

InfiniBand for the enterprise



OPENFABRICS
ALLIANCE

InfiniBand for the enterprise, a customer's perspective.
Eric Ulmer, Accenture

Agenda



- Company Overview
- Our Platform/Environment
- Our Rewards in using the OpenFabrics Software
- Wish List
- Q&A (Will try and maximize allotted time for this portion)

Accenture Software Utility Services



➤ We are an Applications Service Provider

▪ Example customers:

- Navitaire
 - Reservations, Crew Scheduling, Online Bookings, Customer Relationship Management, Route Solver
- Insurance packages
- Telecommunication company software packages
 - Mobile telephone number portability
- Utility company management software packages
 - Natural Gas and Electric trouble ticket management
- Custom stock market analysis and trading packages

Accenture Software Utility Services



Accenture Software Utility Services



➤ 8 Data Centers

- 2 in United States
- 3 in the United Kingdom
- 3 in Australia
- 400+ UNIX/Linux Servers
- 2000+ Windows Instances
- Heavy I/O Microsoft SQL shop (350+ instances)
- 1.6PB of disk total.
- 100% Network Appliance primary and nearline disk (54 filers)
 - NFS for VMWARE and UNIX databases, iSCSI and FCP for Windows
- Data Domain being used for VTL
- HP, SUN and IBM servers

Navitaire



- High Transactional Volume
- Total value of web-based travel transactions exceeds Amazon.com transactions
- Daily credit authorizations reach \$135M
- Execute 750,000 batch jobs monthly
- High Service level Agreement penalties:
 - Outages
 - Application response time
 - We have no scheduled maintenance windows. We pay to take Airline customers down.

NAVITAIRE \$11.6B
amazon.com® \$10.7B
2006



Our Platform(s)

- Traditional HP 5XX & 3XX Servers (Web/App/DB)
 - Gig-E/Mostly iSCSI/4-16CPU/16-128GB RAM
 - Windows 2003 Std and Enterprise Edition; RHEL 3/4

- IBM X460 and X3950(DB)
 - Windows 2003 Enterprise Edition
 - Emulex FCP/Teamed Gig-E/16-32 core/128GB RAM

- SUN Sparc (Misc)
 - Gig-E and 100Mb ethernet connected. Various generations of models; Solaris 8-10

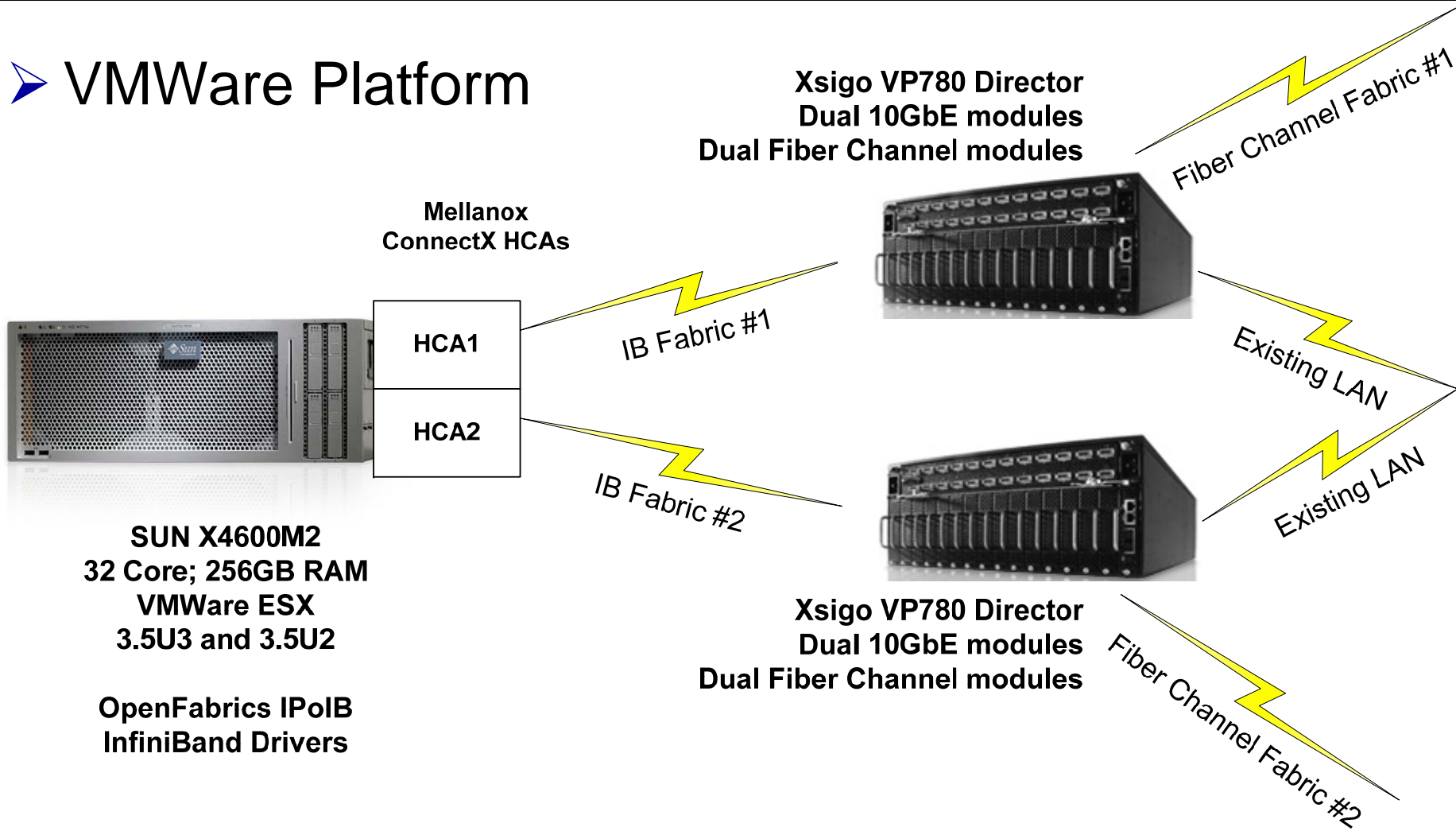
- SUN X86 4600M2 (VMWare)
 - ConnectX - DDR; VMWare ESX 3.5U3; 32 Cores; 256GB Ram each
 - Xsigo VNIC driver on top of OF driver stack.

Our Platforms

- Cisco Ethernet Switching and Routing equipment
 - Primarily GigE to servers; 10G for trunks and Filer connectivity
 - Dual Switches; HSRP (Full redundancy)
 - Servers patched to core via patch panels
- Brocade and Cisco fiber switches
 - Dual fabrics; models vary by data center and generation.
- InfiniBand based LAN/SAN virtualization
 - Two separate IB fabrics and Xsigo director cores per data center
 - Used as our backbone for all VMWare machines
 - Our VMWare farms do not have ethernet connections except for remote consoles; 100% InfiniBand based VMWare deployment.
 - Each VMware server has one connection to each InfiniBand fabric via separate ConnectX HCAs.
 - Xsigo provides virtual network and fiber HBAs to the vmware servers over InfiniBand via their VP780 director modules.
 - We are using Mellanox Silicon with the OpenFabrics ESX drivers

Our Platforms

➤ VMWare Platform



Rewards using OF components

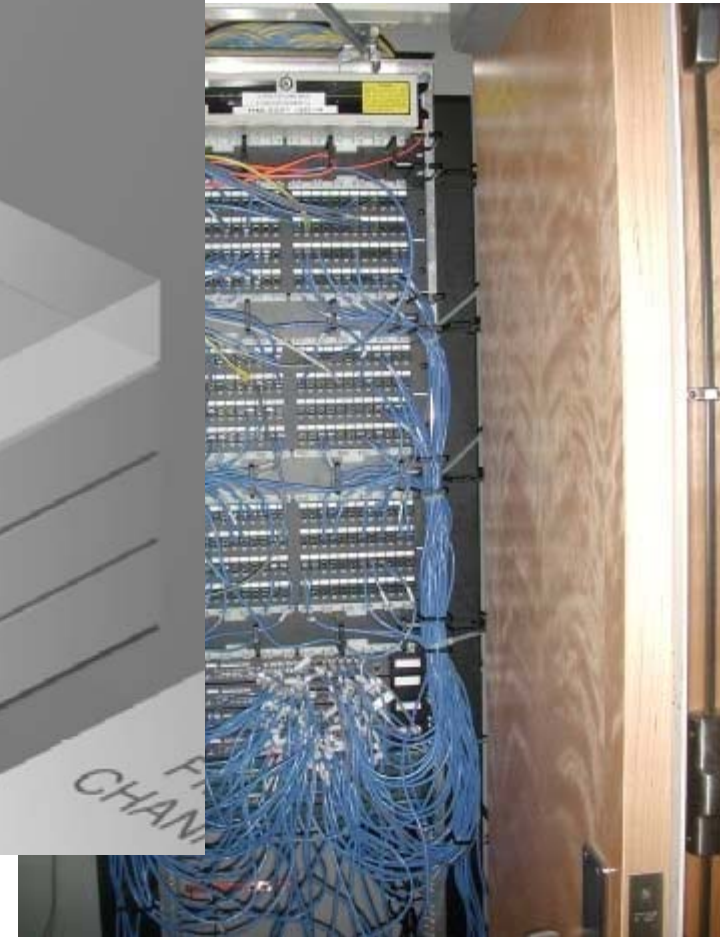
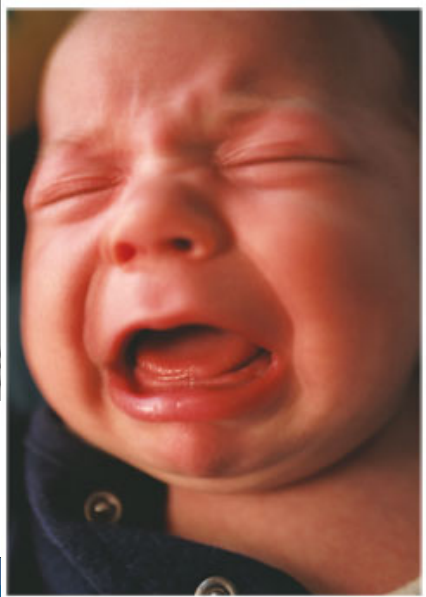
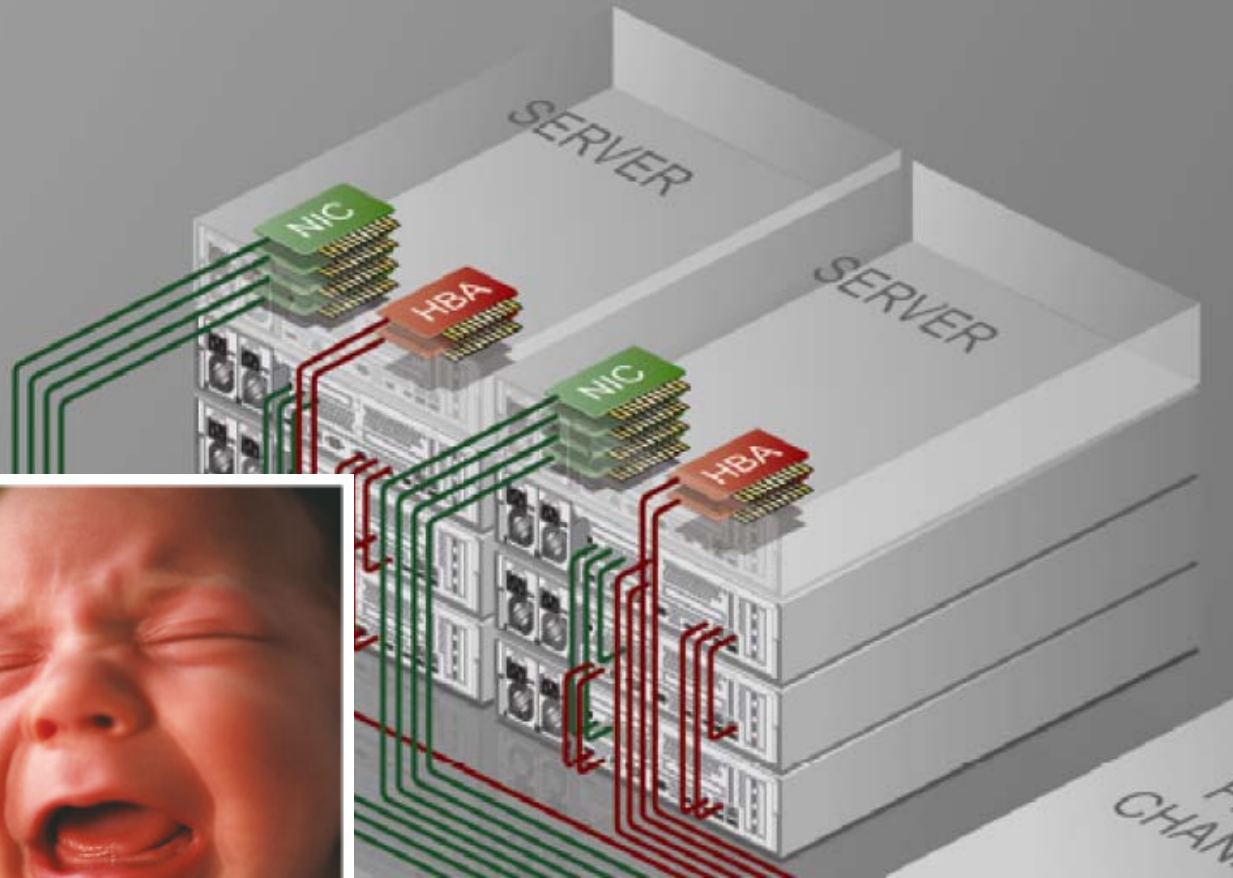
➤ Cost Savings!

- Neutral to our present 1GbE deployment with DDR(20g)
- Cost is less than 1GbE if using Single Data Rate InfiniBand(SDR IB) (10g)
- Cost 4-12X lower than 10GbE with Quad Data Rate InfiniBand(QDR IB) edge(40g) and DDR(20g) Core

➤ Reduced deployment time!

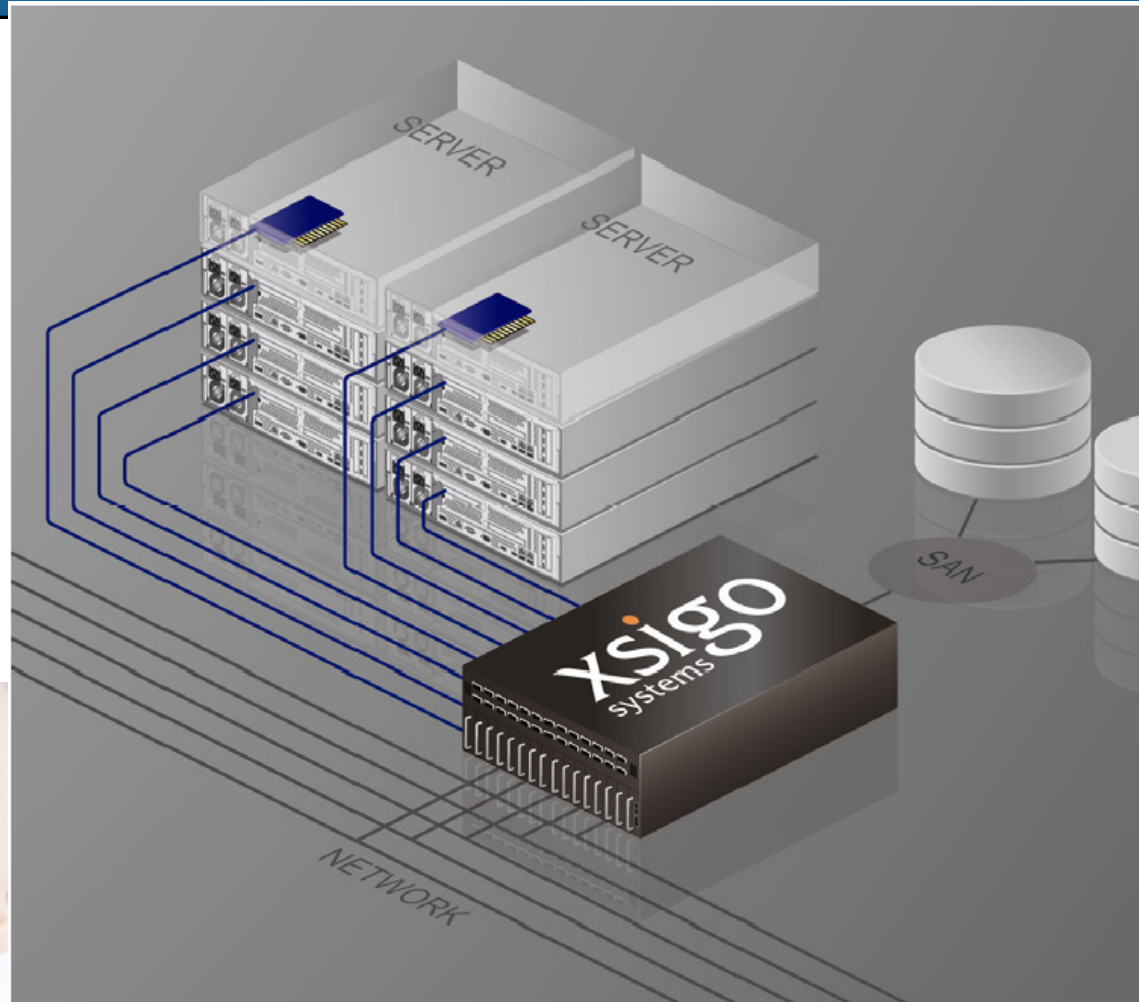
- Two quick connections to an IB edge switch –vs- multiple GigE and FC patch connections. (Not to mention the time and planning in installing the actual patch panels)
- Can deploy “data center in a carton”
- 2-4 hours vs potentially 2 months in some cases.

BEFORE



Picture not of actual Navitaire patch panel

AFTER



Rewards using OF components

➤ Eye to the future

- We are using Solid State disk storage in production in one of our data centers today
 - 8TB usable of Texas Memory Systems RamSan 500 storage front-ended by NetApp V6080 filer today.
 - We foresee a future without “LUNS” on servers
 - Why access RAM (Flash or otherwise) as a disk device?
 - Traditional Ethernet is obviously not the right choice for extended or RDMA memory access
 - InfiniBand and OpenFabrics are obvious existing (and more importantly, shipping) choices for transmitting at high data rates at the lowest cost.
 - Our goal is to be off spinning disk for all primary applications by 2012.

Rewards in using OF components



➤ Eye to the future

- We are testing communication from the application tier to database server tier using Microsoft winsock direct and IPoIB from the WinOF driver package.
- Our database servers have large RAM caches & they can bury a standard ethernet network queue. This causes a troublesome square toothed performance curve; unpredictable response times.
- WinOF provides:
 - **Lower latency**
 - Higher data rates (7.2Gb/sec SDR vs our current 420Mb/sec Quad GieG NIC Team)
 - We're excited about QDR; our purchases are road mapped.
 - **Lower CPU consumption** (2% vs 40% with NIC Team)
 - Lower cost
 - Lower complexity
 - A predictable sine wave shaped performance curve

Challenges (WinOF)

- Limited Windows HCL and “designed for” logo certification of drivers for enterprise editions of Windows. Very HPC centric in orientation.
 - If it’s not blessed by a Microsoft Logo it cannot be used in the Enterprise.
 - Enterprise customers have six to seven digit support contracts with Microsoft which are useless w/o Microsoft blessed software.
- Installation and documentation is centered on experienced computer users.
 - Needs to be GUIfied and dumbed down for Enterprise/End-User Windows admins

What's missing in the ecosphere?

- No existing mainstream enterprise storage vendor doing native IPoIB today.
 - Lots of niche HPC offerings; nothing for the enterprise
 - Should be top priority to get EMC, NetApp, SUN or HDS onboard with native IB connections.
- RDMA/Connected mode type connectivity is needed to unlock the performance potential of the solid state storage era.

Wish List for the ecosphere

- Windows Certifications for WinOF
 - Clustering catalog entries
 - Base installer and documentation to be more like other windows software installs (Next->Next->Next)
 - GUID->MAC mapping always auto-magic in WinOF
- Vendors need to make sure their products are submitted to Microsoft for inclusion in their catalog (WHQL)
- Enterprise Storage Vendor (we need one!)
 - One vendor is on the fence now
 - Contact your storage vendor and ask for InfiniBand
- Connected Mode or NFS/RDMA IO to VMWare
 - Customers contact your VMWare rep and ask for InfiniBand connectivity
- VLAN and Private VLAN like separation and security
 - Required to separate different tiers of the stack; multiple customer segmentation
 - Partition key management?

Q&A

Q&A