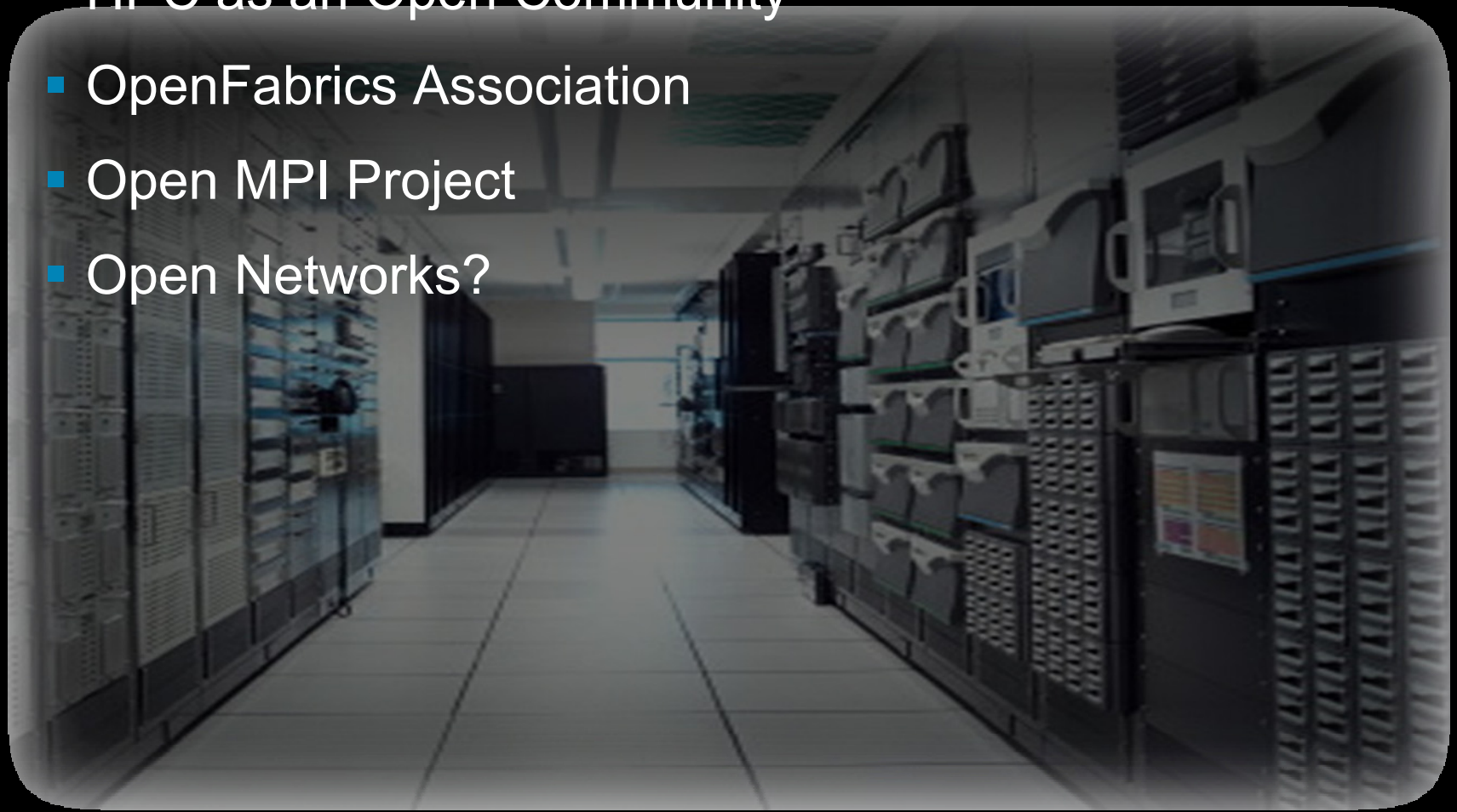
Open Architectures for High Performance Computing

Robert Starmer

Cisco Systems

Overview

- HPC as an Open Community
- OpenFabrics Association
- Open MPI Project
- Open Networks?



High Performance Computing

Two modes of delivering performance

- Tightly coupled applications
  - Latency matters
  - Bisectional Bandwidth matters
  - Networks replace backplanes
  - MPP or Clusters
- Loosely coupled applications
  - Bandwidth can be important
  - Scale is important
  - Clusters rule

HPC Evolution: Late 80's

Parallel HPC	Fast Vector HPC	
Massively Parallel Processors (MPP's) Government users	Now the "classic" HPC Supercomputer Government users	

HPC Evolution: Early 90's

Parallel is the future		
Massively Parallel Processors (MPP's) Government users A few small Beowulf clusters (4-16 nodes) Academic users		

HPC Evolution: Late 90's

Leading Edge HPC	Mid-Range HPC	
Many Beowulf clusters (hundreds nodes), high-speed proprietary networks Government, research, and academic users	Small Beowulf clusters (4-16 nodes) Leading-edge industry and academic users	

HPC Evolution: Early 00's

Leading Edge HPC	Mid-Range HPC	Commodity HPC
Complex Linux clusters (hundreds of nodes), high speed networks Dedicated HPC architectures Government, research, and academic users	Simple Linux clusters (tens of nodes), high speed networks Leading-edge industry and acedemic users	Small Multi-processor machines (SMPs) Many industry users

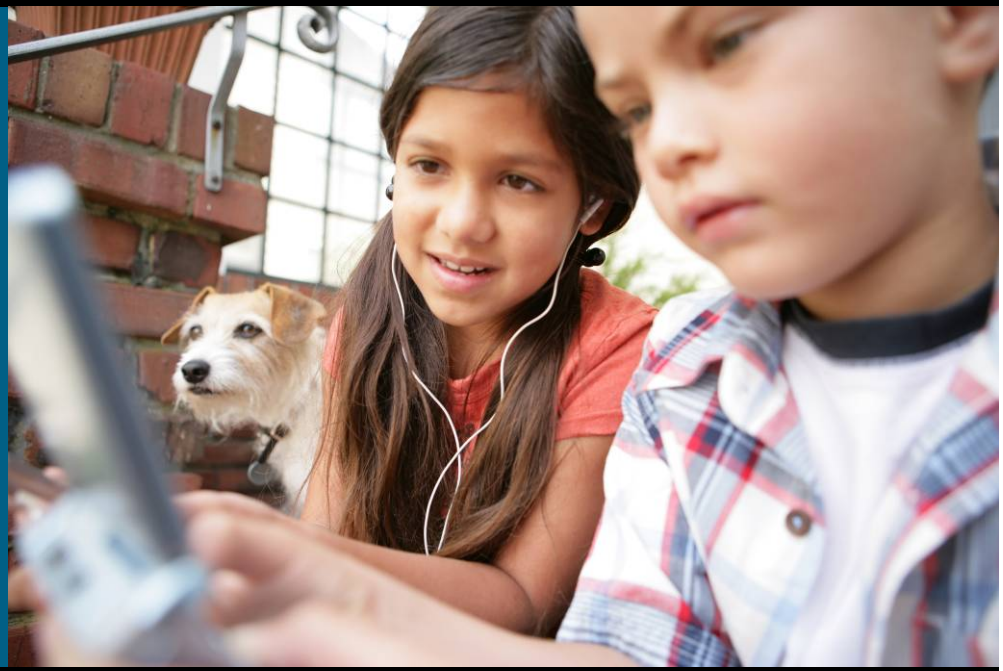
HPC Evolution: Late 00's

Leading Edge HPC	Mid-Range HPC	Commodity HPC
Complex Linux clusters of multi-cores (thousands of nodes, tens of thousands of processors), complex high speed networks Dedicated HPC architectures Government, research, and academic users	Linux clusters of multi-cores (hundreds of nodes, thousands of processors), high speed networks Leading-edge industry and acedemic users	Linux clusters of multi-cores (tens of nodes), high-speed networks Many industry users Growing vendor market

In search of the Trickle-Down Effect

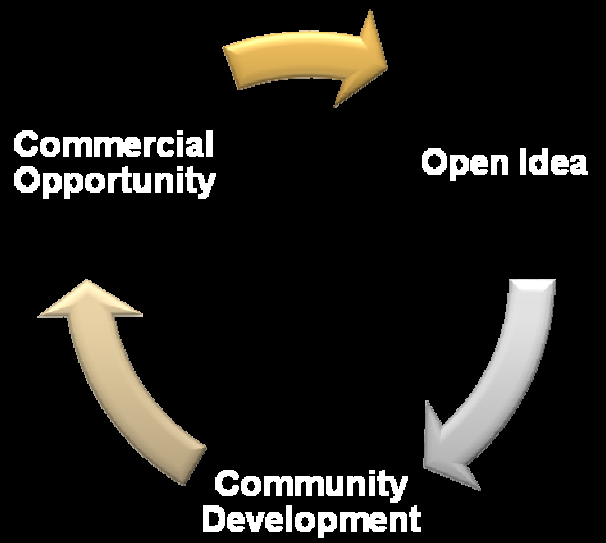
- Leading edge – cutting edge early adopters
  - Want absolute best performance
    - …but willing to pay the price for it
    - Monetary premium for newest technologies
    - Hardware and software immaturity / instability
- Research proposals written in anticipation of new technologies (!)
- Takes years (if ever) for these technologies to go from “Leading Edge” to “Commodity”
  - *Open source architectures are changing this paradigm*

Open Source



Why Open Source?

- Open source in HPC helps everyone
  - Plenty of room for research / new ideas
  - Open environment allows access and information transfer
  - Feed them back into production / commodity products
- Shorten the cycle from research to commodity
- Researchers have the ideas -- industry has the production capability to bring it to the masses
  - There are smart people in both!



Case Studies: OpenFabrics, Open MPI

- Software stacks used in HPC
  - OpenFabrics: network driver stack and tools (initially InfiniBand)
  - Open MPI: Portable high performance MPI-2 implementation
- Changing the paradigm
  - Both packages used in leading edge, mid-range, and commodity HPC environments
- Dramatically shortening the time required from “idea” to “product”
  - Ideas are good; practical applications are great
  - How long before your Grandmother is using what used to be closeted acedemic compute architectures?

The OpenFabrics Alliance

- Not-for-profit organization to advance RDMA technologies
  - 26 industry members
  - 8 academic / research members
- Targeting all areas of high-speed networking
  - HPC community
  - High-capacity storage
  - Datacenter applications
- Active community
  - Collaborating on code base
  - Publishing research papers (still many unanswered questions!)

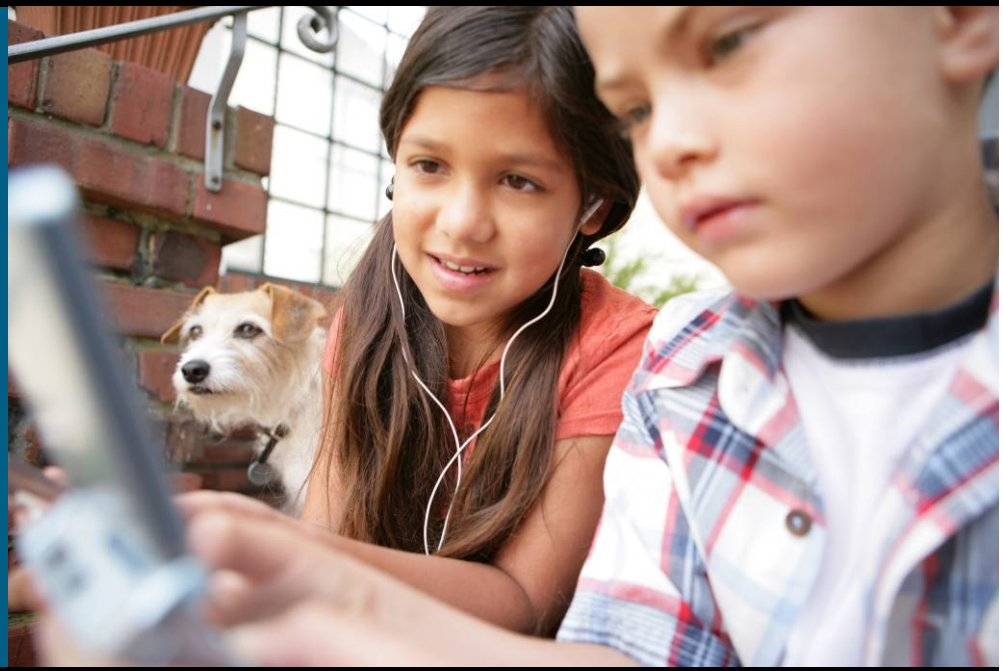
OpenFabrics Enterprise Distribution

- Release vehicle for OpenFabrics code base
  - Soon to release v1.3
- Community release process
  - QA-checked / certified by industry partners
  - Paid support options available
- Now a mature, stable, and scalable software project
  - New features continually being developed
  - Influenced by the research community

Open MPI

- Originated as research / academic project
  - Consolidate best ideas from several prior MPI implementations
  - Provide unified user experience across different platforms
  - Engage the HPC community for MPI expertise
- Cisco was first corporate member
  - Now has 14 voting members, 9 non-voting contributors
  - 10 industry members
  - 13 academic / research members
- Active community
  - Collaborating on code base
  - Publishing research papers (still many unanswered questions!)

Open Network?



HPC Evolution: Late 80's

Parallel HPC	Fast Vector HPC	
Massively Parallel Proces Goverr	What people often ear r" Government users	

Proprietary Networks
Proprietary Backplanes

HPC Evolution: Early 90's

Parallel is the future	Still kicking	
Massively Parallel Proce	Now the "classic" computer	
Government users	Government users	
A few small Beowulf		
Academic users		

Proprietary Networks

Standards Based Networks

HPC Evolution: Late 90's

Leading Edge HPC	Mid-Range HPC	
Many Beowulf nodes), high-speed proprietary networks Government, research, and academic users	Small Beowulf lusters (4-16 nodes) industry and academic users	

Proprietary Networks

Standards Based Networks

HPC Evolution: Early 00's

Leading Edge HPC	Mid-Range HPC	Commodity HPC
Complex Linux nodes), high speed networks Dedicated HPC architectures Government, research, and academic users	Simple Linux clusters ens of nodes), high spee Leading-edge industry	Small Multi-processor machines (SMPs) Many industry users

Proprietary Networks

Standards Based Networks

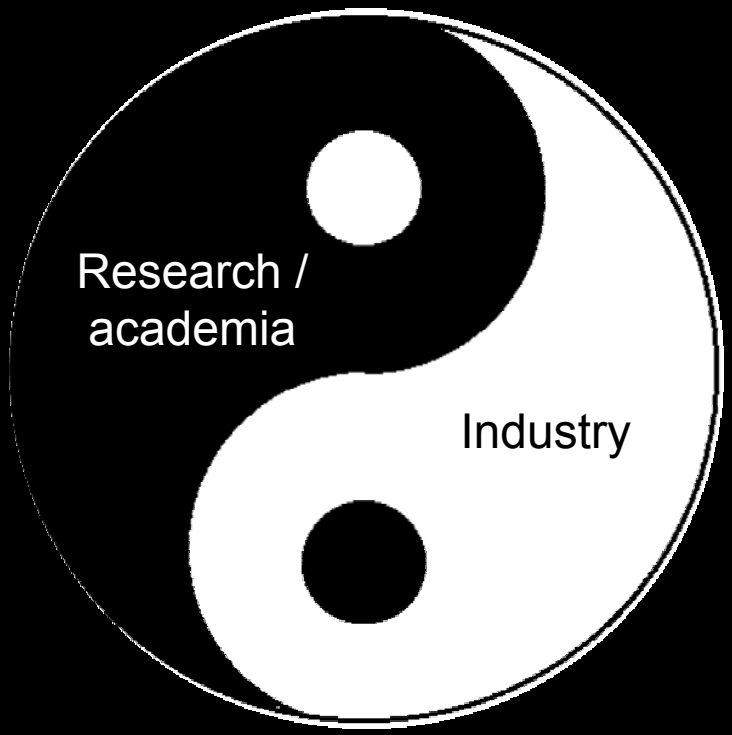
HPC Evolution: Late 00's

Leading Edge HPC	Mid-Range HPC	Commodity HPC
Complex Linux clusters of multi-cores (thou tens of thousands of processors), complex high speed networks Dedicated HPC architectures Government, research, and academic users	Linux clusters of multi-cores (hundreds of processors), high speed networks Leading-edge industry	Linux clusters of multi-cores (tens of d networks Many industry users Growing vendor market

Standards Based Networks

What We Have Learned

- HPC knowledge is not centralized
  - Get everyone who cares to participate
  - Different viewpoints and agendas are good!
- Ensure that everyone can "win" in some way
  - Research papers, products, PR, etc.
- Open source is not free
  - Someone pays for humans and equipment
  - Cannot be completely self-serving



What We Have Learned

- Consensus is good
  - But not always possible
- Open sources does not guarantee quality
  - We pay attention to what we care about
  - Careful balance between strict QA controls and independent academic freedom
- Open source is like a marriage
  - It takes a lot of work
  - Must find a good balance
  - But the end result is great