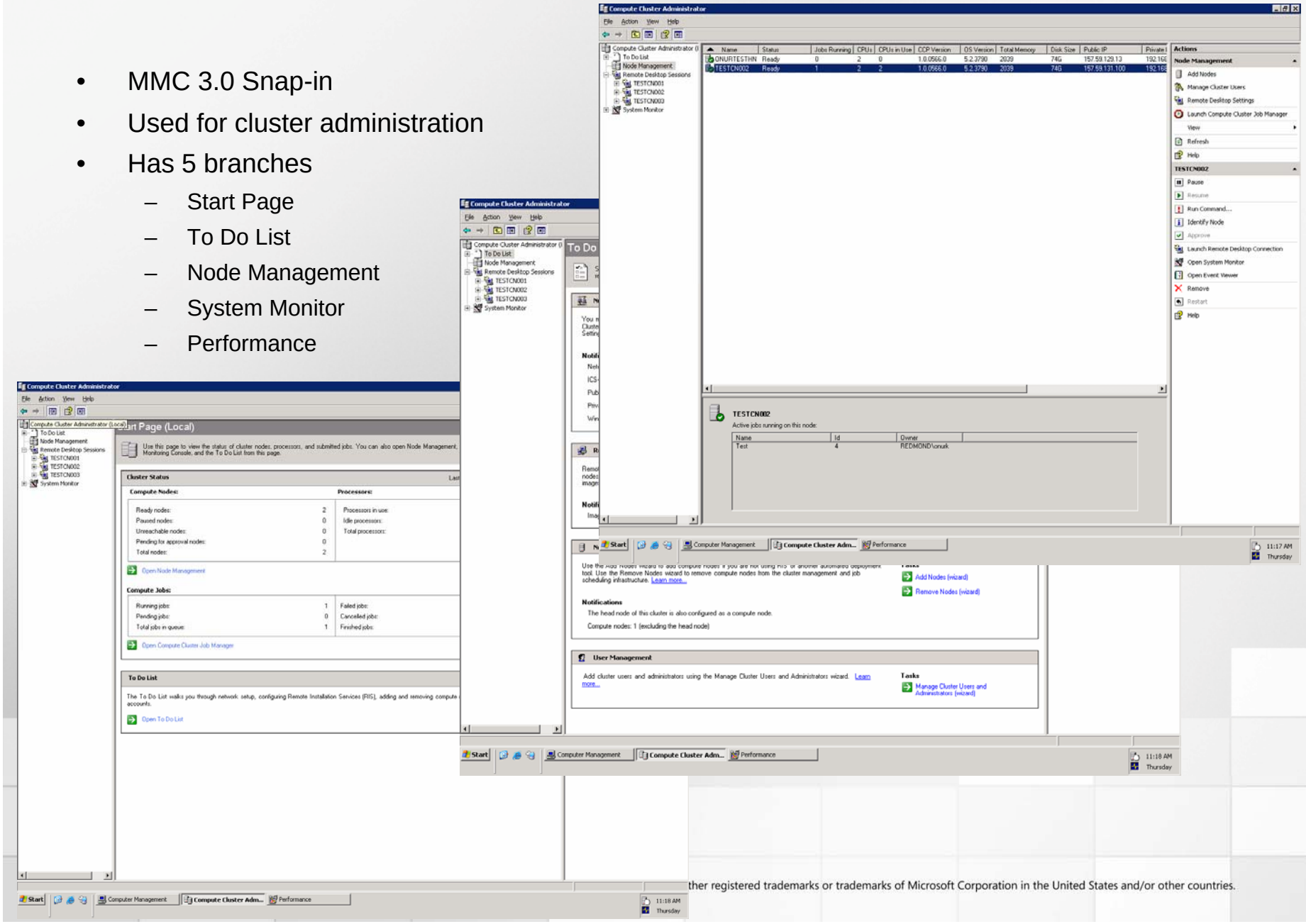
Supports specific actions



Microsoft Windows Compute Cluster Server 2003

Compute Cluster Administrator

- MMC 3.0 Snap-in
- Used for cluster administration
- Has 5 branches
  - Start Page
  - To Do List
  - Node Management
  - System Monitor
  - Performance



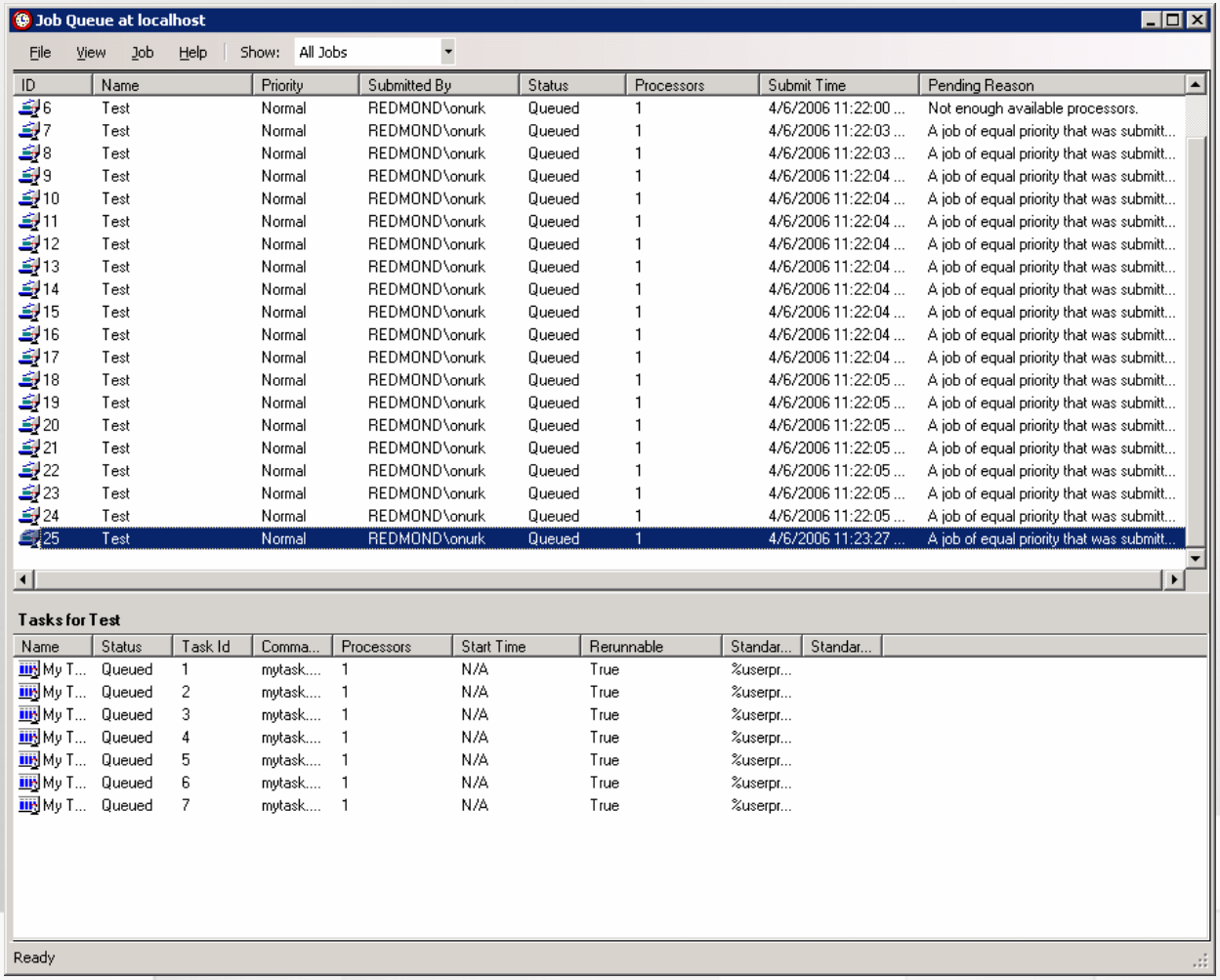
Compute Cluster Administration

Microsoft® Windows® Compute Cluster Server 2003

HPC goes mainstream

Compute Cluster Job Manager

- Win32 App
- Centrally manage the entire job queue
- Print queue manager like experience
- Ability to save and submit jobs using templates
- Automatically create “Parametric Sweep”



Microsoft Windows Compute Cluster Server 2003

Job Scheduler

Microsoft® Windows® Compute Cluster Server 2003

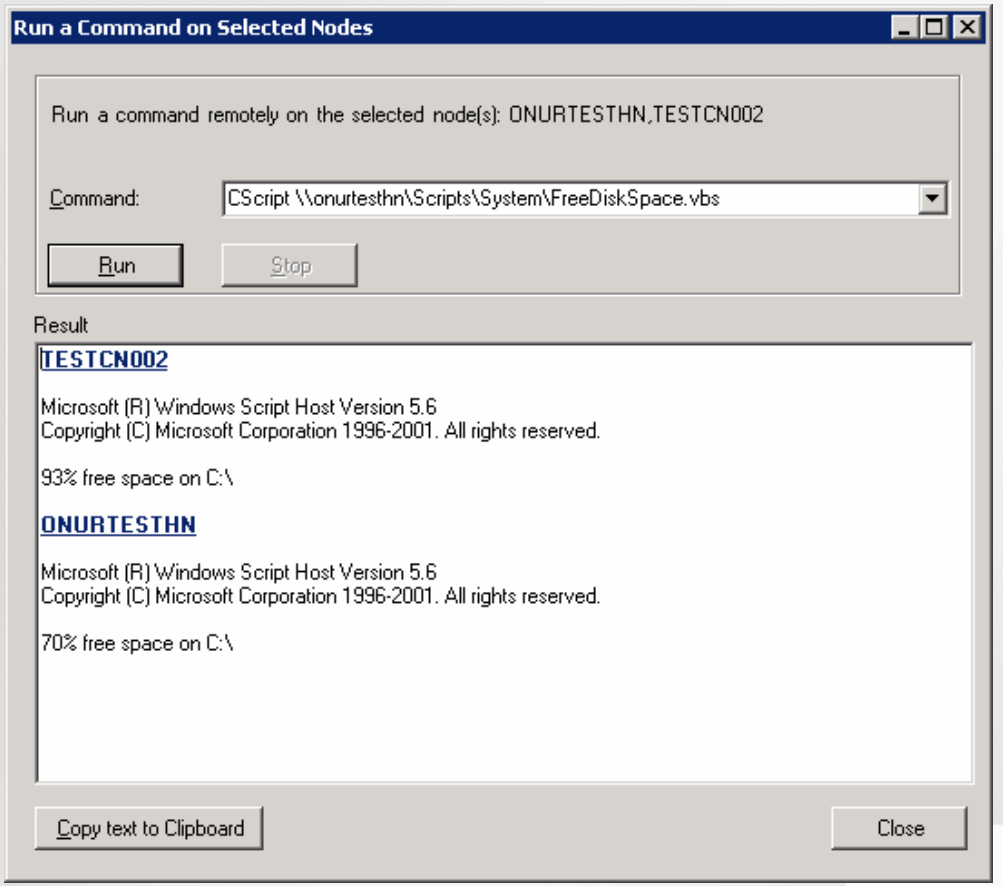
HPC goes mainstream

Add Administrators and Users

- **Cluster Administrator**
  - Cluster administrators have full access to all nodes on the cluster and automatically become members of the local Administrators group on each node in the cluster.
  - Cluster administrators can use the Compute Cluster Administrator to add and remove compute nodes and add or remove cluster users.
  - Administrators can use the Compute Cluster Job Manager or issue commands from the command-line interface (CLI) to submit jobs and manage the job queue.
- **Cluster User**
  - Cluster users can submit jobs by using the Compute Cluster Job Manager or the CLI. Users will be able to see the entire job queue and can only modify their own jobs.
  - Jobs will run using the user credentials of the person who submitted the job; as a result, the job can only access resources and services during job execution under the rights of that user.

Remote Command Execution

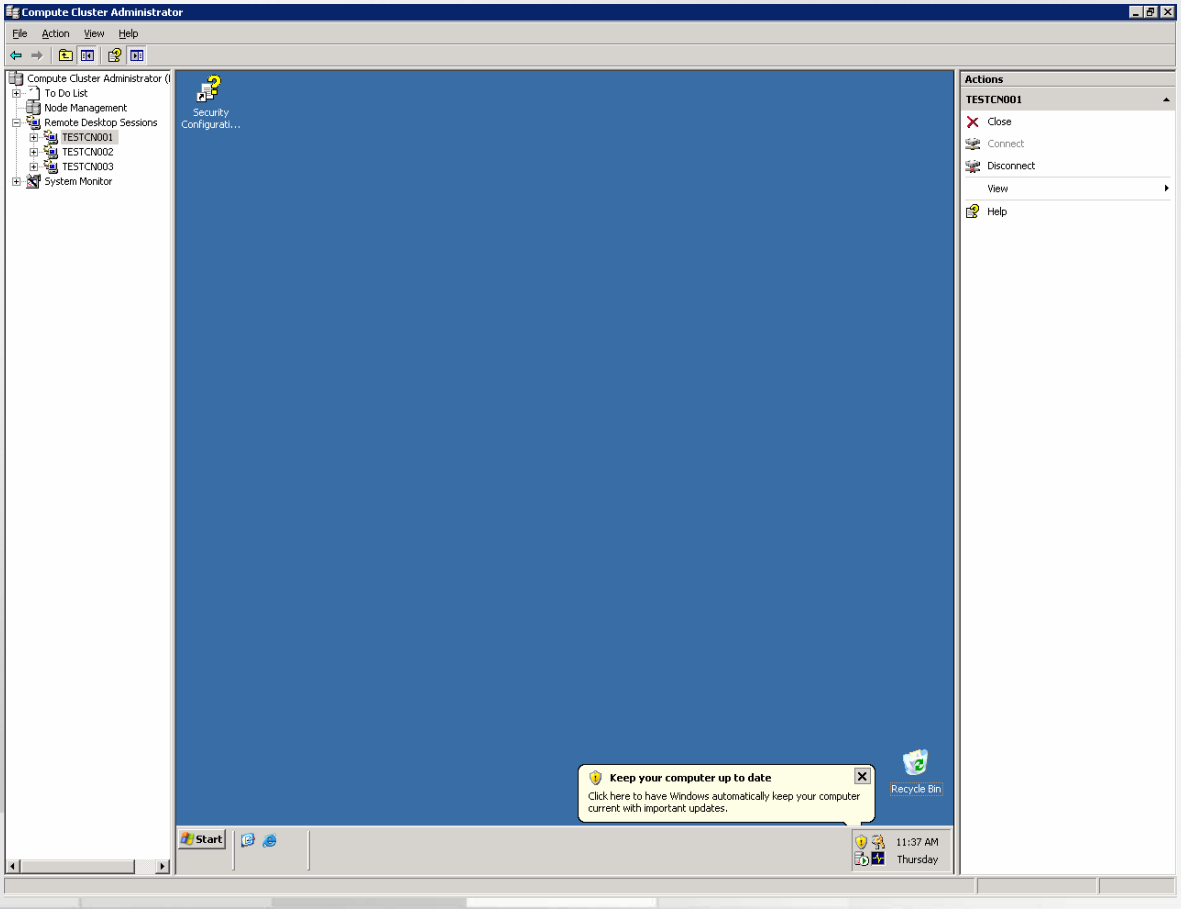
- **Admin can issue a command line action on one or more selected compute nodes.**
- **The command is immediately run through the Job Scheduler on a priority basis. The command can be any valid command line argument.**
- **The credentials of the user submitting the command line action is used to execute the action.**
- **Can be used to run scripts and programs on compute nodes.**

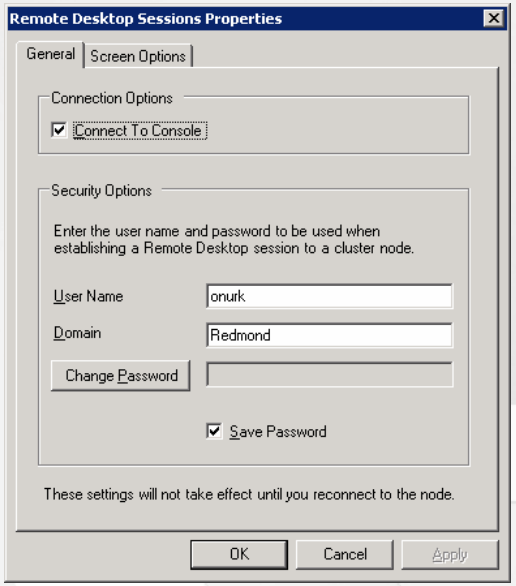


Microsoft Windows Compute Cluster Server 2003

Launch Remote Desktop Connection

Admins can create Remote Desktop sessions to multiple compute nodes, switching between each node to perform required operations.





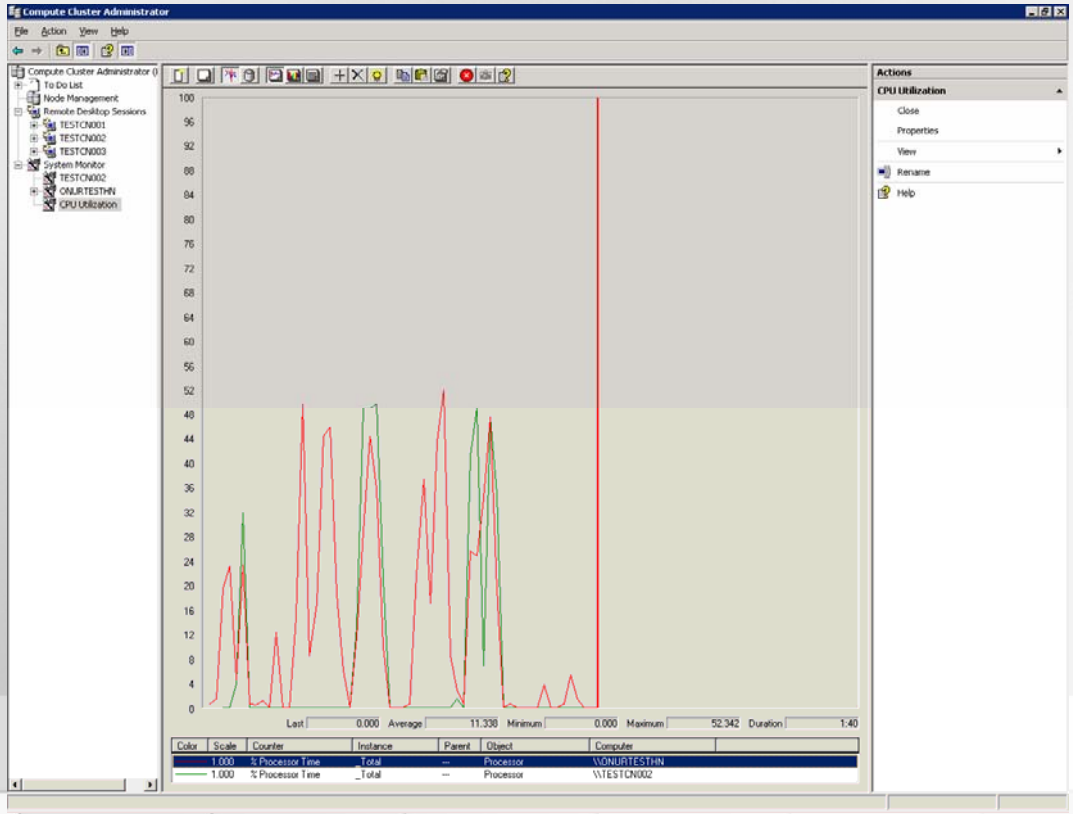
Microsoft Windows Compute Cluster Server 2003

Monitoring Compute Clusters

- Built-in capability
- Queue monitoring
  - # of running, pending, failed, canceled, finished jobs in queue
  - # of running, pending , failed, canceled, finished tasks in the queue
- Node monitoring
  - # of total nodes
  - # of idle, paused, unreachable and ok nodes
  - Percentage of idle, paused, unreachable and ok nodes
  - # of CPUs in cluster
  - # of CPUs in use in cluster
- Base OS performance counters (CPU, memory, logical/physical disk, network…etc)

Open System Monitor

- **CCS provides two categories of performance objects:**
  - **Compute Cluster: cluster-specific performance counters**
  - **Compute Nodes: node specific performance counters**
- **Each object exposes a rich set of performance counters to monitor different characteristics of the cluster and nodes.**



Microsoft Windows Compute Cluster Server 2003

Monitoring Using Microsoft Server Performance Advisor

- SPA: Lightweight monitoring tool (free)
- Use templates
  - Compute Cluster Compute Node (CN) template
  - Compute Cluster Head Node (HN) template
- SPA can be run on nodes using remote command execution
- Run Compute Cluster CNs template on CNs
- Run Compute Cluster HNs template on CNs
- Detail Report

Monitoring Using Microsoft Operations Manager 2005

- MOM 2005: Enterprise Monitoring Tool
- Performance and trend analysis
- Alerts and Notifications
  - % Processor Time-_Total>%40 on any nodes
  - Number of queued jobs>30
- Computer group health diagram

Microsoft Windows Compute Cluster Server 2003

Script Center for Compute Cluster Scripting

- Provides compute cluster scripting overview
- Script repository

http:/technet.microsoft.com/scriptcenter/hubs/ccs.mspx

Microsoft Windows Compute Cluster Server 2003

Summary

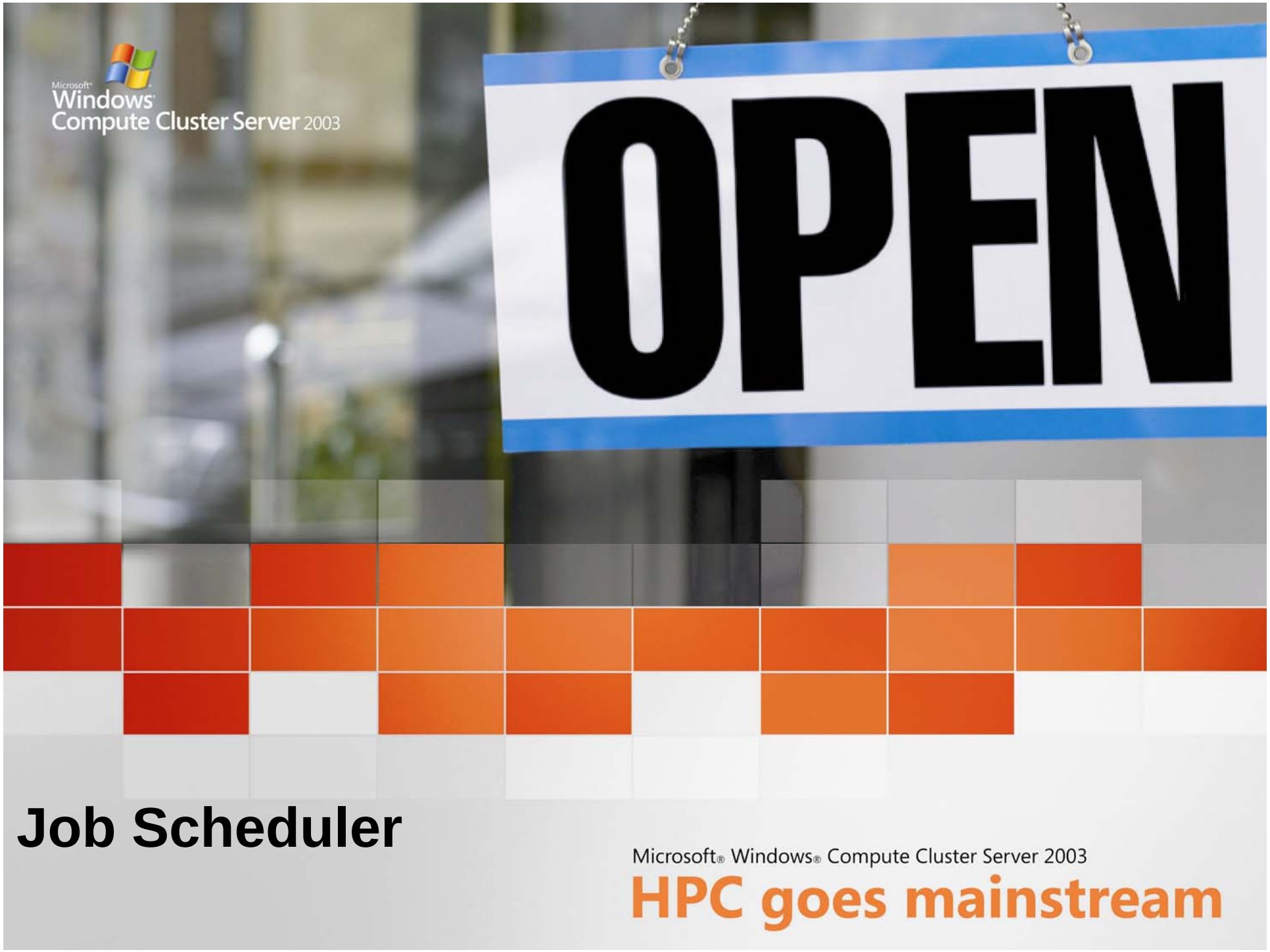
- Deploying and operating clusters is not complex
- Does not require the end-user to stitch together and manage multiple pieces of hardware and software.
- Windows Compute Cluster Server 2003 removes administrative barriers preventing broad adoption of HPC solutions
  - Familiar environment and integration with standard tools
  - Convenient job scheduler
  - Parallel debugging capabilities and full support of MPI standards
- Windows Compute Cluster Server 2003 makes HPC accessible to all scientists, engineers, and businesses

Microsoft Windows Compute Cluster Server 2003

External Resources

- Microsoft HPC Web site
  - http://www.microsoft.com/hpc/
- Windows Server x64 information
  - http://www.microsoft.com/64bit/, http://www.microsoft.com/x64/
- Windows Server System information
  - http://www.microsoft.com/wss/
- Get the Facts Web site
  - http://www.microsoft.com/getthefacts
- GotDotNet Gallery for HPC
  - http://www.gotdotnet.com/codegallery/codegallery.aspx?id=0c12237e-58e3-4a9f-9637-ec6a4b9efe4d
- Blog
  - http://blogs.msdn.com/hpc

Microsoft Windows Compute Cluster Server 2003



Job Scheduler

Microsoft® Windows® Compute Cluster Server 2003

HPC goes mainstream

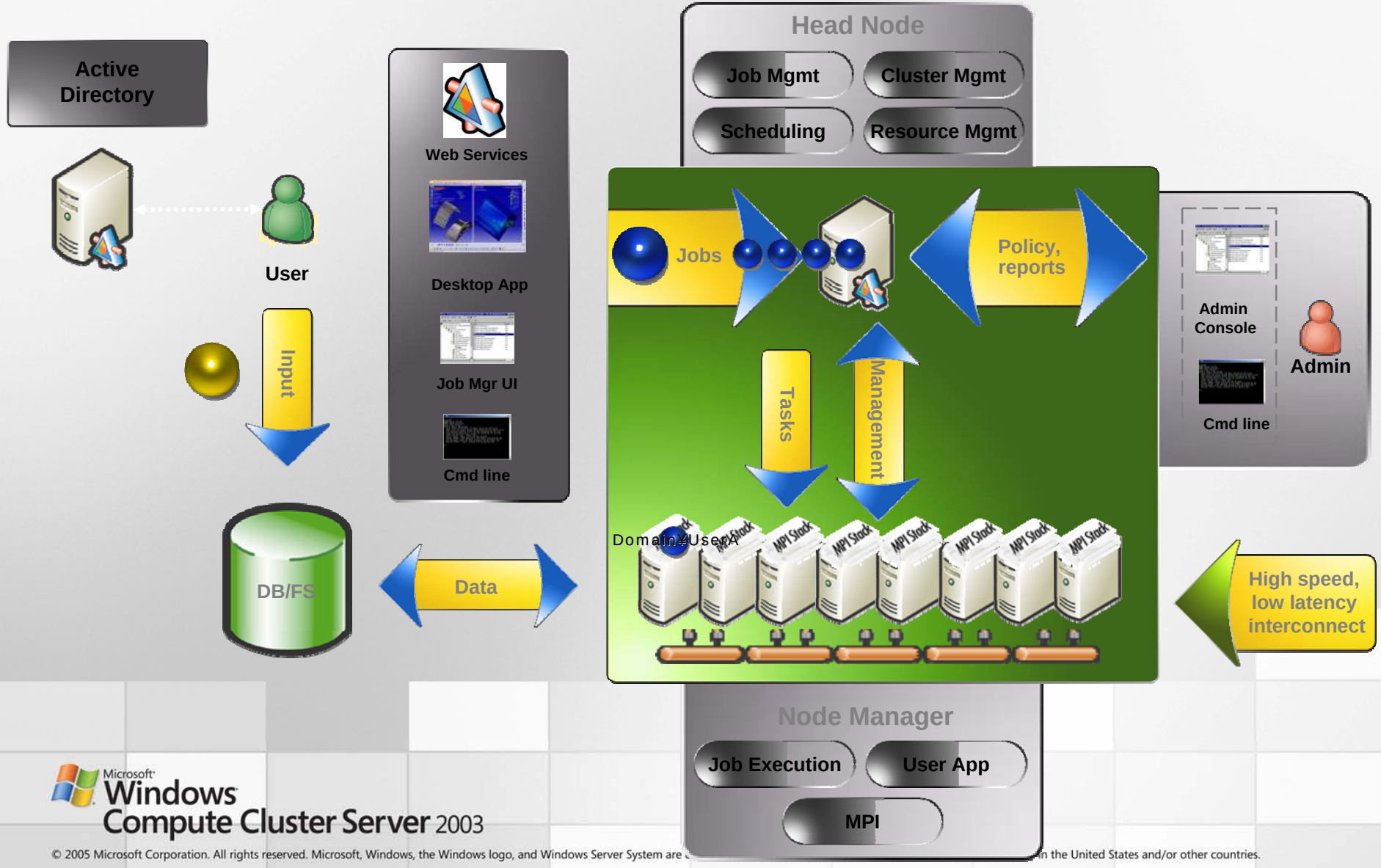
Windows Compute Cluster Server 2003

Mission: Deliver the easiest to deploy and most cost effective solution for solving scaled-out business, engineering and scientific computational problems.

Windows Server 2003, Compute Cluster Edition	+	Compute Cluster Pack	=	Microsoft Windows Compute Cluster Server 2003
• Support for high performance hardware (x64bit architecture) • RDMA support for high performance interconnects (Gigabit Ethernet, Infiniband, Myrinet, and others)		• Support for Industry Standards MPI2 • Integrated Job Scheduler • Cluster Resource Management Tools		• Integrated Solution out-of-the-box • Leverages investment in Windows administration and tools • Makes cluster operation easy and secure as a single system

Microsoft Windows Compute Cluster Server 2003

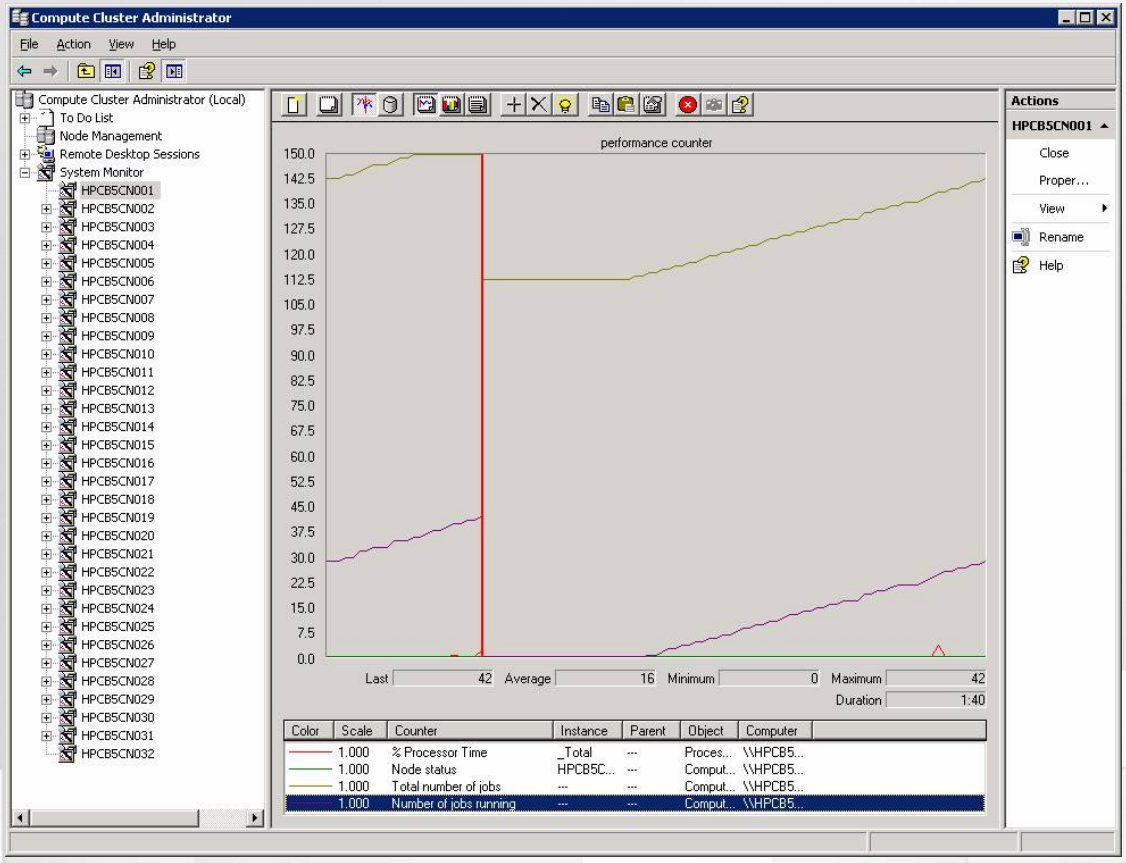
How CCS works



Microsoft Windows Compute Cluster Server 2003

Integrated Job Scheduling and Management

- Integrated queue monitoring
- Integrated node add/remove operations
- Integrated administrative command and job scheduling



Microsoft Windows Compute Cluster Server 2003

Integrated Parallel Programming

- Message Passing Interface (MPI)
  - Works in shared memory (single chassis) and cluster (multiple chassis) situations.
  - Requires moderate-advanced level of programming
- OpenMP
  - Works in shared memory situations only- inside one physical chassis
  - Easy to program- simply identify parallelizable loops with pragmas

Microsoft Windows Compute Cluster Server 2003

Partners

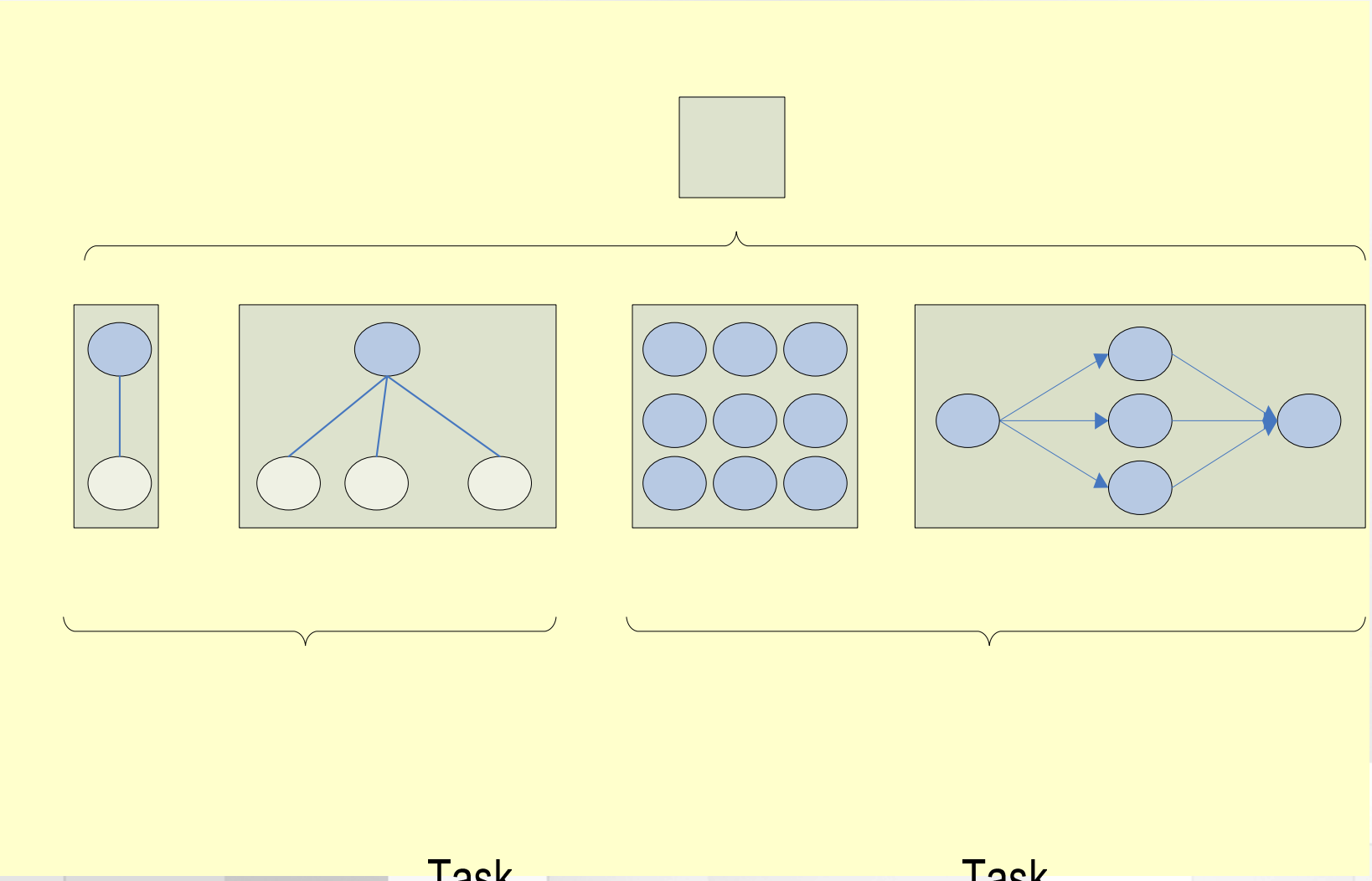
Microsoft Windows Compute Cluster Server 2003

Concepts

Objects	Descriptions
Cluster	The top-level organizational unit of CCP; comprising a set of nodes, a queue and jobs.
Queue	An organization unit which provides queuing and scheduling of jobs. There is only one queue in a CCP cluster and it contains pending, running and completed jobs. Completed jobs are purged periodically from the queue.
Job	A collection of tasks, using a task to perform a computation, initiated by a user. Jobs are used to reserve resources for subsequent use by one or more tasks.
Task	A task represents the execution of a program on some compute nodes. A task can be a serial program (single process) or an MPI program with multiple processes
Node	A single compute node, with one or more processors

Microsoft Windows Compute Cluster Server 2003

Common Job / Task Types

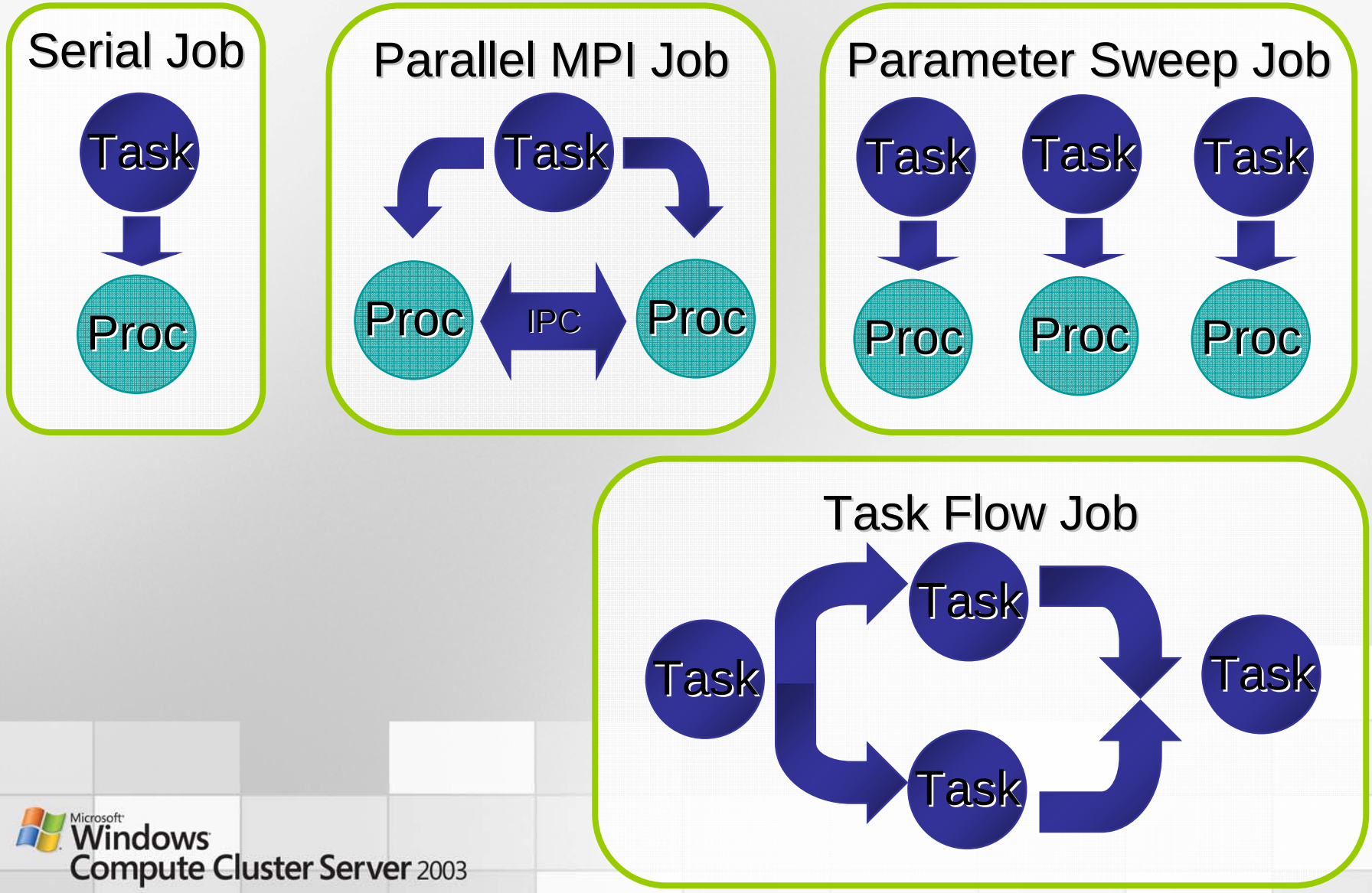


Task Task Task

Task

Proc Proc Proc Proc Task

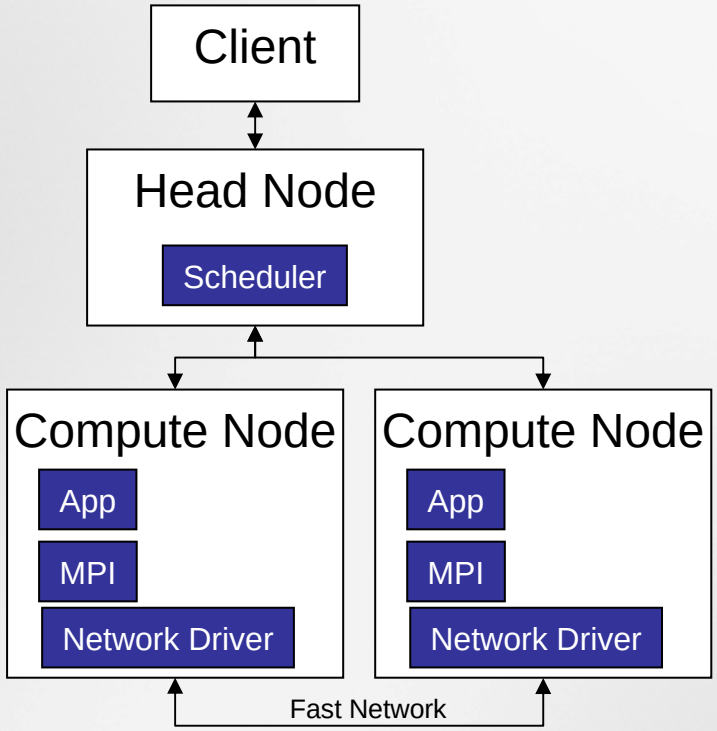
Job/Task Conceptual Model



Microsoft Windows Compute Cluster Server 2003

What Happens After I Click “Submit Job”?

- Task Execution
  - Scheduler orchestrates
    - Node allocation to the tasks
    - Timing, execution, and clean-up
    - Error Recovery
      - Re-try
      - Routing “around” un-responsive nodes
  - Within a security context
    - Compute nodes authenticate as the user
    - Secure client-scheduler-compute
    - Node communication
- The “Other” Layers
  - Application
  - Message Passing Interface (MPI): API for messaging between compute nodes cooperating on a task
  - Networking: drivers that enable fast communication via WinSock Direct



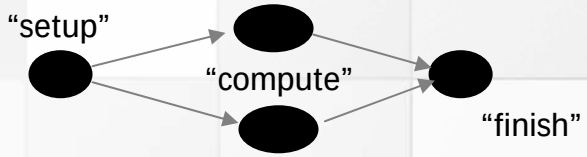
Microsoft Windows Compute Cluster Server 2003

Job Submission Syntax Examples

```
rem Run a serial job
job submit /stdout:%PDCSHARE%¥batchpi10.out %PDCSHARE%¥batchpi.exe 10

rem Run a parallel job
job submit /stdout:%PDCSHARE%¥batchpi100.out /numprocessors:2 mpiexec
   %PDCSHARE%¥batchpi.exe 100

rem Run a parametric sweep job
for /F "usebackq tokens=4" %j in (`job new`) do set JOBID=%j
for /L %i in (100, 100, 1000) do job add %JOBID%
   /stdout:%PDCSHARE%¥batchpi%i.out %PDCSHARE%¥batchpi.exe %i
job submit /id:%JOBID%
```



Microsoft Windows Compute Cluster Server 2003

Command Line Interface

Job	`job new [job_terms]`	create a job container
	`job add jobID [task_terms]`	add tasks to a job
	`job submit /id:jobid`	submit job / tasks
	`job submit [job_terms][task_terms]`	submit a job
	`job cancel jobID`	cancel a job
	`job modify [options]`	modify a job
	`job requeue JobID`	requeue a job
	`job list`	list jobs in the cluster
	`job listtasks`	list tasks of a job
	`job view JobID`	view details of a job
Task	`task view`	view details of a task
	`task cancel`	cancel a task
	`task requeue`	requeue a task
Cluscfg	`cluscfg view`	view details of a cluster
	`cluscfg params/setparams`	view/set config params
	`cluscfg listenvs/setenv`	list / set cluster wide env
	`cluscfg delcreds/setcreds`	set /delete user cred

Microsoft Windows Compute Cluster Server 2003

Task Terms

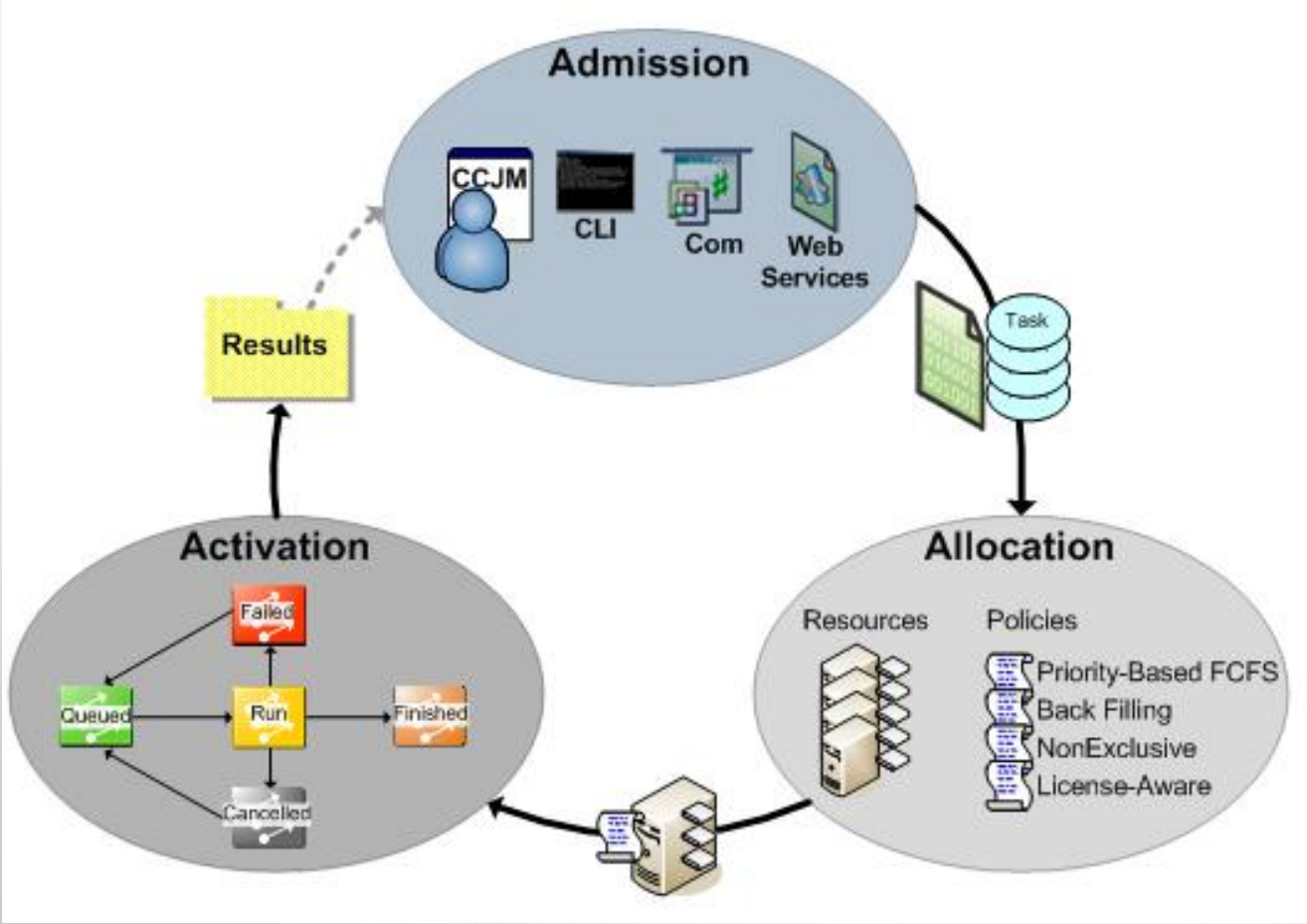
Task terms	Descriptions
cmdline	Command line of the task
env	Environment variables for the task
depend	Inter-task dependencies
rerunnable	Specifies that a task can be rerun after failure
workdir	Start-up folder of the task (or current working directory)
runtime	Specify the runtime of the task
numprocessors	Number of processors a task requires
name	Specifies the job name to associate with the job (in status displays)
project	Specifies the project name to associate with the job for accounting purposes
lic	Specifies the license features required for the job to run
priority	Highest, AboveNormal, Normal, BelowNormal, Lowest
askednodes	Specific list of nodes the task runs on
exclusive	non-exclusive allocation of nodes a task
stdin	Redirect standard input of the task from a file
stdout	Redirect standard output of the task to a file
stderr	Redirect standard error of the task to a file

Job Terms

Job Term	Description
numprocessors	Specifies the number of compute processors to be reserved across a set of nodes
askednodes	Specifies a list of nodes
priority	`Highest` Cluster Admin `AboveNormal` Cluster Admin `Normal` Cluster User `BelowNormal` Cluster User `Lowest` Cluster User
runtime	Specifies the wall clock runtime limit of the job
exclusive	Specifies whether the nodes are allocated to jobs exclusively
name	Specifies the job name to associate with the job (in status displays)
project	Specifies the project name to associate with the job for accounting purposes
license	Specifies the license features and amt required for the job to run

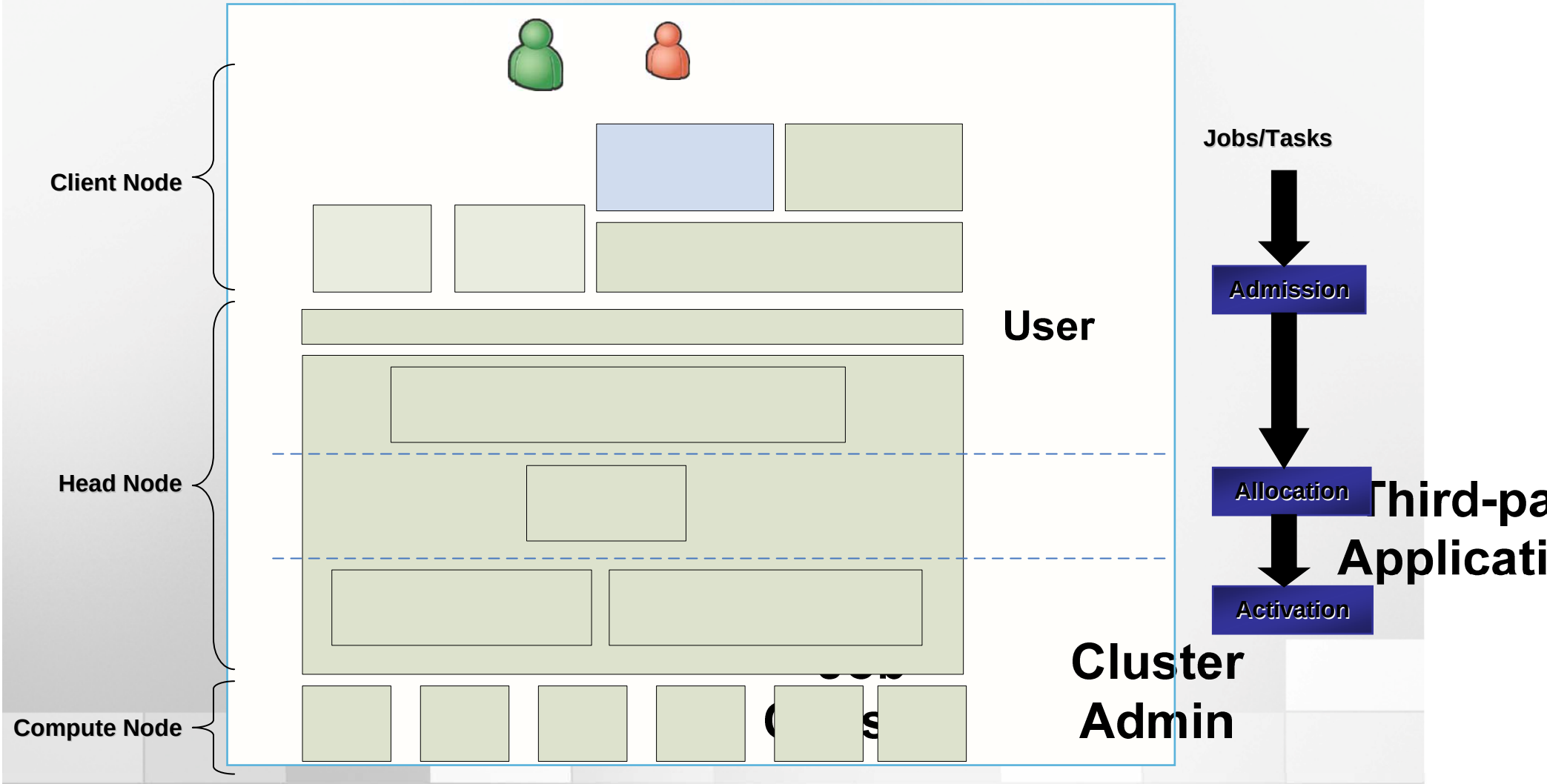
Job Manager UI

Job Life Cycle



Microsoft Windows Compute Cluster Server 2003

Job Scheduler Stack

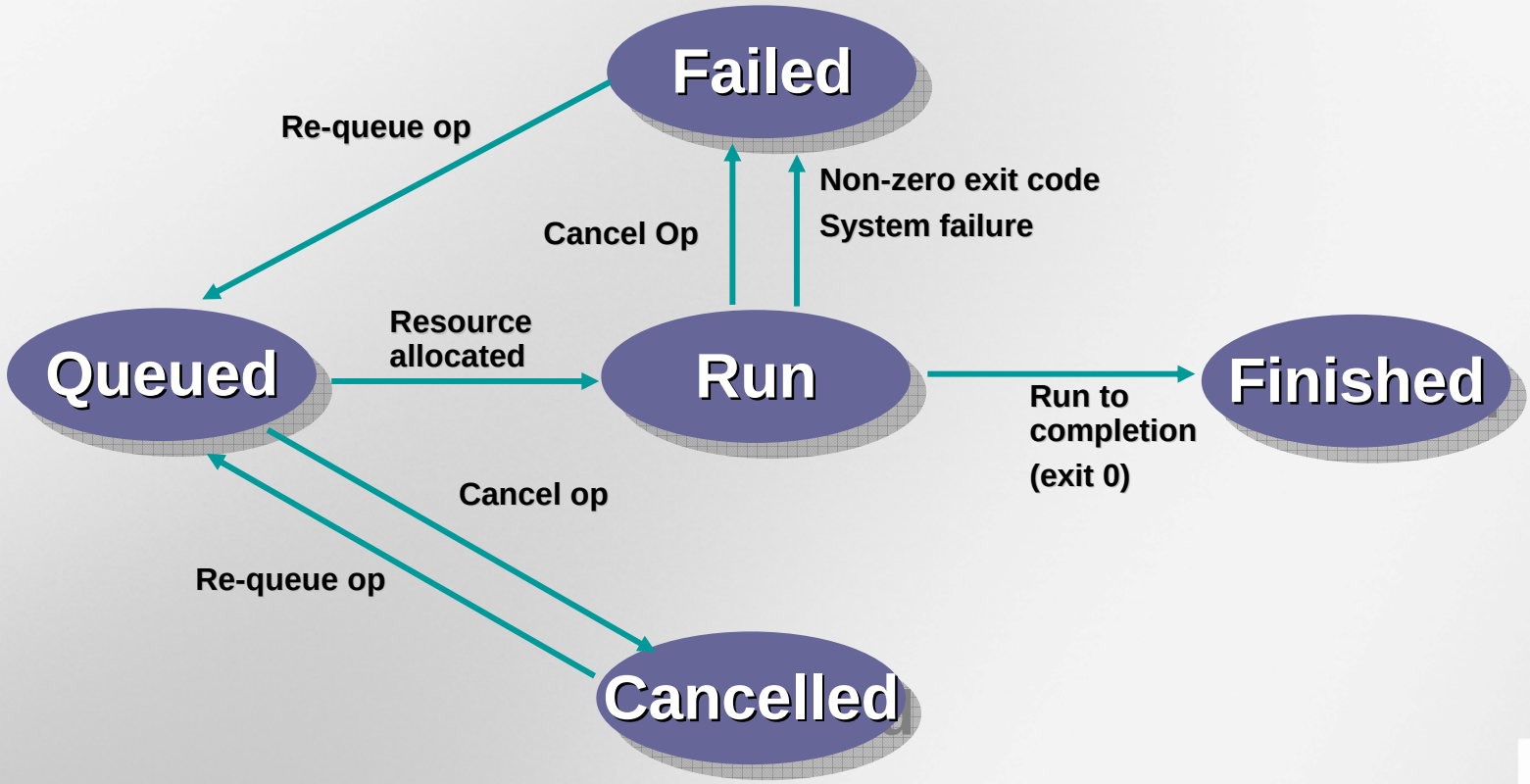


Microsoft Windows Compute Cluster Server 2003

Job Scheduler

- Job ordering
  - Priority-based first-come, first-serve (FCFS)
  - Backfill supported for jobs with time limits
- Resource allocation
  - License-aware scheduling through plug-ins
  - Parallel application node allocation policies
- Extensible
  - Third parties can extend by using standard submission and execution
  - Partnering with third parties for deeper integration in the future
- Job submission
  - Jobs submitted via several methods (e.g. Web, command line)
- Cleanup
  - Jobs handled uniformly, facilitates easy and complete cleanup

Task State Transition



Microsoft Windows Compute Cluster Server 2003

Pi Demo

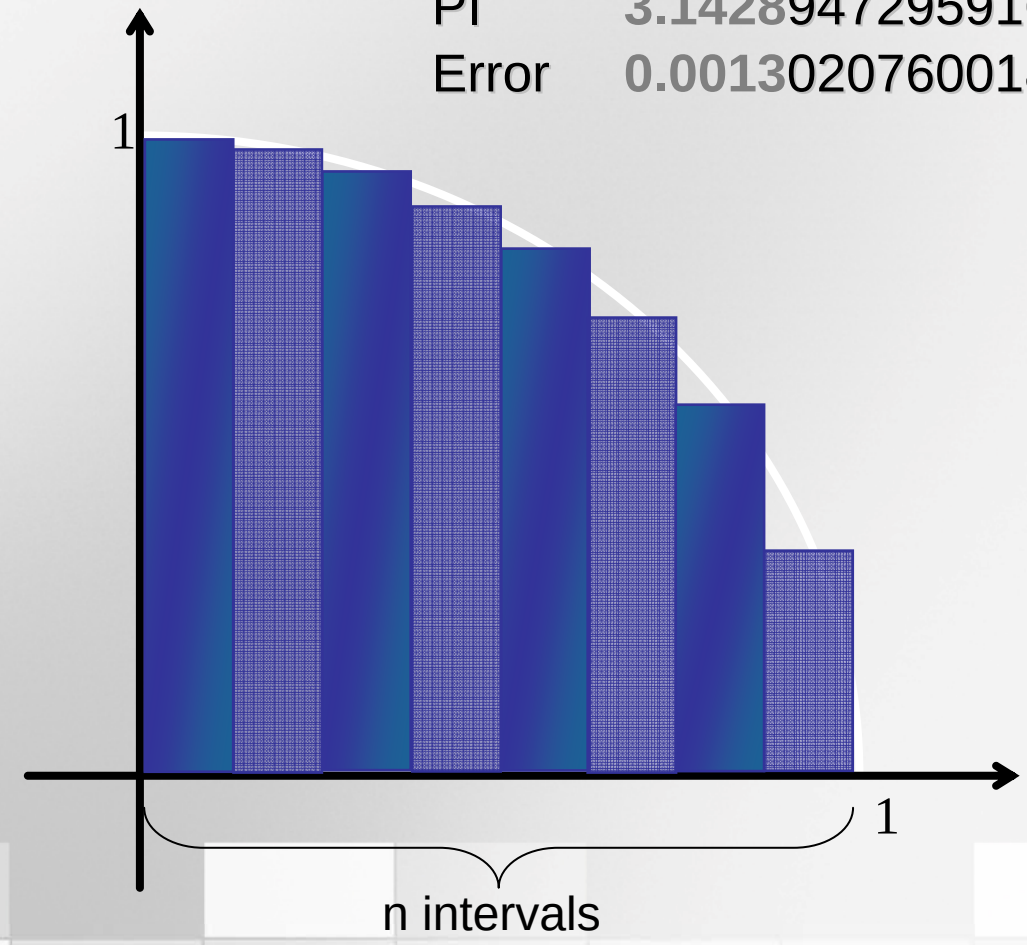
Show the CLI

- Run the Pi calculation program
- Submit the Pi program
- View job status
- Show the results

Microsoft Windows Compute Cluster Server 2003

Example: Calculate Pi

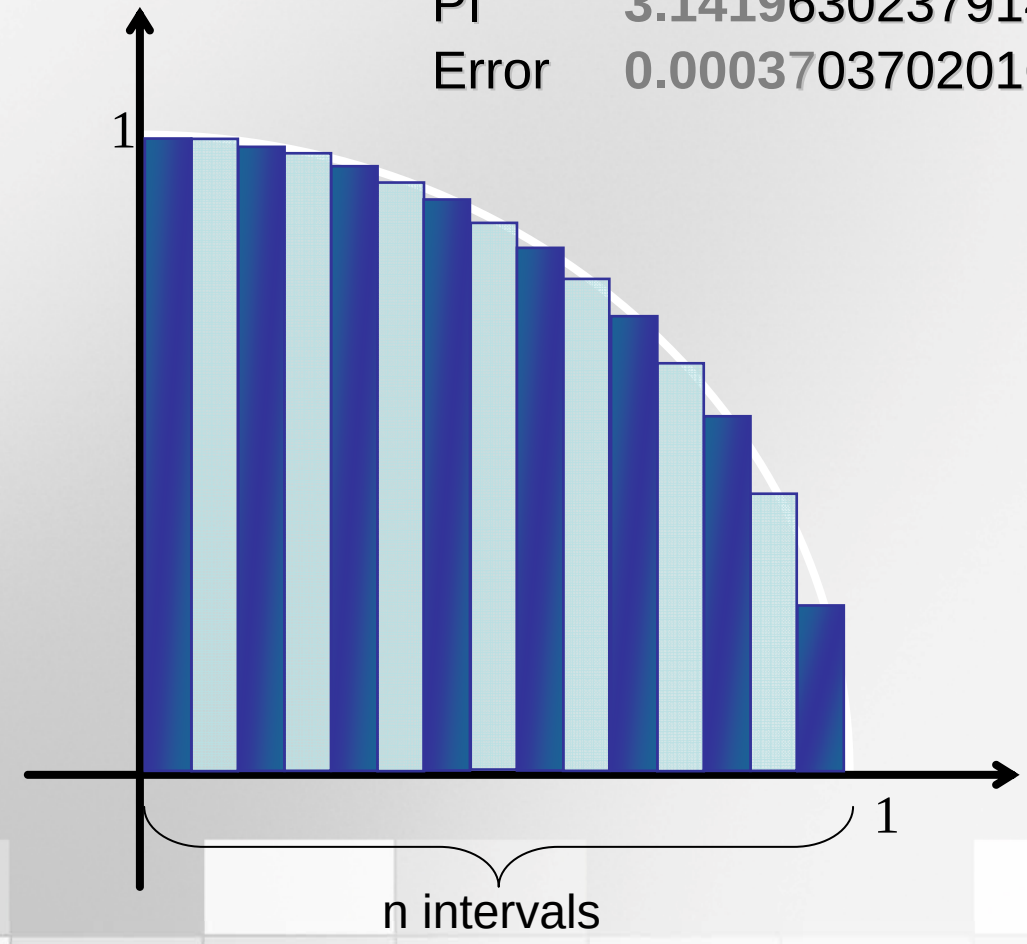
PI 3.1428947295916889
Error 0.0013020760018958



Microsoft Windows Compute Cluster Server 2003

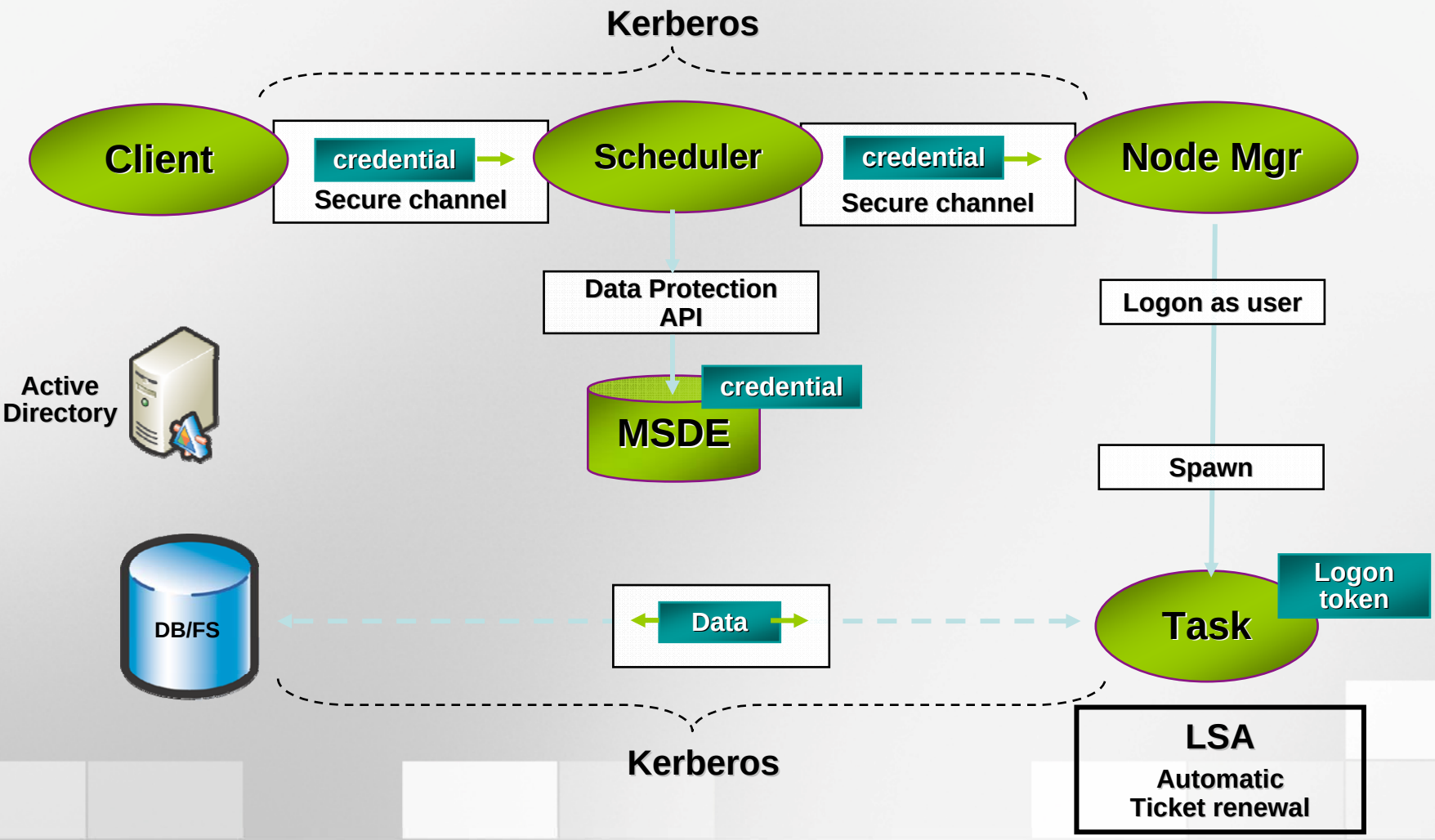
Example: Calculate Pi in Parallel

PI 3.1419630237914182
Error 0.0003703702016251



Microsoft Windows Compute Cluster Server 2003

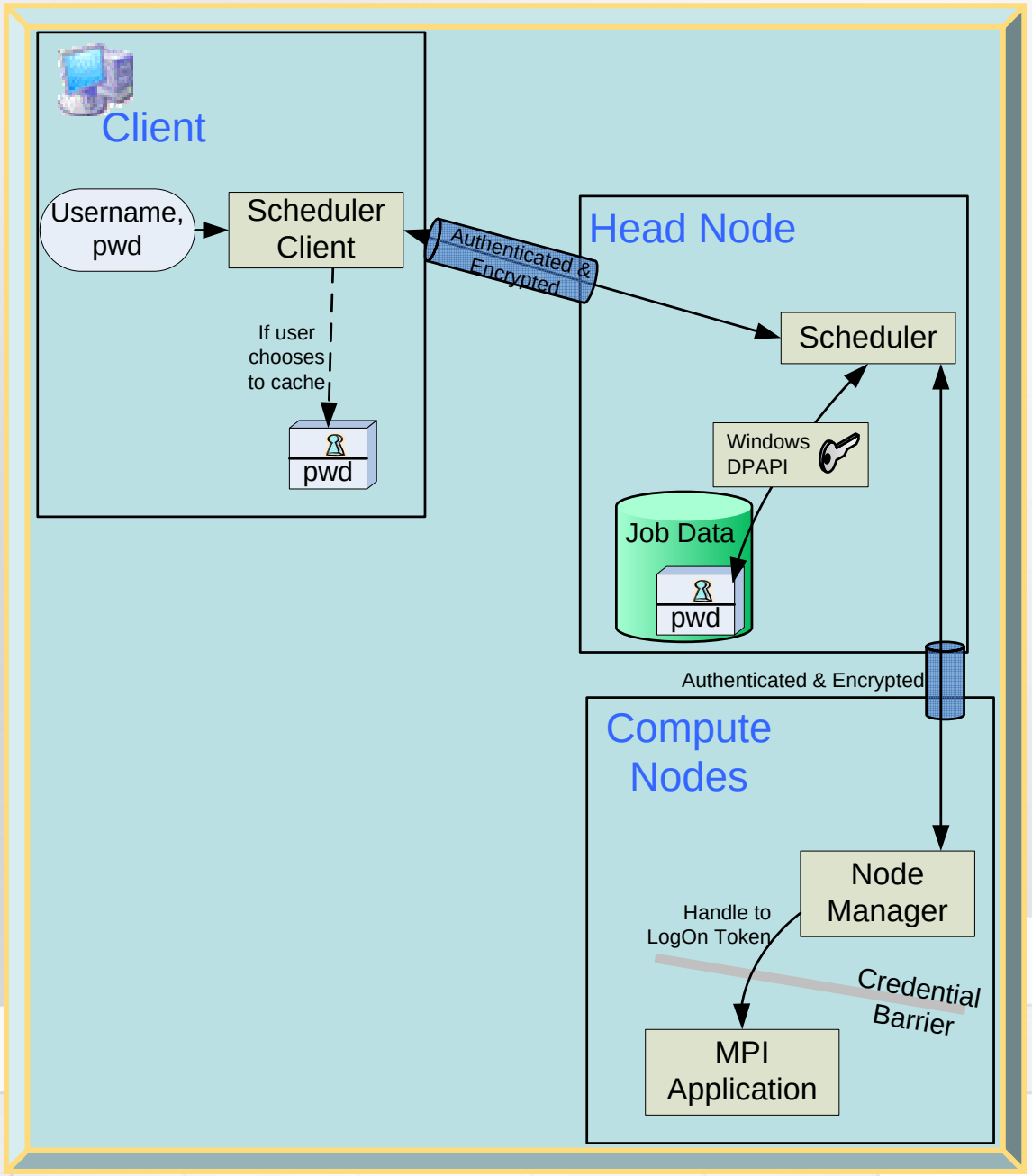
End-To-End Security



Microsoft Windows Compute Cluster Server 2003

Credential Handling

- Credential stored with job data in scheduler
  - Communications are authenticated and encrypted
  - Credential is erased after job is complete
- Credential optionally cached on client
  - Encrypted with key known only to scheduler



Microsoft Windows Compute Cluster Server 2003

Demo

- Run the cmd.exe scripts
  - Serial job
  - Parallel job
  - Parametric sweep job
  - Flow job

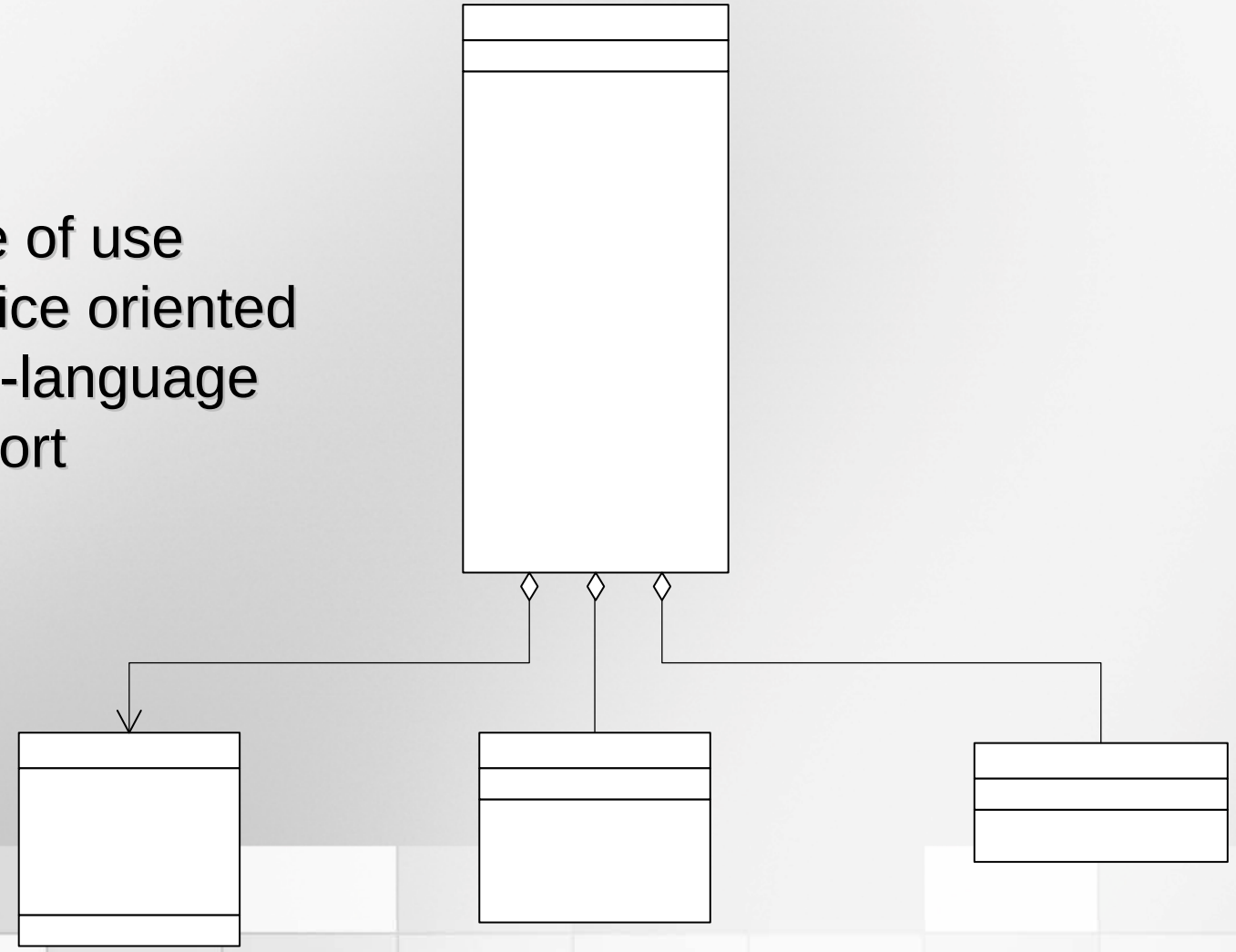
Serial Job, Parallel Job, Parametric Sweep Job, and Flow Job

Microsoft® Windows® Compute Cluster Server 2003

HPC goes mainstream

Scheduler Object Model

- Ease of use
- Service oriented
- Multi-language support



C

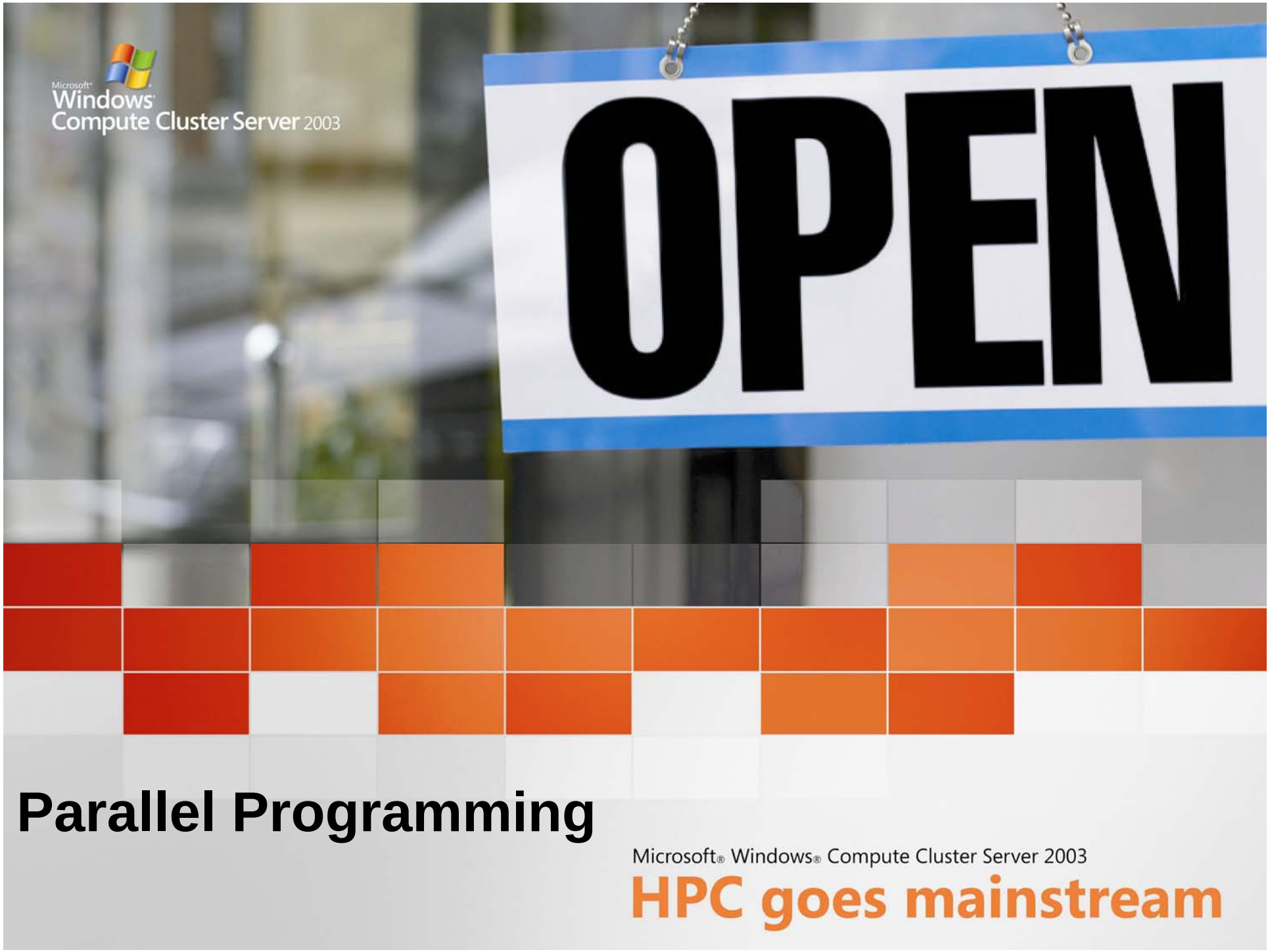
+Connect()
+ApproveNo
+PauseNod
+ResumeNo
+CreateJob(
+CreateJobI
+ListJobs()
+QueueJob(
+SubmitJob
+CancelJob
+RequeueJo
+ModifyJob(
+SetCached
+DeleteCach
+GetJobRes
+CreateTas

Microsoft Windows Compute Cluster Server 2003

External Resources

- Microsoft HPC Web site
  - http://www.microsoft.com/hpc/
- MSDN documents
  - http://msdn.microsoft.com/library/default.asp?url=/library/en-us/ccpsdk/ccp/microsoft_compute_cluster_pack.asp
- Windows Server x64 information
  - http://www.microsoft.com/64bit/, http://www.microsoft.com/x64/
- Windows Server System information
  - http://www.microsoft.com/wss/
- Get the Facts Web site
  - http://www.microsoft.com/getthefacts
- GotDotNet Gallery for HPC
  - http://www.gotdotnet.com/codegallery/codegallery.aspx?id=0c12237e-58e3-4a9f-9637-ec6a4b9efe4d
- Blog
  - http://blogs.msdn.com/hpc

Microsoft Windows Compute Cluster Server 2003

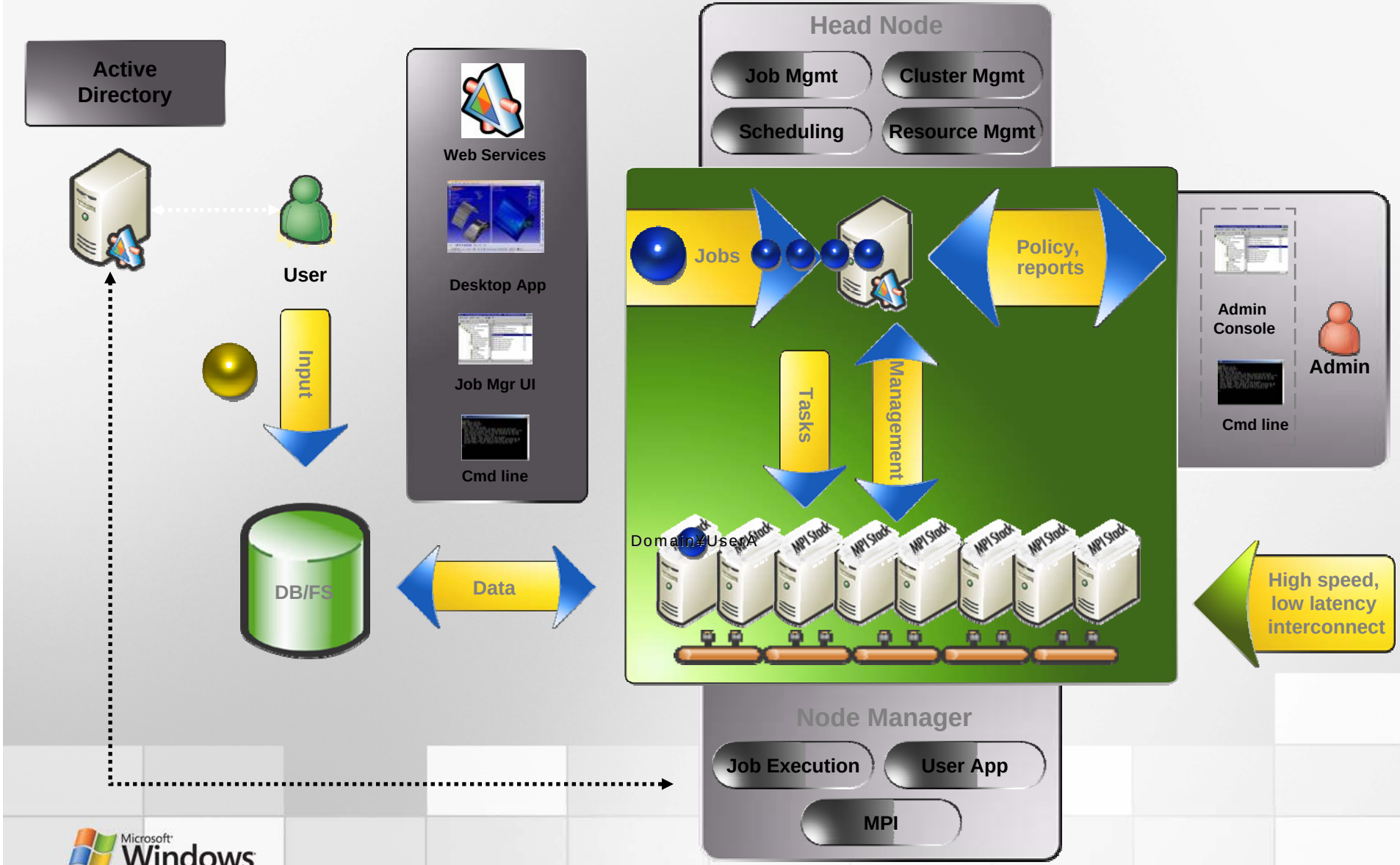


Parallel Programming

Microsoft® Windows® Compute Cluster Server 2003

HPC goes mainstream

Windows Compute Cluster Server



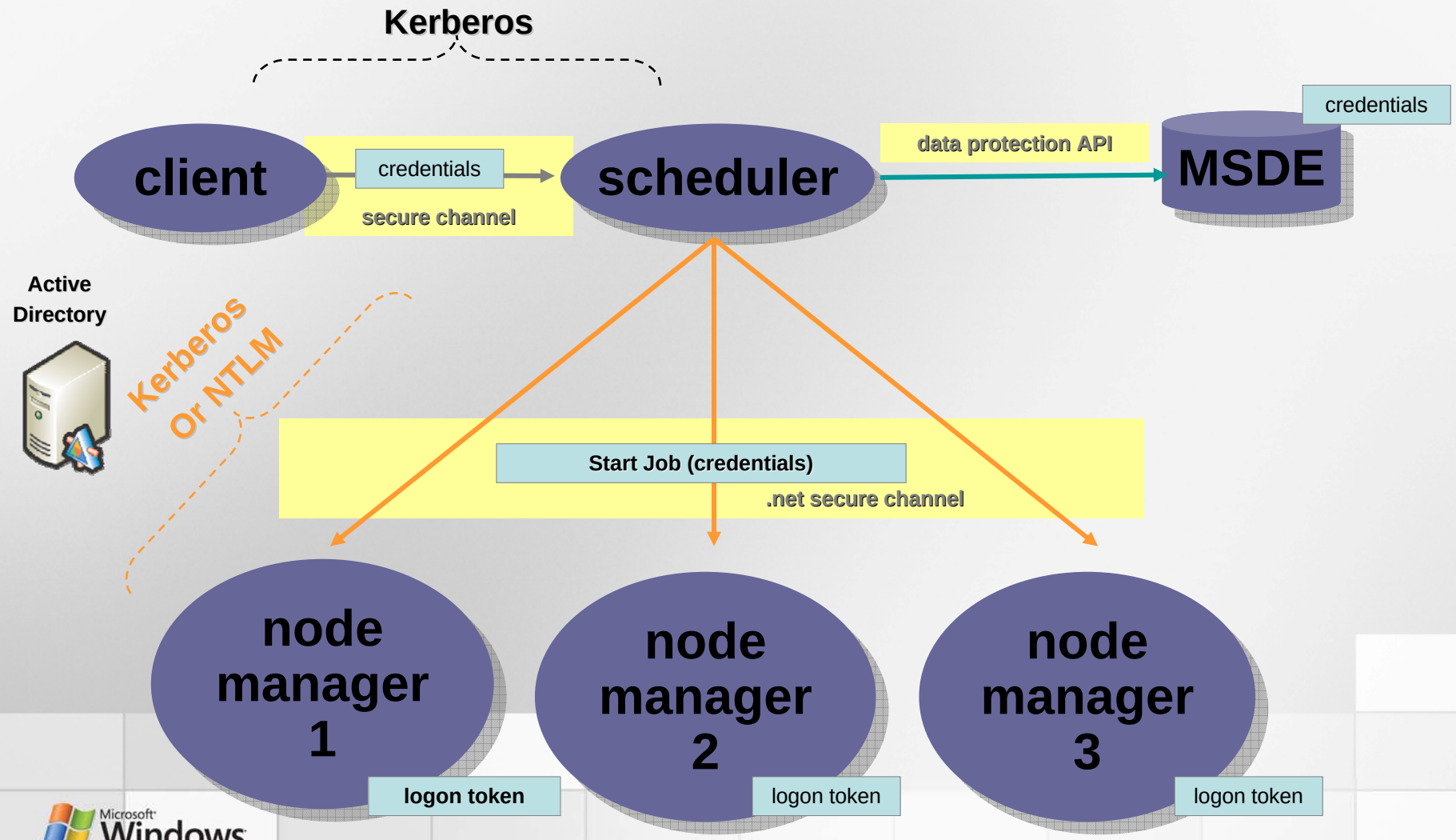
Microsoft Windows Compute Cluster Server 2003

Compute Cluster Server's Developer Environment

- Compute Cluster Server's Scheduler software
  - Programmatic job submission/control
- Compute Cluster Server's MPI software
  - Derived from Argonne National Lab's MPI-2 implementation (MPICH2)
  - MS MPI consists of 2 parts
    - For ISVs: Full-featured API of160+ functions (can do much work with ~10 functions!)
    - For Users: Command-line (mpiexec) or GUI tool to launch jobs
  - Can use ANY MPI STACK on CCS
- Visual Studios 2005
  - New parallel debugger!
  - MPI & Open Message Passing (OpenMP) support

Secure MPI Activation

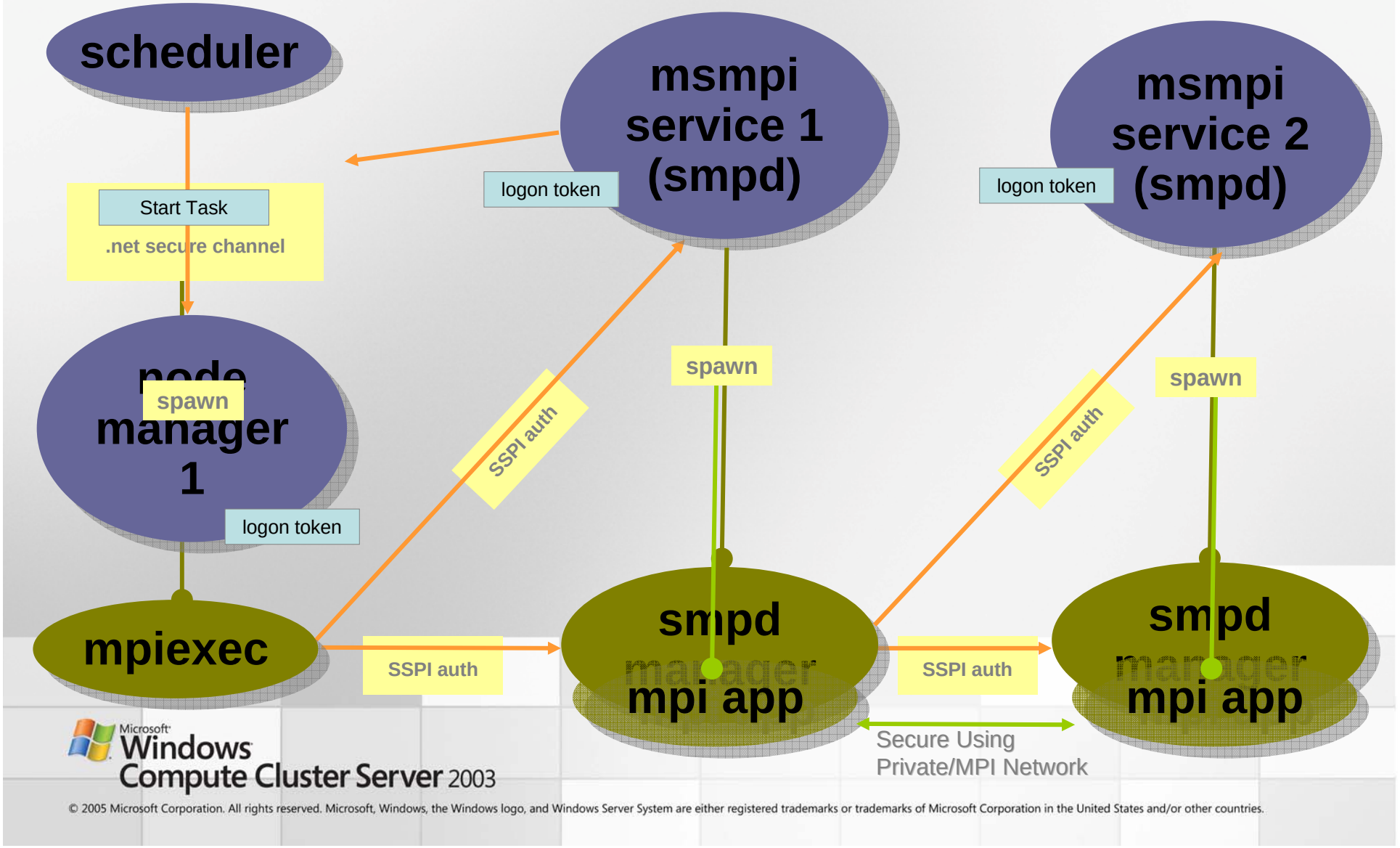
Resource Allocation



Microsoft Windows Compute Cluster Server 2003

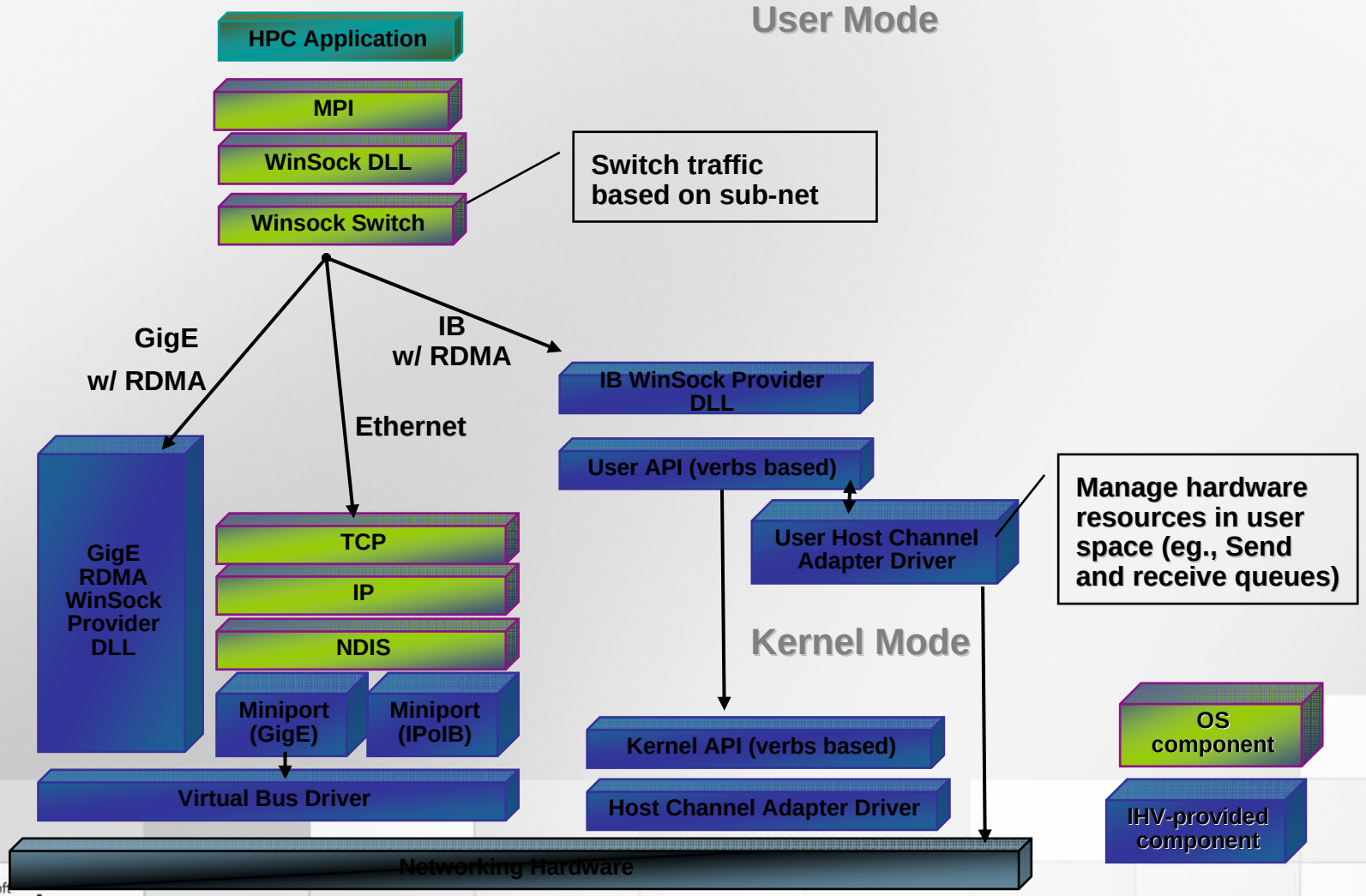
Secure MPI Activation

Task Activation



Microsoft Windows Compute Cluster Server 2003

MS MPI Uses Winsock Direct for Low Latency w/ User Flexibility



Microsoft Windows Compute Cluster Server 2003

Message Passing

How Parallel Programming…Gets Parallel

2 Popular Methods of Message Passing for HPC

- OpenMP
  - Works in shared memory situations only- inside one physical chassis
  - Easy to program- simply identify parallelizable loops with pragmas
- Message Passing Interface (MPI)
  - Works in shared memory (single chassis) and cluster (multiple chassis) situations.
  - Ample MPI API
    - Fine-grained control of messaging between processes
    - Collect and reduce data in a single call
  - Requires moderate-advanced level of programming
    - Harder than OpenMP but MUCH easier than sockets

OpenMP Support in VS2005

- A specification for multithreaded programs
  - Eases hyper-threading
- Conformance to the OpenMP 2.5 standard
- Support for .NET and OpenMP together
  - Compiler generates MSIL for OpenMP code
- Simple to program with OpenMP
  - Set of simple #pragmas and runtime routines
    - #pragma omp parallel
    - Start with sequential code and parallelize by adding #pragmas
- Most value
  - Parallelizing large loops without loop-dependencies
    - Can do more, but that’s the big win

About MPI

- Early HPC systems (Intel's NX, IBM's EUI, etc) were not portable
- The MPI Forum organized in 1992 with broad participation by
  - vendors: IBM, Intel, TMC, SGI, Convex, Meiko
  - portability library writers: PVM, p4
  - users: application scientists and library writers
- MPI is a standard specification, there are many implementations
  - MPICH and MPICH2 reference implementations from Argonne
  - **MS MPI based on (and compatible with) MPICH2**
  - Other implementations include LAM-MPI, OpenMPI, MPI-Pro, WMPI
- Why did MS HPC team choose MPI?
  - MPI has emerged as **de-facto** standard for parallel programming
- MPI consists of 3 parts
  - **Full-featured API of 160+ functions**
  - **Secure process launch and communication runtime**
  - **Command-line (mpiexec) to launch jobs**

Fundamental MPI Features

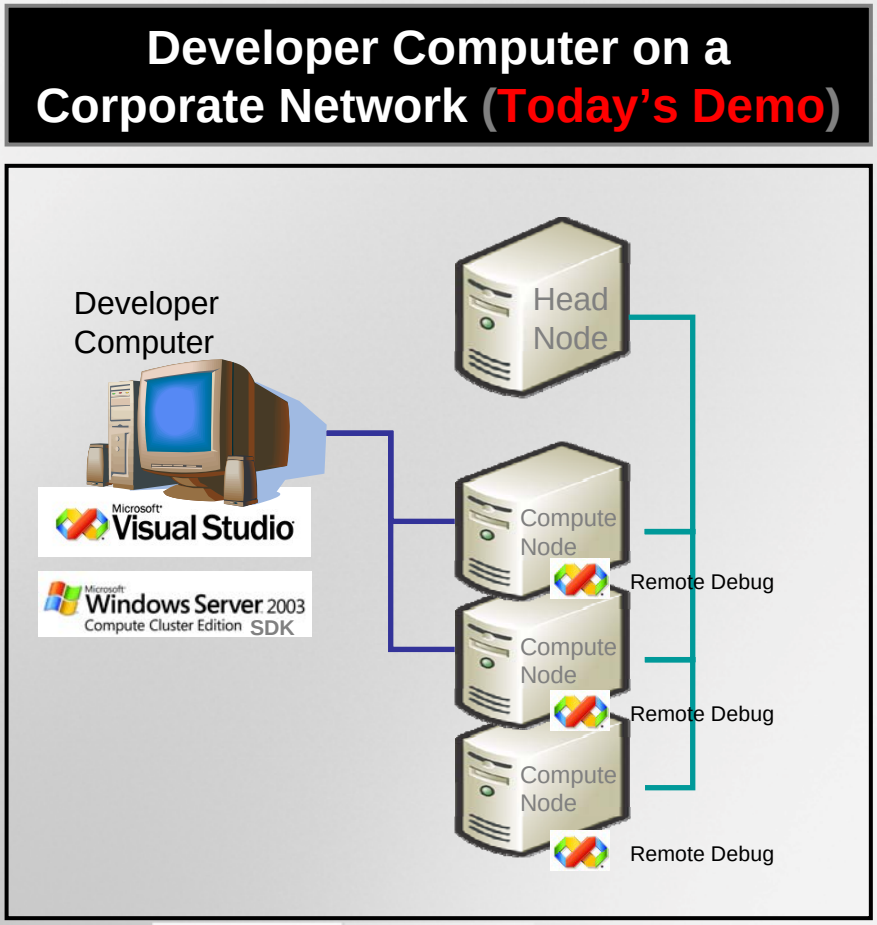
Programming with MPI

- Communicators
  - Groups of nodes used for communications
  - MPI_COMM_WORLD is your friend
- Rank (a node's ID)
  - Target communications
  - Segregate work
- Collective Operations
  - Collect and reduce data in a single call
  - sum, min, max, and/or, etc
- Fine control of comms and buffers if you like
- MPI and derived data types

Launching Jobs

- MPIexec arguments
  - # of processors required
  - Names of specific compute nodes to use
  - Launch and working directories
  - Environment variables to set for this job
  - Global values (for all compute nodes- not just the launch node)
  - Point to files of command line arguments
  - Env MPICH_NETMASK to control network used for this MPI job

A Developer's View of Cluster Connectivity



Microsoft Windows Compute Cluster Server 2003

Standard, Integrated Development

- Cluster scheduler programmable via .NET and DCOM
- Visual Studio 2005 – OpenMP, Parallel Debugger
- Support for MPI2 based on a reference implementation
  - Standardized MPI stack
  - User credentials secured in job scheduler and compute nodes
  - Performance improvements that take advantage of advanced Windows I/O architecture
  - Support for multiple advanced interconnect types via Winsock Direct architecture
  - Microsoft provided stack reduces application/MPI incompatibility issues

Visual Studio® 2005

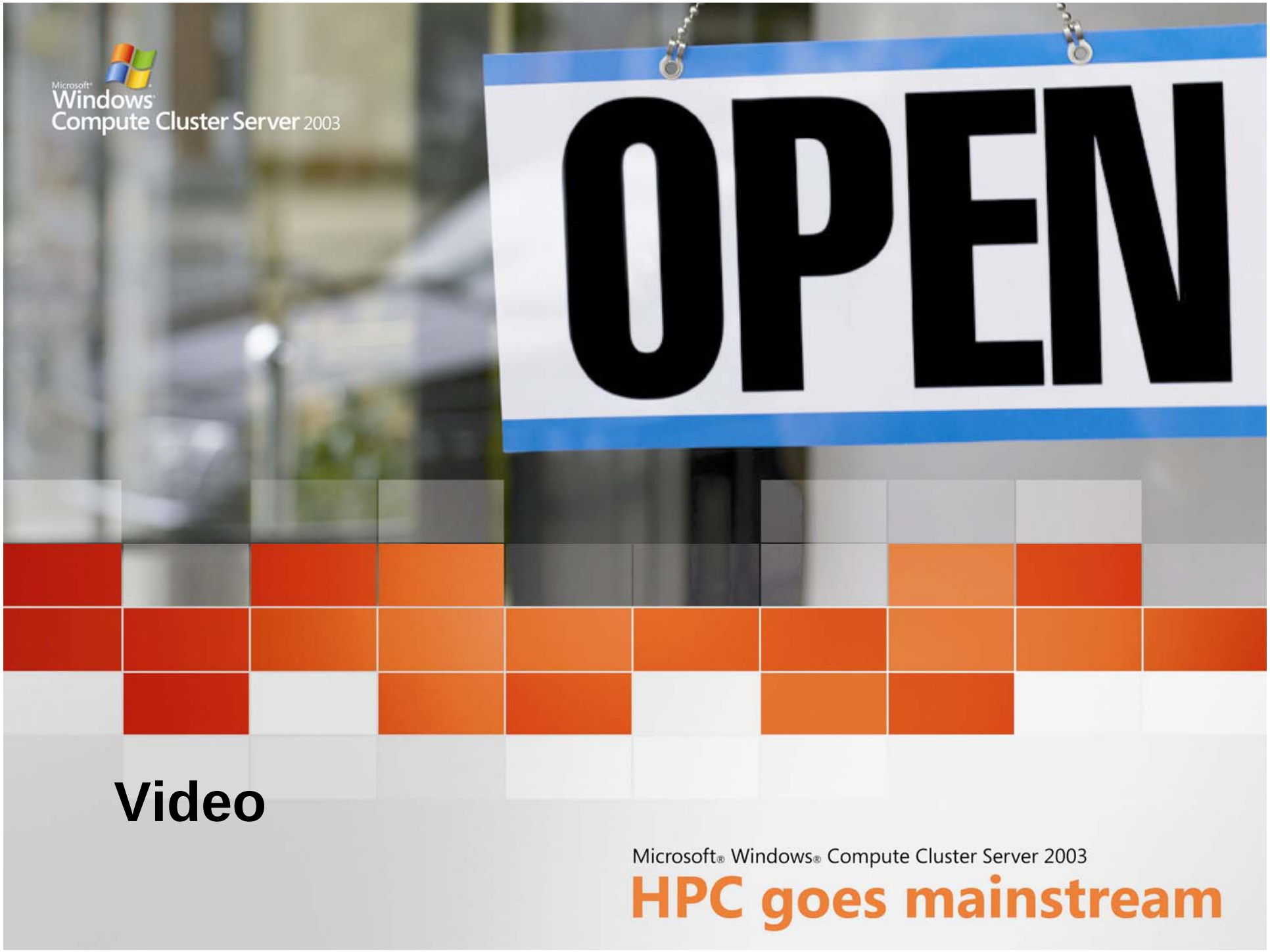
Engineered to provide support for development of HPC applications.

Microsoft Windows Compute Cluster Server 2003

Parallel Debugger in Visual Studio

- Basic features to debug MPI applications
- Features:
  - Automatic attach to MPI processes from IDE
  - Process level stepping
  - Process breakpoints
  - Process sensitive expression evaluation

Microsoft Windows Compute Cluster Server 2003



Video

Microsoft® Windows® Compute Cluster Server 2003

HPC goes mainstream

Summary

- Windows Compute Cluster Server 2003 is Microsoft's first step towards makes HPC accessible to all scientists, engineers, and businesses
- Windows Compute Cluster Server 2003 delivers …
  - Faster time-to-insight through simplified cluster deployment, job submission and status monitoring
  - Better integration with existing Windows infrastructure allowing customers to leverage existing technology and skill-sets
  - Familiar development environment allows developers to write parallel applications from within the powerful Visual Studio IDE

Microsoft Windows Compute Cluster Server 2003

CCS & MS MPI Information

- Microsoft HPC web site
  - http://www.microsoft.com/hpc
- Argonne National Lab's MPI website
  - http://www-unix.mcs.anl.gov/mpi/
- Tutorial from Lawrence Livermoore National Lab
  - http://www.llnl.gov/computing/tutorials/mpi/
- Tuning MPI Programs for Peak Performance
  - http://mpc.uci.edu/wget/www-unix.mcs.anl.gov/mpi/tutorial/perf/mpiperf/index.htm
- Microsoft HPC newsgroup
  - microsoft.public.windows.hpc

Microsoft®

Your potential. Our passion.™

Microsoft Windows Compute Cluster Server 2003