Using Infiniband as a High Speed LAN Interconnect

NATIONAL CENTER
FOR COMPUTATIONAL SCIENCES

presented by

Makia Minich

Oak Ridge National Laboratory
U.S. Department of Energy

Summary

- What are we trying to accomplish?
- Why choose Infiniband?
- Network Summary
- Current Uses
- Proposed Uses
- Problems We've Seen
- Needs

NCCS Network Roadmap

- Shifting to a hybrid InfiniBand/Ethernet network
- InfiniBand based network helps meet the bandwidth and scaling needs for the center
- Wide-Area network will scale to meet user demand using currently deployed routers and switches

1000 TF
200 GB/s LAN
4 GB/s WAN

2009

250 TF
200 GB/s LAN
3 GB/s WAN

2008

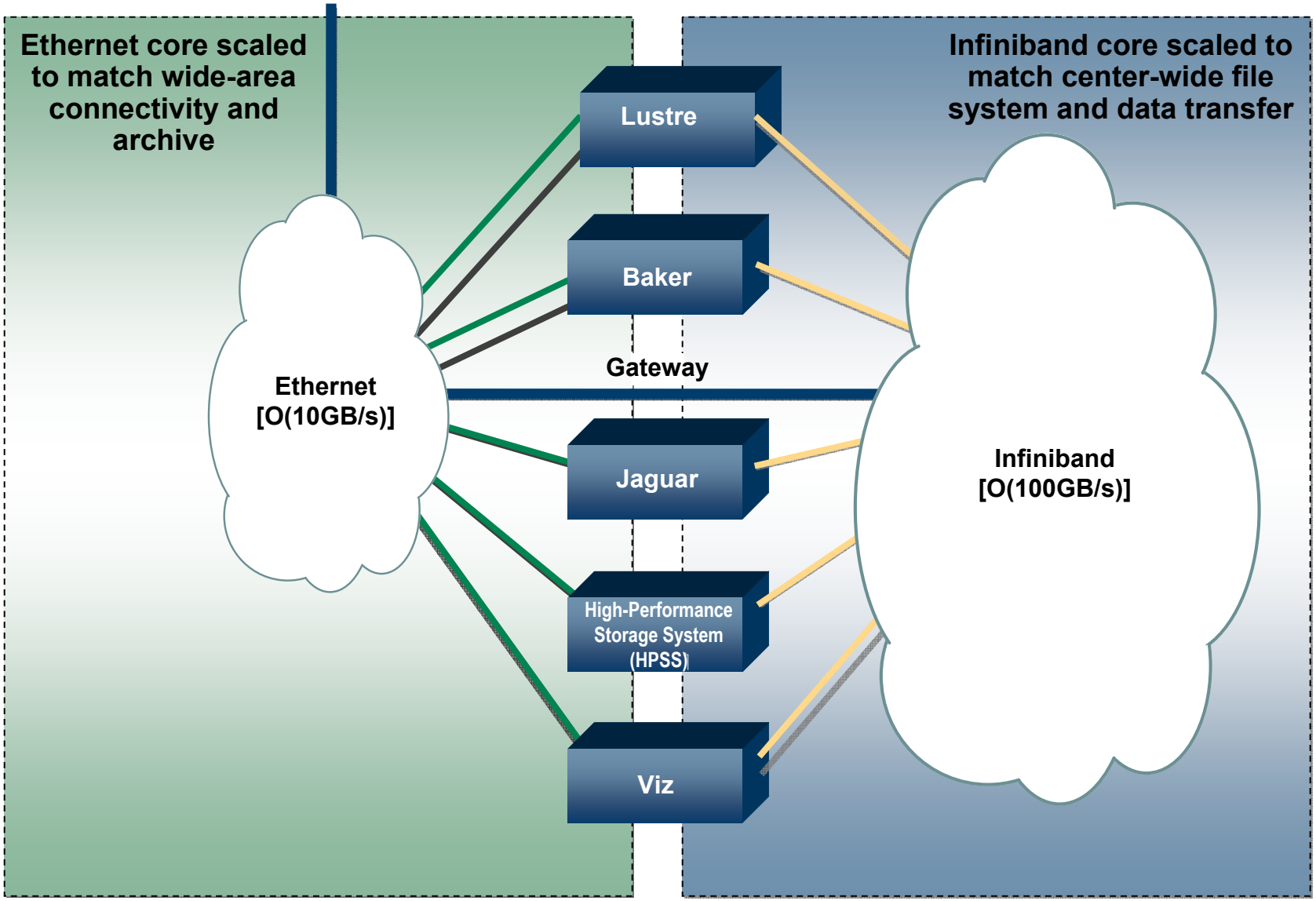
100 TF
60 GB/s LAN
3 GB/s WAN

2007

Why Infiniband

- Cost effective for needed bandwidth
- Ongoing stack development
- Better defined future roadmap (is ethernet going 40G or 100G next?)
- Scalability

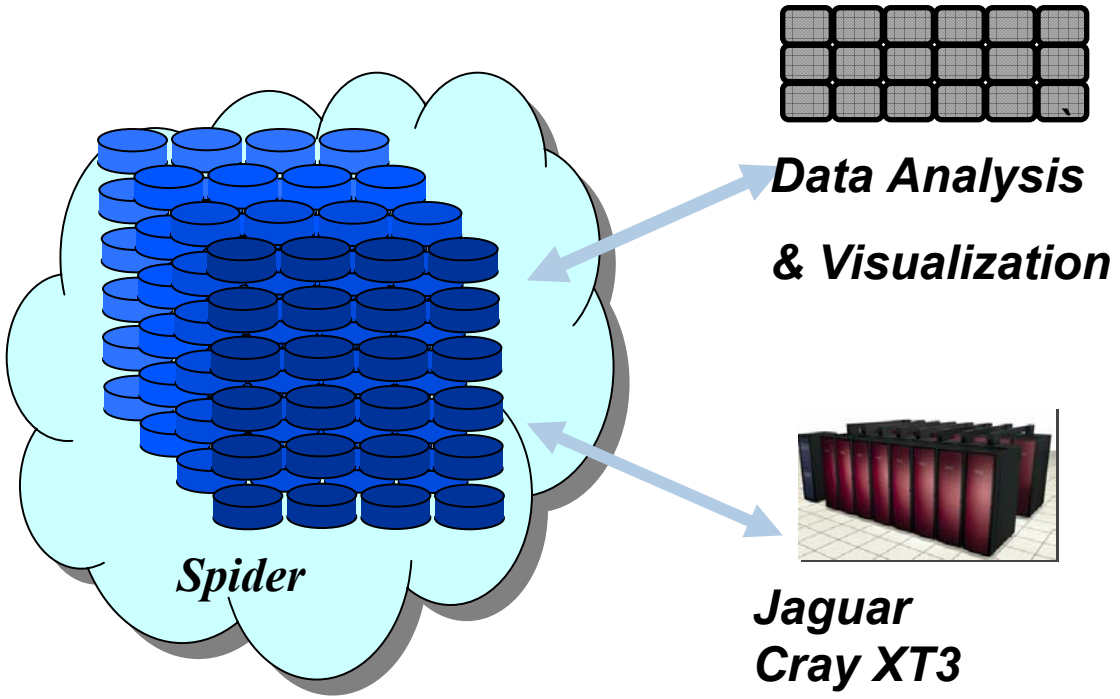
Network Summary



Current Uses

Lustre Results:

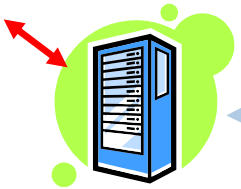
- Single Client – 800MB/s
- Multiple Clients – 500-600MB/s per router



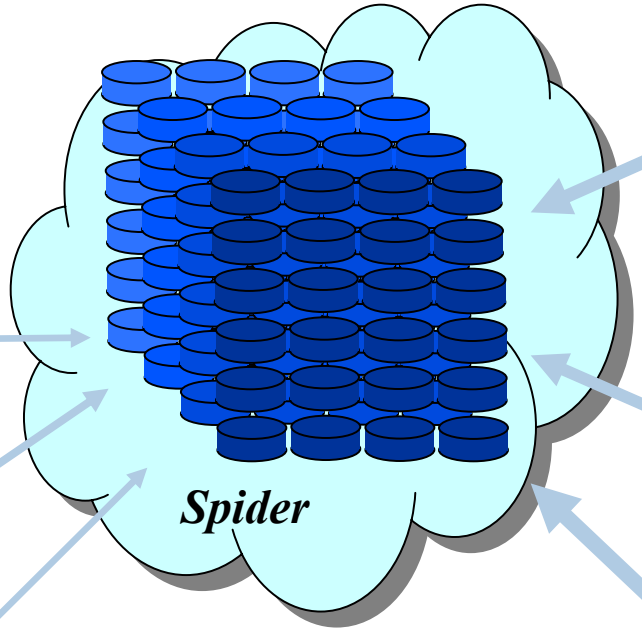
2007
- 1 PB
- 30 GB/s (aggregate)

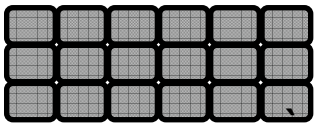
Proposed Uses

Phoenix
Cray X1E



NFS Servers

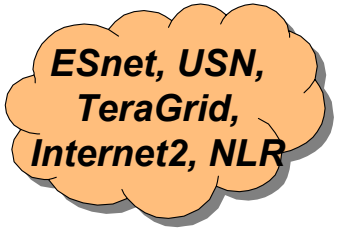




Data Analysis
& Visualization

Jaguar
Cray XT3

HPSS



2007
- 1 PB
- 30 GB/s (aggregate)

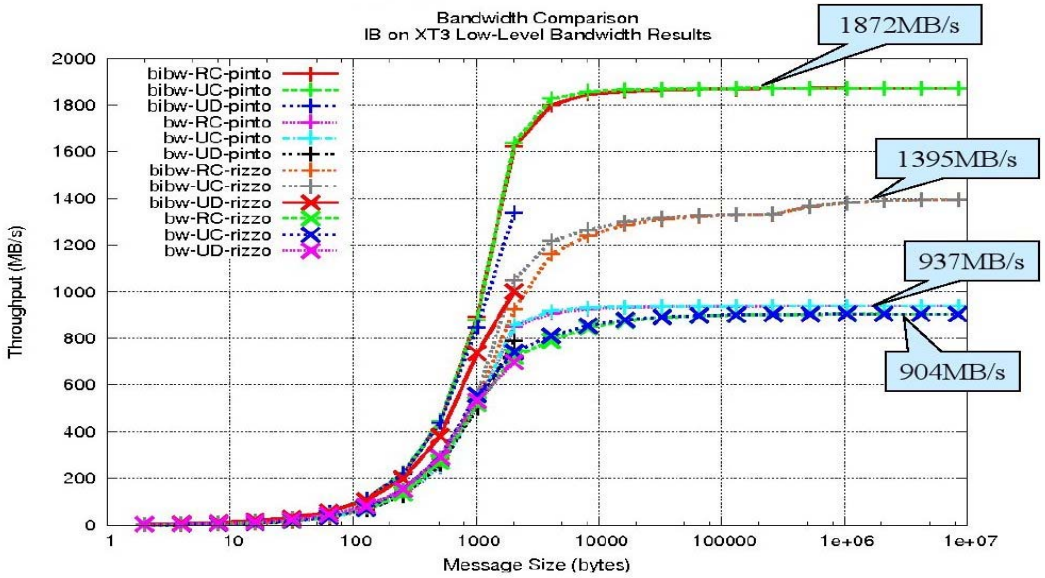
2008
- 10 PB
- 200 GB/s (aggregate)

Baker

Problems We've Seen

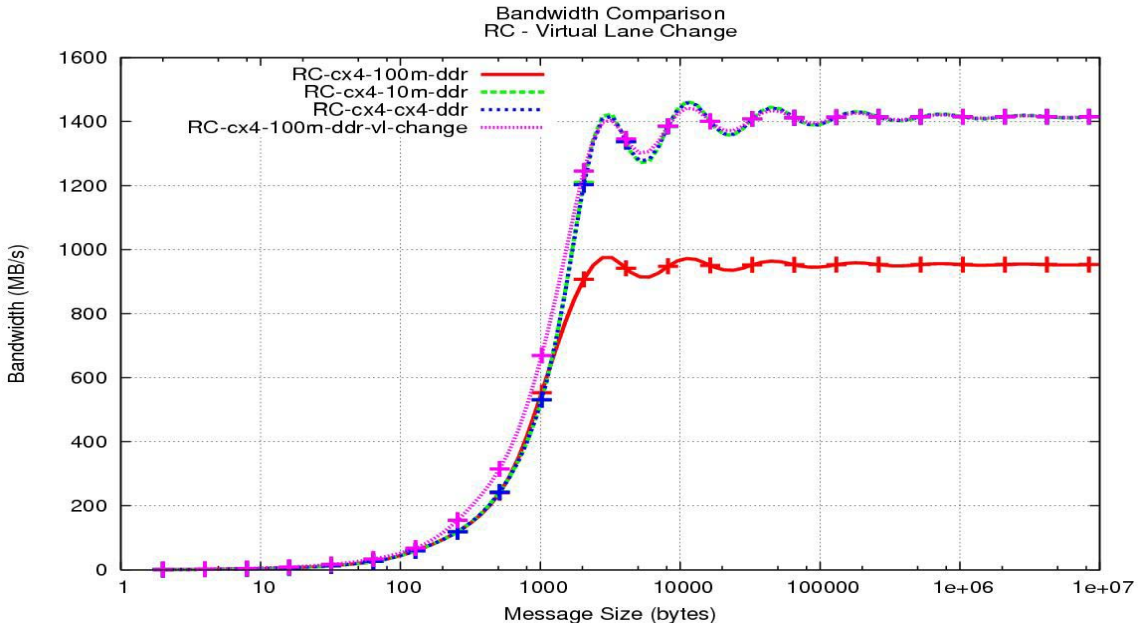
IB on the XT3/XT4

- Needed to get the IB stack to load on the XT3
  - Source code issues
- Performance Issues
- Lack of system-based debugging



Distance

- Two story machine room
- Longest distance around 60m
- Did extensive testing with Intel Connects Cables



Subnet Issues

- Mixed IB network (various versions)
  - MAD Storm
- Link speed negotiation
  - firmware (switch and hca)
  - cables
- Virtual lane resource handling
  - cable distances

Needs (Existing or Otherwise)

- Functioning QoS
- VLAN-like Subnetting (to allow for network segregation and security administration)
- Subnet Routing
- IPoIB to 10GbE Routing
- Traffic pattern inspection/Utilization Information
  - How much data over each link
  - Destination of that data

Questions/Links?

- Makia Minich (minich@ornl.gov)
- National Center for Computational Sciences - http://www.nccs.gov
- Jobs - http://jobs.ornl.gov