HyperTransport® CONSORTIUM

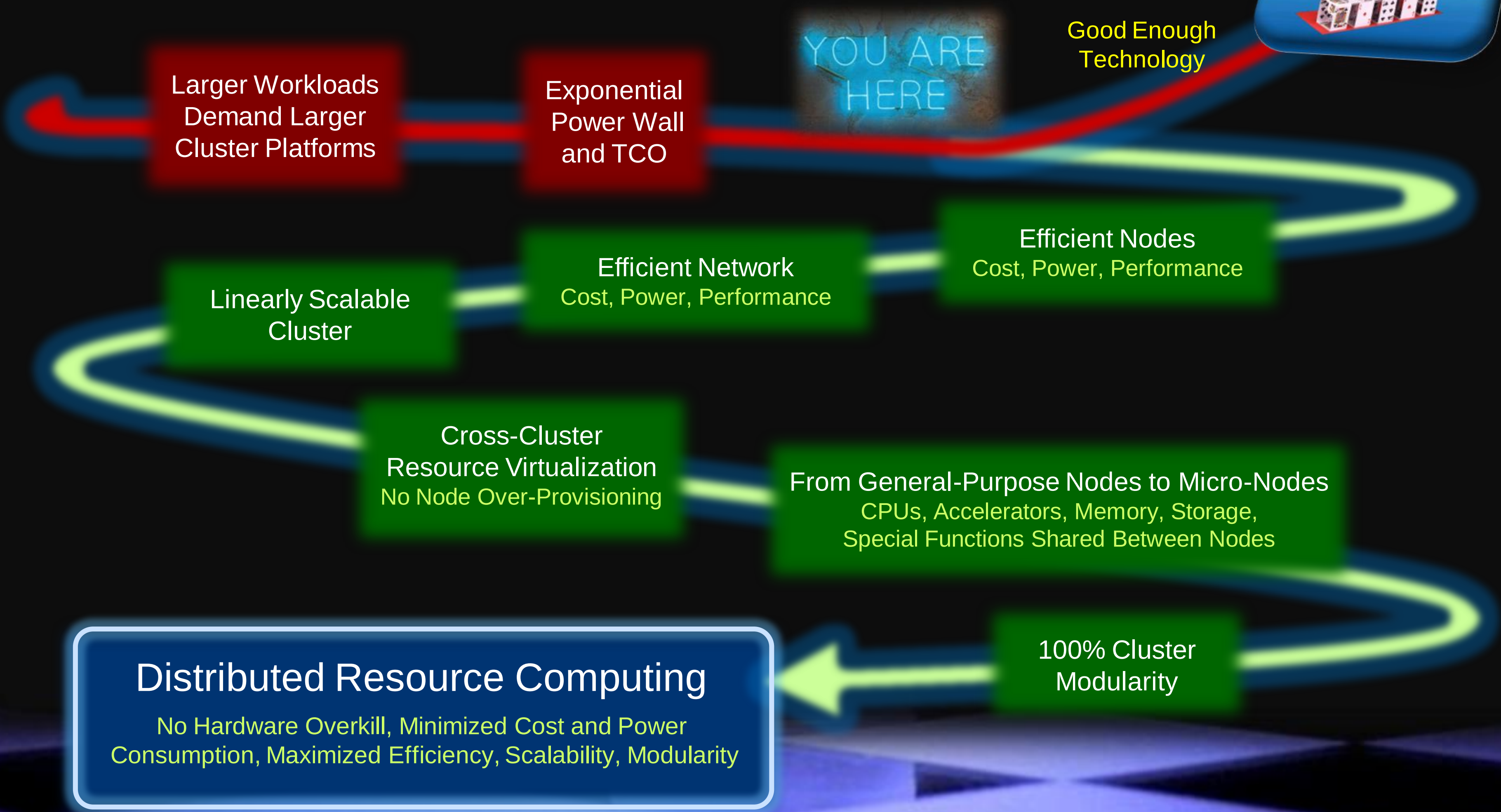
Time to Think Differently

“Good Enough” Clustering Technology is Out

Distributed Resource Computing is In

Full Cluster Resource Virtualization, Minimized TCO
and Power Consumption, Maximized Efficiency, Scalability, Modularity

From Clustering to Distributed Resource Computing



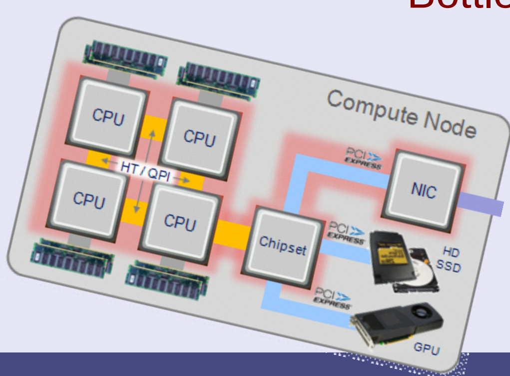
From Clustering to Distributed Resource Computing

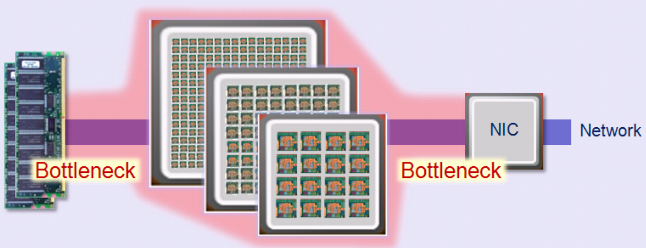
Efficient Nodes
Cost, Power, Performance

Scale Down of High-End CPU Power Footprint in Progress

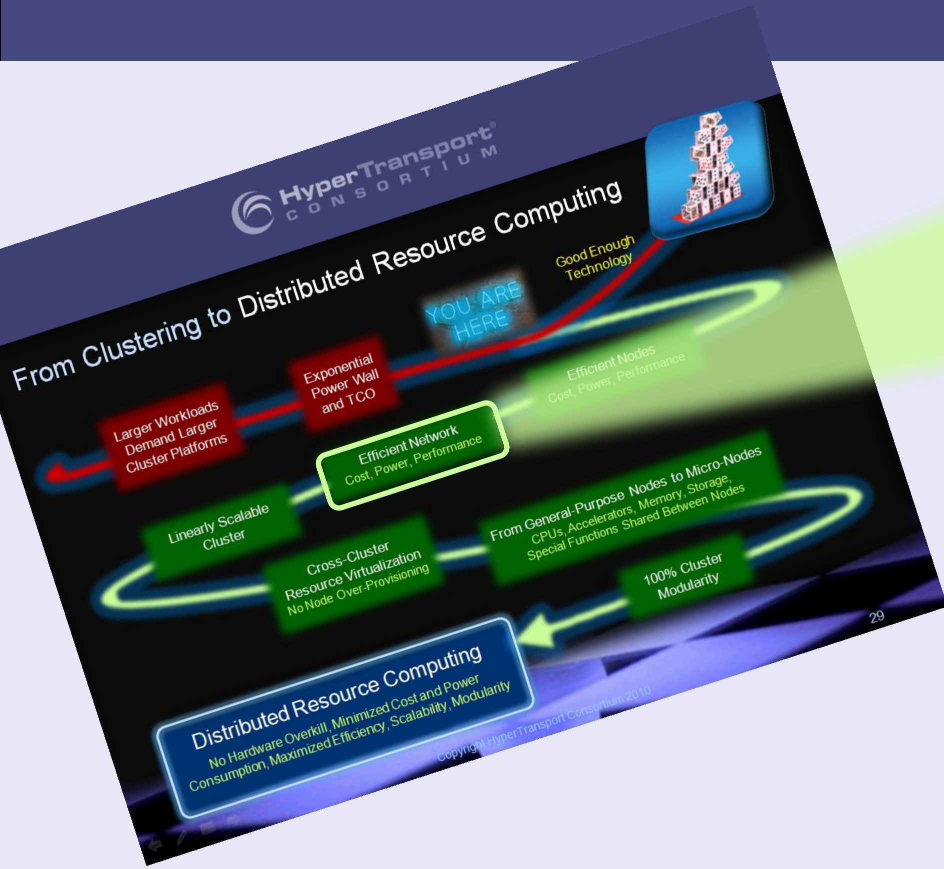
Power-Miser, Low-End, x86 and Non-x86 Processors Vying as Cost- and Power-Conscious CPU Choices for Clusters

Core Proliferation at Threshold of Pain Level for Power Consumption and Performance Bottlenecks





From Clustering to Distributed Resource Computing (cont.)



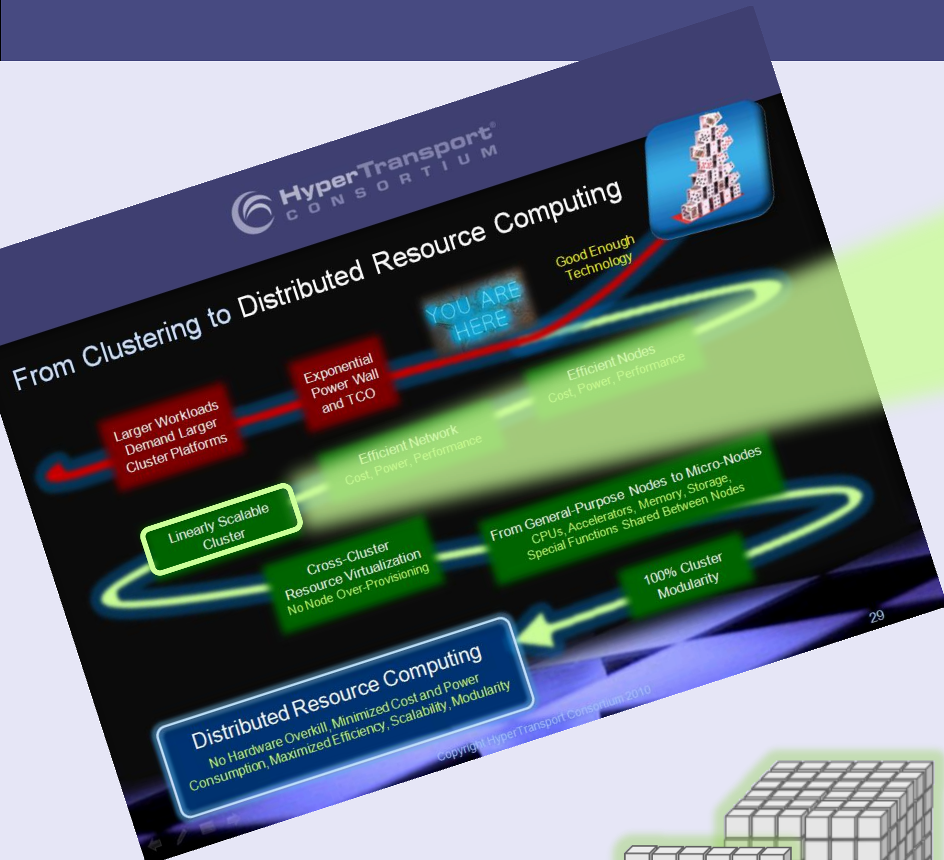
Efficient Network
Cost, Power, Performance

New-Gen of Ethernet and InfiniBand Switches Deliver Some Form of Power/Cost Reduction

Switched Fabrics Remain Power Hogs

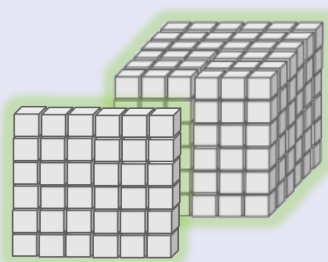
Torus Topologies' Cost, Power, Performance Efficiency Now Extending from Supercomputers to Commercial Cloud

From Clustering to Distributed Resource Computing (cont.)



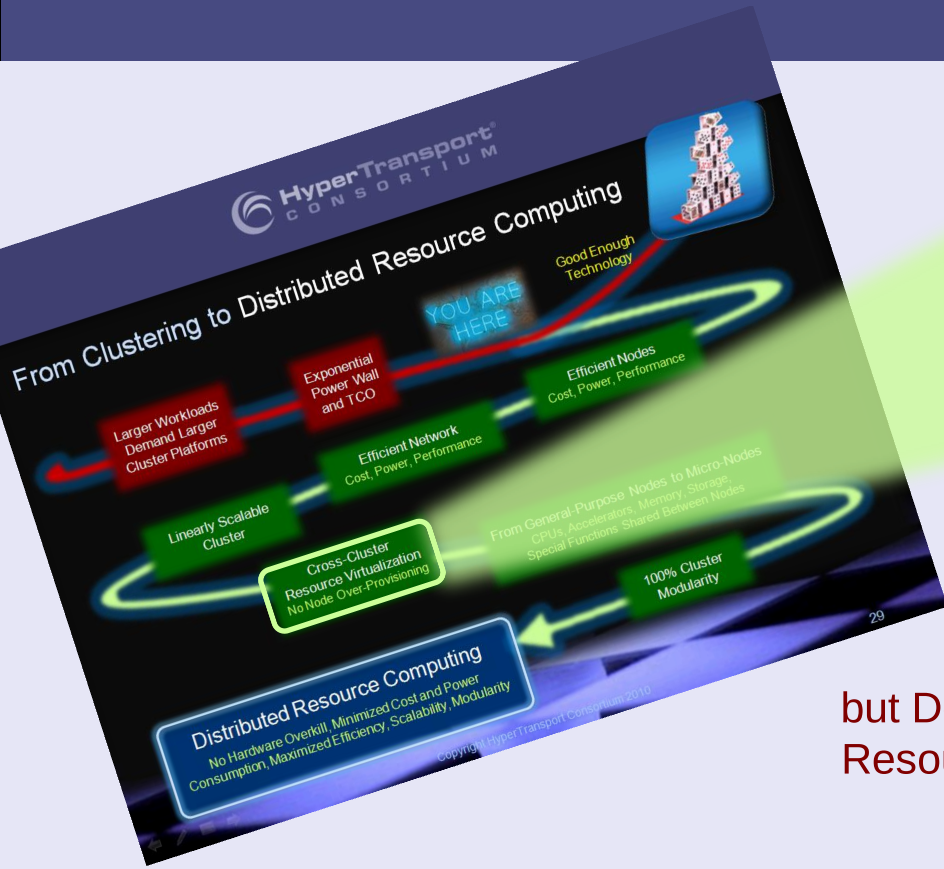
Linearly Scalable Clusters

Fat Tree Network Scalability Limited by Switch Port Capacity – i.e. an Increase in Ports per Switch Increases the Max Node Count Supported. Conversely and in Proportional Measure an Increase in Ports per Switch Decreases Both Cluster Scaling Linearity and Granularity



Torus Network Scalability Not Limited by Any Hardware Configuration, Scales Linearly and Granularity is Maximized

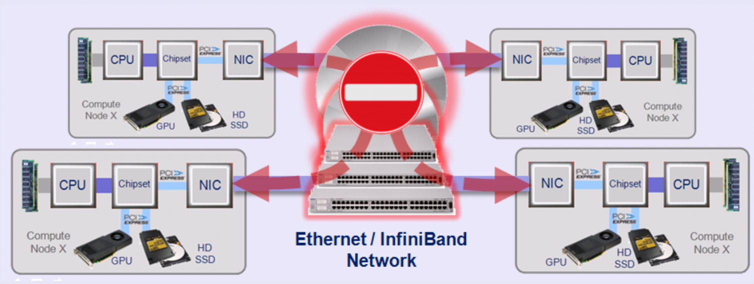
From Clustering to Distributed Resource Computing (cont.)



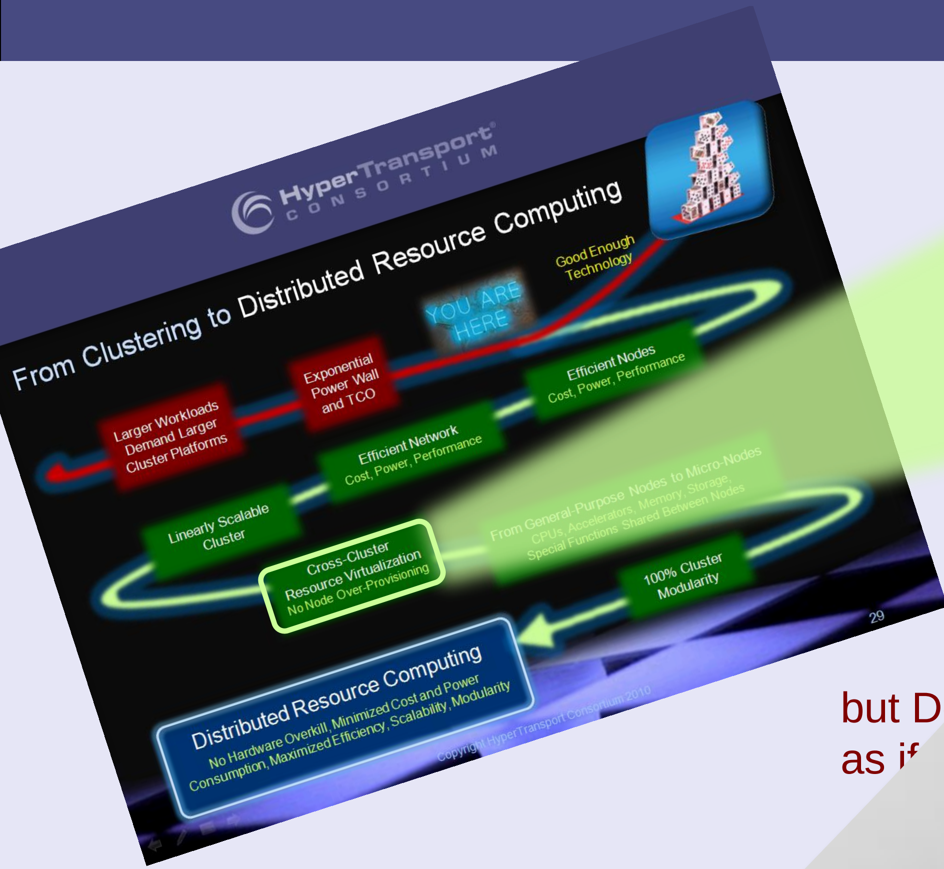
Cross-Cluster
Resource Virtualization
No Node Over-Provisioning

Cache Coherent Architectures Deliver Full Cross-Cluster Hardware and Software Virtualization to Small Clusters (10s of Nodes)

Message-Passing Protocols Used with Ethernet and InfiniBand Networks Allow Nodes of Large Clusters to Exchange Data but Do Not Allow them to Efficiently Hardware-Share Resources as if they were a Single Large Node



From Clustering to Distributed Resource Computing (cont.)



Cross-Cluster
e Virtualization
rovisioning

Cache
Cross-
Virtua

odes

Me
F

y

Any Size Clusters

Until Today

but Do
as if

ols

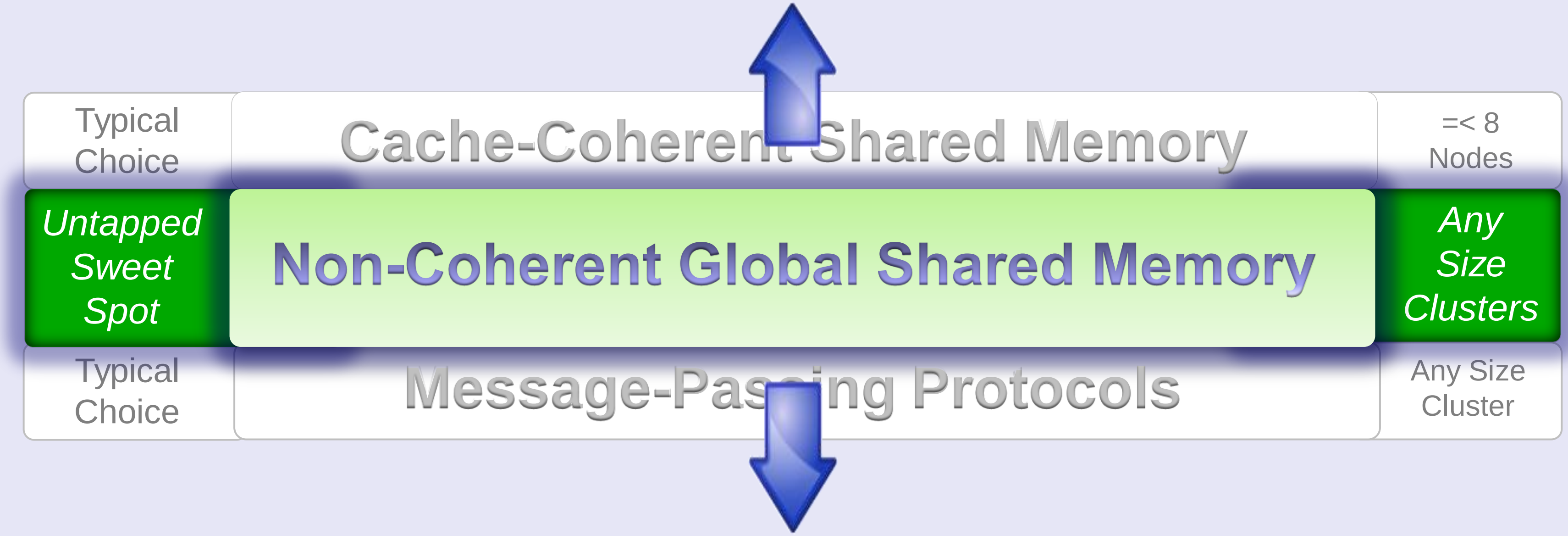
Any Size Cluster

er Footprint
astructure Costs
st Operational Costs

From Clustering to Distributed Resource Computing (cont.)

Sweet Spot Between Architectural Extremes

High Performance
High Node Scalability
Cross-Cluster Hardware Virtualized Resources



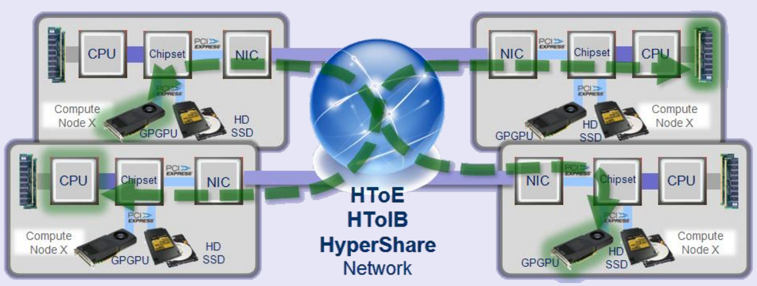
Lowest Power Footprint
Lowest Infrastructure Costs
Lowest Operational Costs

From Clustering to Distributed Resource Computing (cont.)

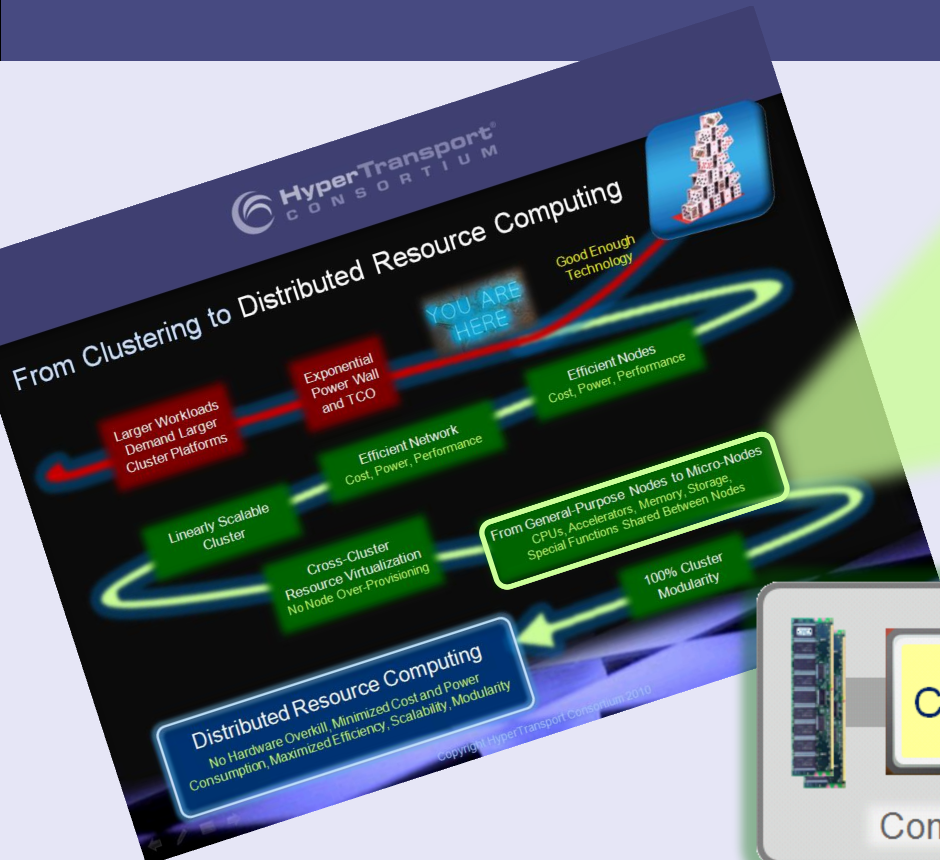
From General-Purpose Node to Micro-Node
CPU, Acceleration, Memory, Storage,
Special Functions Shared Between Nodes

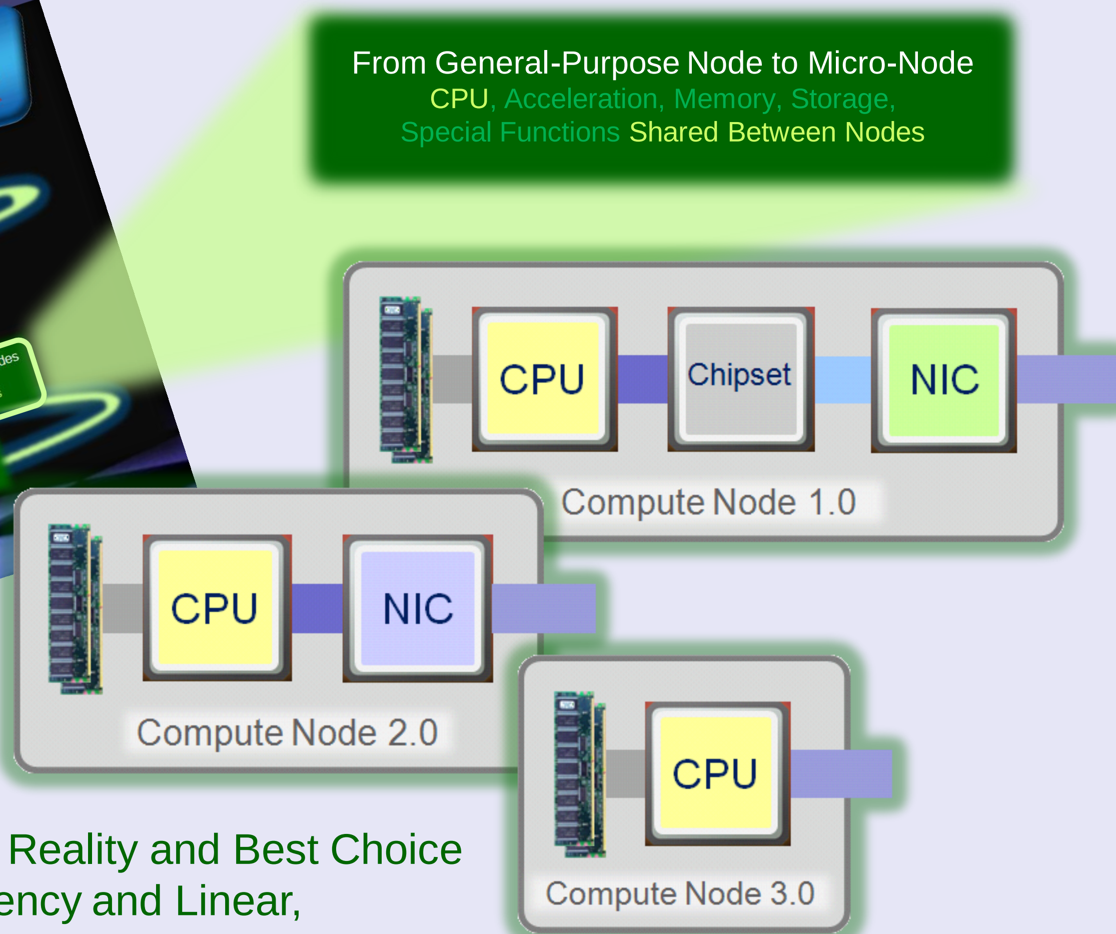
With Cluster Resource Sharing, Nodes Can Go from General Purpose / Multi-Function to Single Function for Lowest TCO

Non-Coherent Global Shared Memory Empowers Mid to Large Size Clusters with Support for Single Function Micro-Nodes and the Ability to Forgo Node Over-Provisioning Typical of Message-Passing-Based Systems



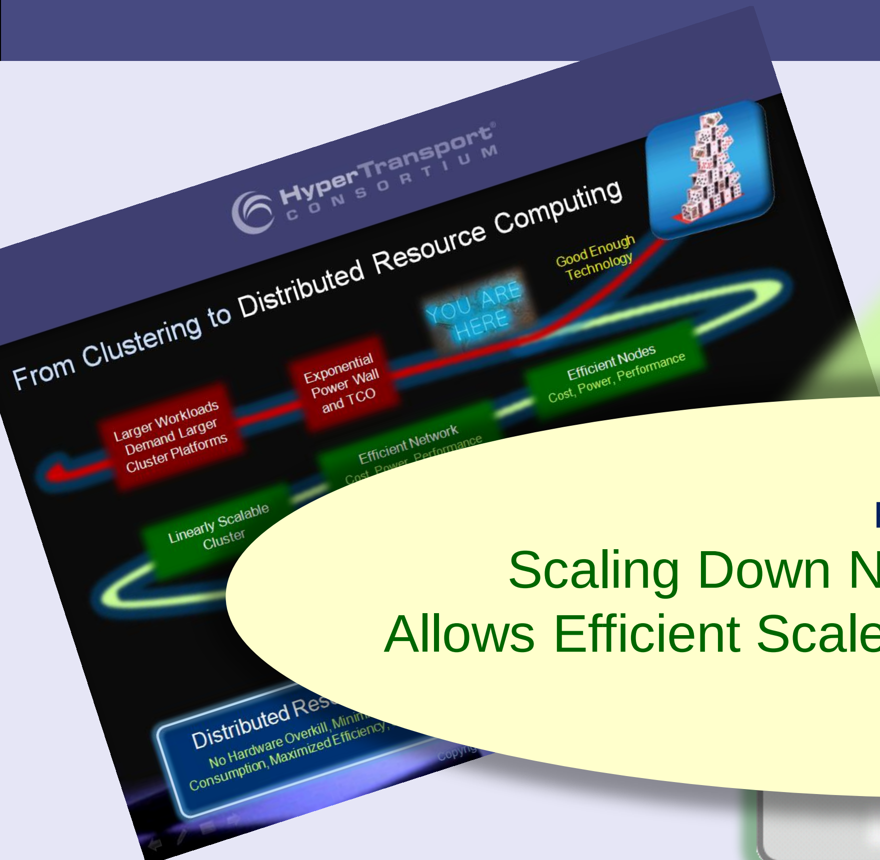
From Clustering to Distributed Resource Computing (cont.)



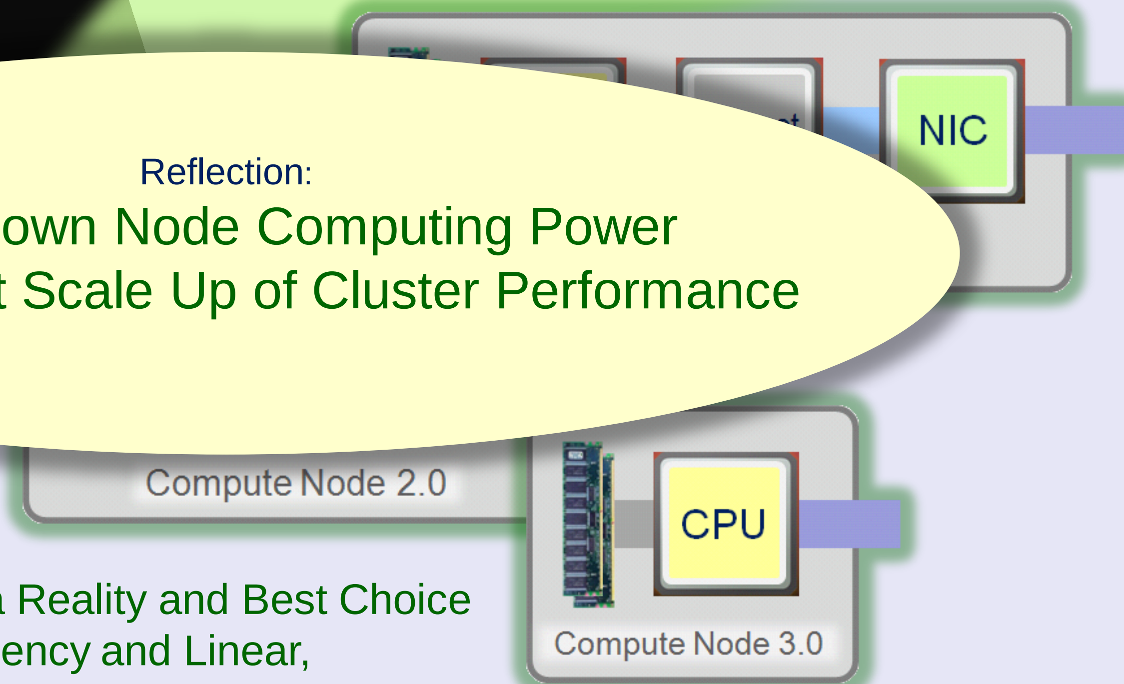


Single Socket CPU Micro-Nodes a Reality and Best Choice for Minimized Hardware, Top Efficiency and Linear, Granular Scaling of Computing Power

From Clustering to Distributed Resource Computing (cont.)



From General-Purpose Node to Micro-Node
CPU, Acceleration, Memory, Storage,
Special Functions Shared Between Nodes

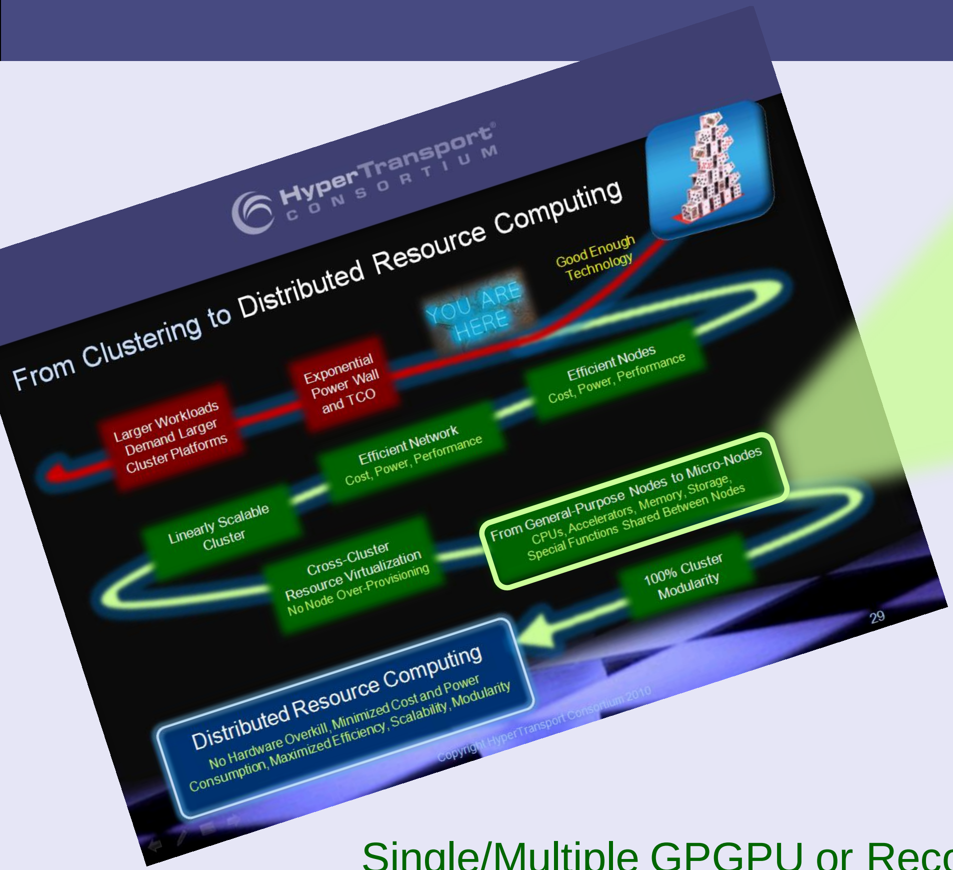


Reflection:

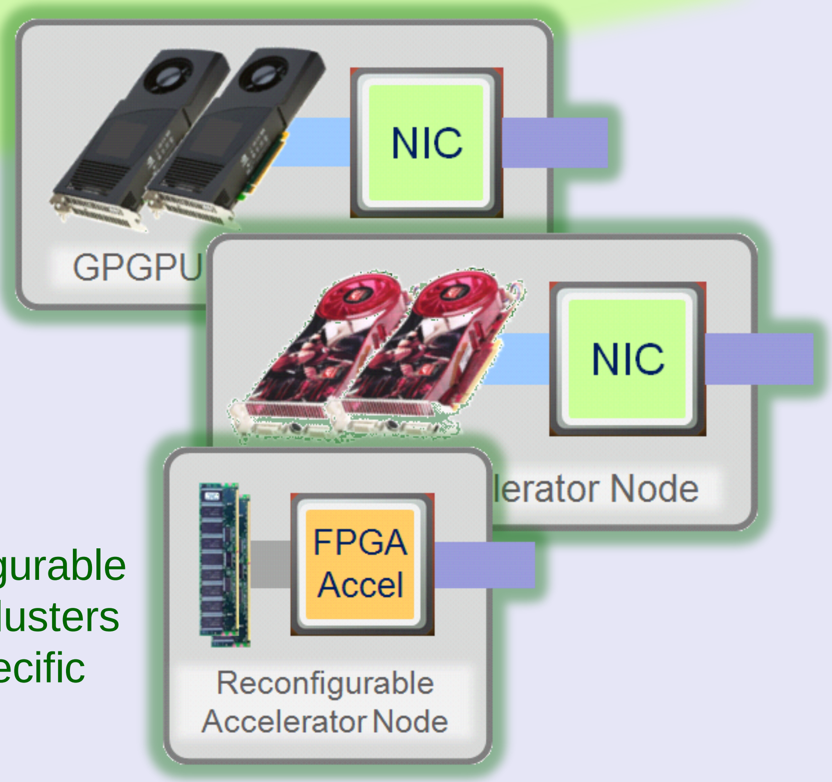
Scaling Down Node Computing Power Allows Efficient Scale Up of Cluster Performance

Single Socket CPU Micro-Nodes a Reality and Best Choice for Minimized Hardware, Top Efficiency and Linear, Granular Scaling of Computing Power

From Clustering to Distributed Resource Computing (cont.)



From General-Purpose Node to Micro-Node
CPU, Acceleration, Memory, Storage,
Special Functions Shared Between Nodes

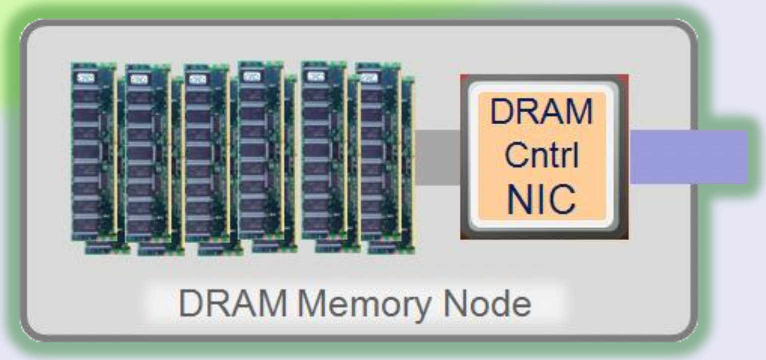


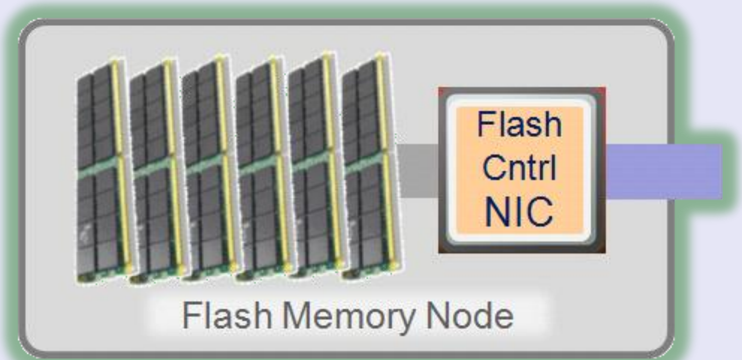
Single/Multiple GPGPU or Reconfigurable Acceleration Nodes Make Large Clusters

Migration from General Purpose to Application-Specific with Off-the-Shelf Technology a Reality

From Clustering to Distributed Resource Computing (cont.)

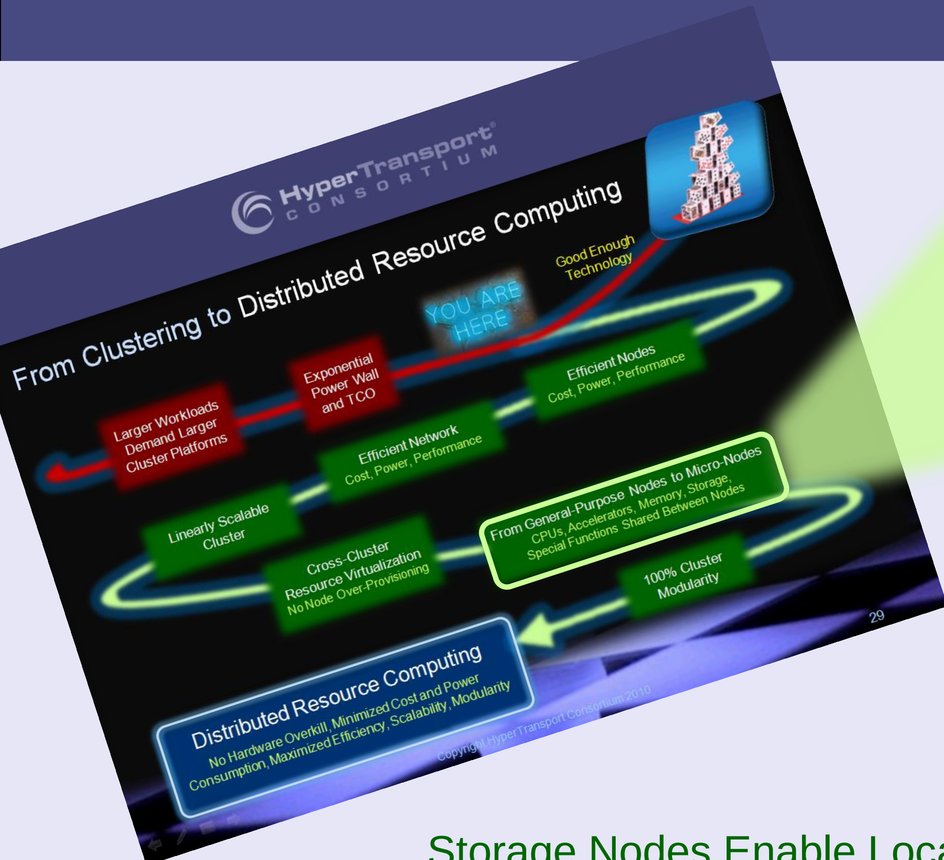
From General-Purpose Node to Micro-Node
CPU, Acceleration, Memory, Storage,
Special Functions Shared Between Nodes



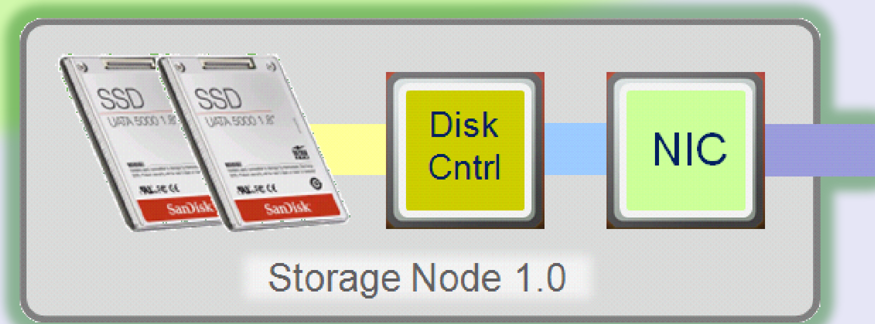


Memory Nodes Empower
Data Base, Data-Intensive Analytics and Informatics Processing with Highly Sought Low Latency Storage Subsystems

From Clustering to Distributed Resource Computing (cont.)

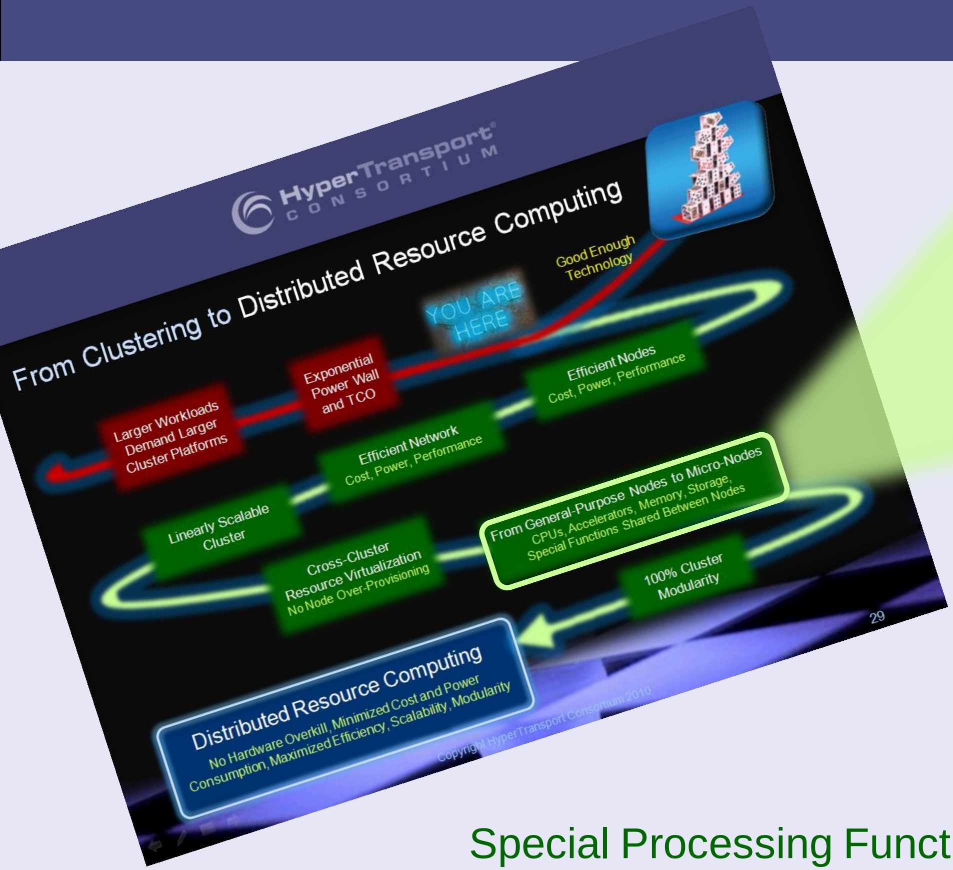


From General-Purpose Node to Micro-Node
CPU, Acceleration, Memory, Storage,
Special Functions Shared Between Nodes

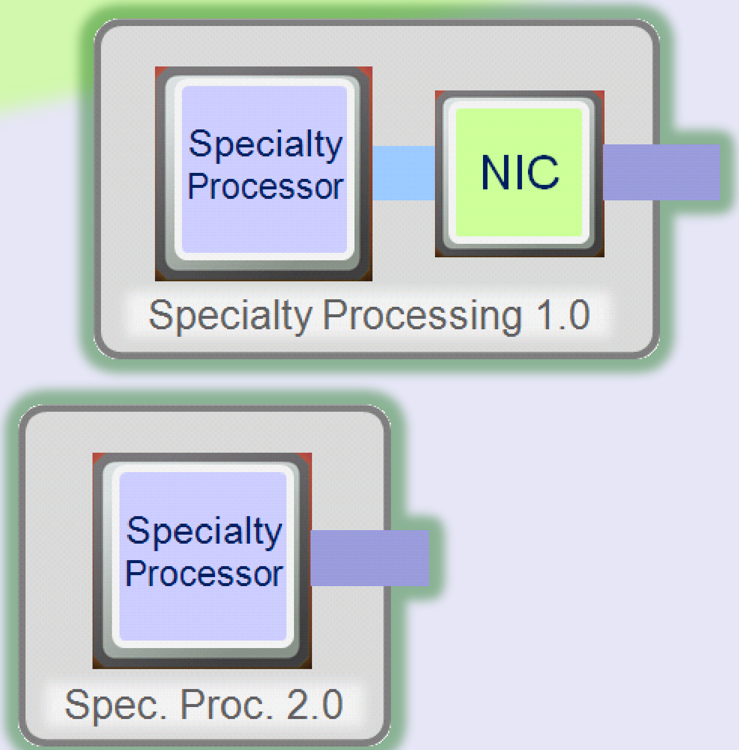


Storage Nodes Enable Localized
Back-Up and Permanent Data Storage
Subsystems to Serve the Entire Cluster

From Clustering to Distributed Resource Computing (cont.)

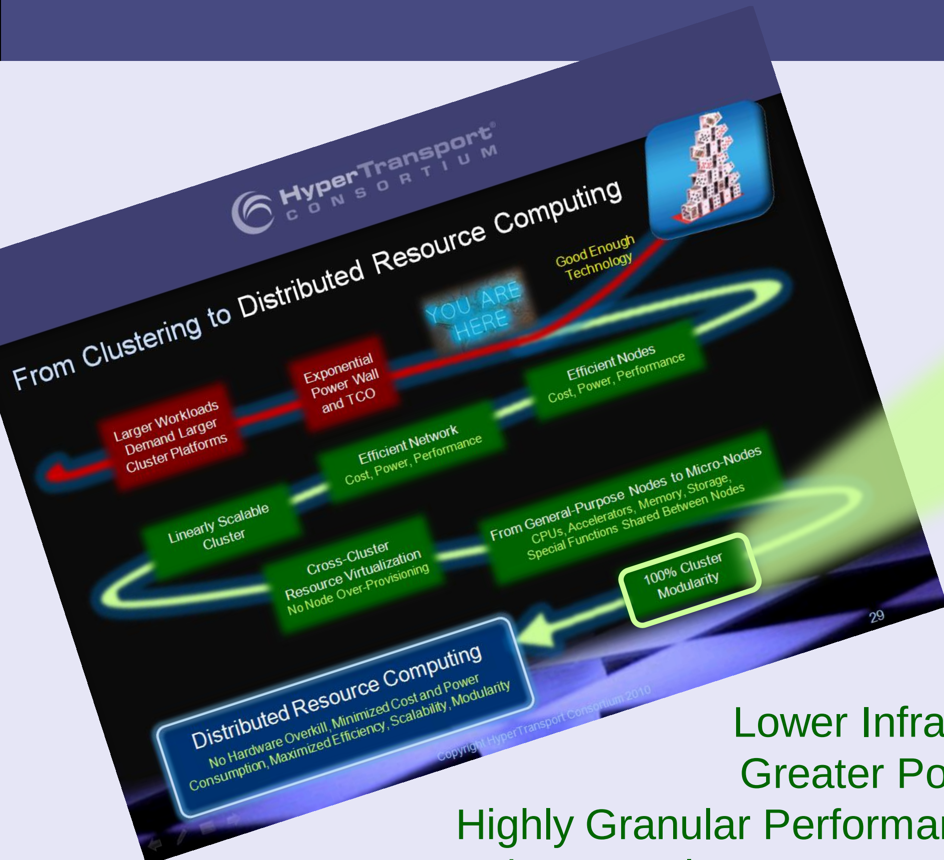


From General-Purpose Node to Micro-Node
CPU, Acceleration, Memory, Storage,
Special Functions Shared Between Nodes



Special Processing Functions,
e.g. Deep Packet Inspection, Security, etc.,
can be Cluster-Centralized to Serve All
Nodes in the Cluster

From Clustering to Distributed Resource Computing (cont.)

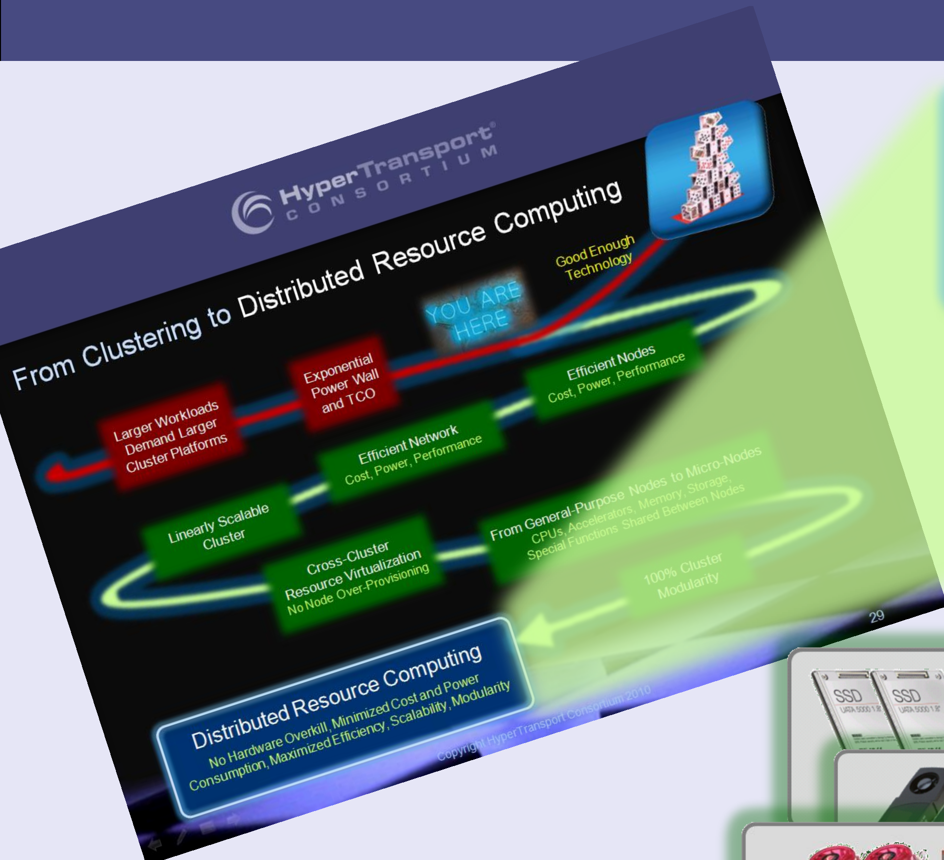


100% Cluster Modularity

Leaner, Modular Nodes Deliver Functional and Cost Breakthroughs to Large Scale Clusters

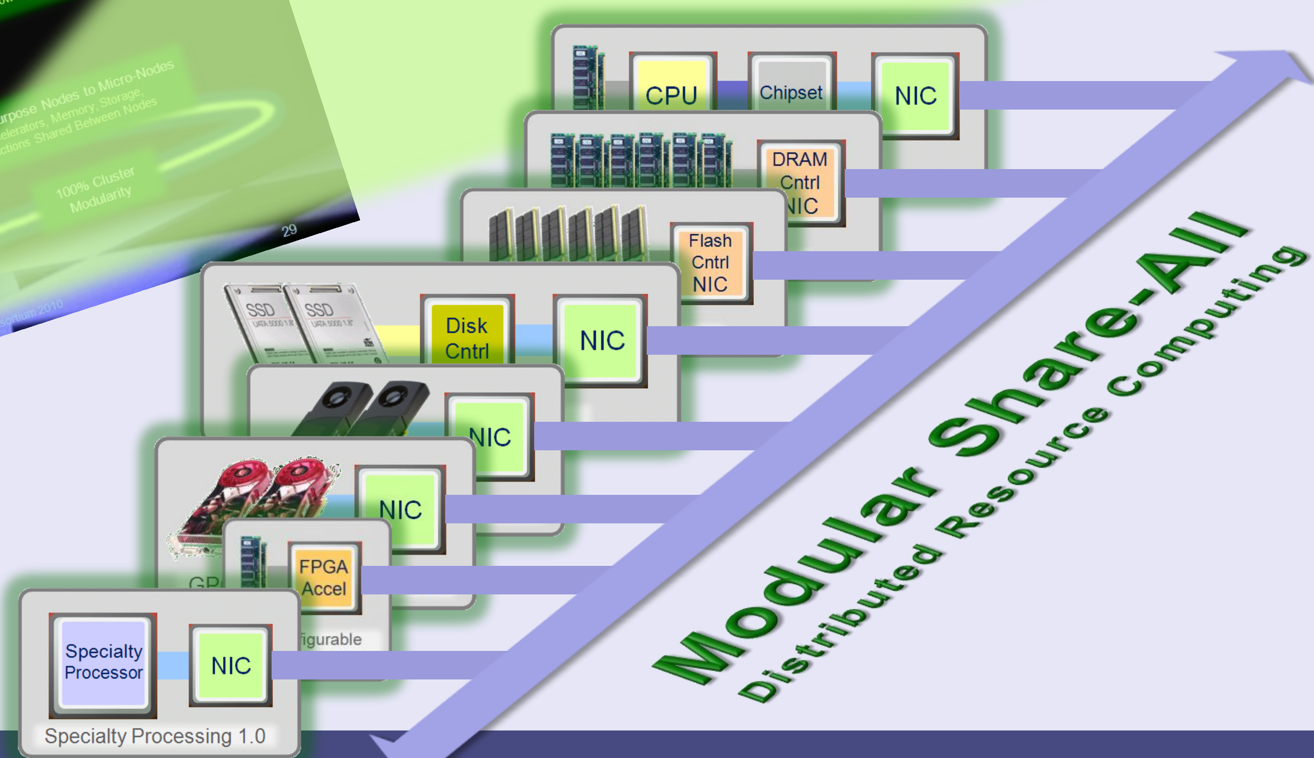
Lower Infrastructure Cost
Greater Power Efficiency
Highly Granular Performance Scalability
Easier to Implement, Less Costly Cooling
Greater Fault-Tolerance
Mission-Critical “Always Up” Capability
Easier System Reconfiguring / Upgrading
Facilitated Servicing and Maintenance

From Clustering to Distributed Resource Computing (cont.)



Distributed Resource Computing

No Hardware Overkill, Minimized Cost and Power Consumption, Maximized Efficiency, Scalability, Modularity



Ultimate Target for Scale-Up Performance and Scaled Down TCO

HyperShare™

Enables the Ultimate

Distributed Resource Computing