InfiniBand in the Lab

Erik Bussink

@ErikBussink

www.bussink.ch

www.VSAN.info

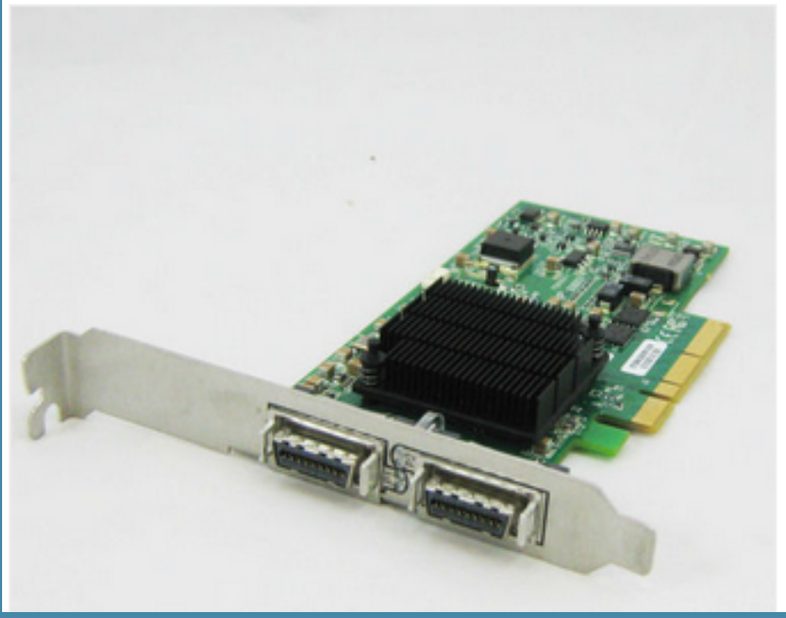
InfiniBand
an alternative
to 10Gb

Fast
Cheap

Price comparison on Adapters

Intel / X520-T2 10Gbp/s Dual Port
PCI-e Server Adapter E10G42BT

$349.00
or Best Offer



HP 452372-001 Infiniband PCI-E 4X DDR Dual Port Storage Host Channel Adapter HCA

Item condition: **Seller refurbished**

Quantity: 1 More than 10 available / 344 sold

Price: **GBP 32.90**
Approximately **US $52.47**

Buy another

Add to list

Shipping: **GBP 14.90 (approx. US $23.76)** Royal Mail International Signed-for |
See details
See details about international shipping here.
Item location: **Huntingdon, United Kingdom**
Ships to: **Europe, Asia, United States, Australia, Canada** See exclusions

Delivery: Estimated between **Thu. Oct. 17 and Wed. Oct. 23**
Seller ships within 1 day after receiving cleared payment.

Payments: PayPal | See details

Returns: 14 days, buyer pays return shipping | Read details

Seller informa
99.4% Positive feedbac
Save this seller
See other items
Visit store: Ch
Registered as a Busine

Connectors

- For SDR (10Gb/s) and DDR (20Gb/s) use the CX4 connectors

- For QDR (40Gb/s) and FDR (56Gb/s) use QSFP connectors

Connectors

Infiniband 10GBs 4X CX4 to CX4 Cable SAS M/M 0.5M LATCH Type DDR

Item condition: **New other (see details)**
"*No original packaging*"

Quantity: 1 More than 10 available / 10 sold

Price: **US $14.96**

Buy It Now

Add to cart

Best Offer:

Make Offer

Add to watch list

Add to collection

100% positive
Feedback

Best offer available

Shipping: See details
Item location: **Guang Zhou, Guang Dong, China**
Ships to: **United States, United Kingdom, Australia** See exclusions

Delivery: Varies

And switches too...

Example... 24 ports DDR (20Gb/s) for £212

Latest firmware (4.2.5) has embedded Subnet Manager

Qlogic SilverStorm 9024-DDR 24 Port 10Gb 20Gb InfiniBand Switch 9024S

Item condition: **Used**
"*WORKING ENVIRONMENT*"

Price: **US $349.99**
Approximately **£212.82**

Buy it now

Best Offer:

Make offer

8 watchers

Add to Watch list

Best Offer available

Collect 212 Nectar points
Get Started | Conditions

Postage: Will post to United Kingdom. Read item description or contact seller for postage options. | See details
Item location: **MONTREAL, Canada**
Posts to: **Worldwide**

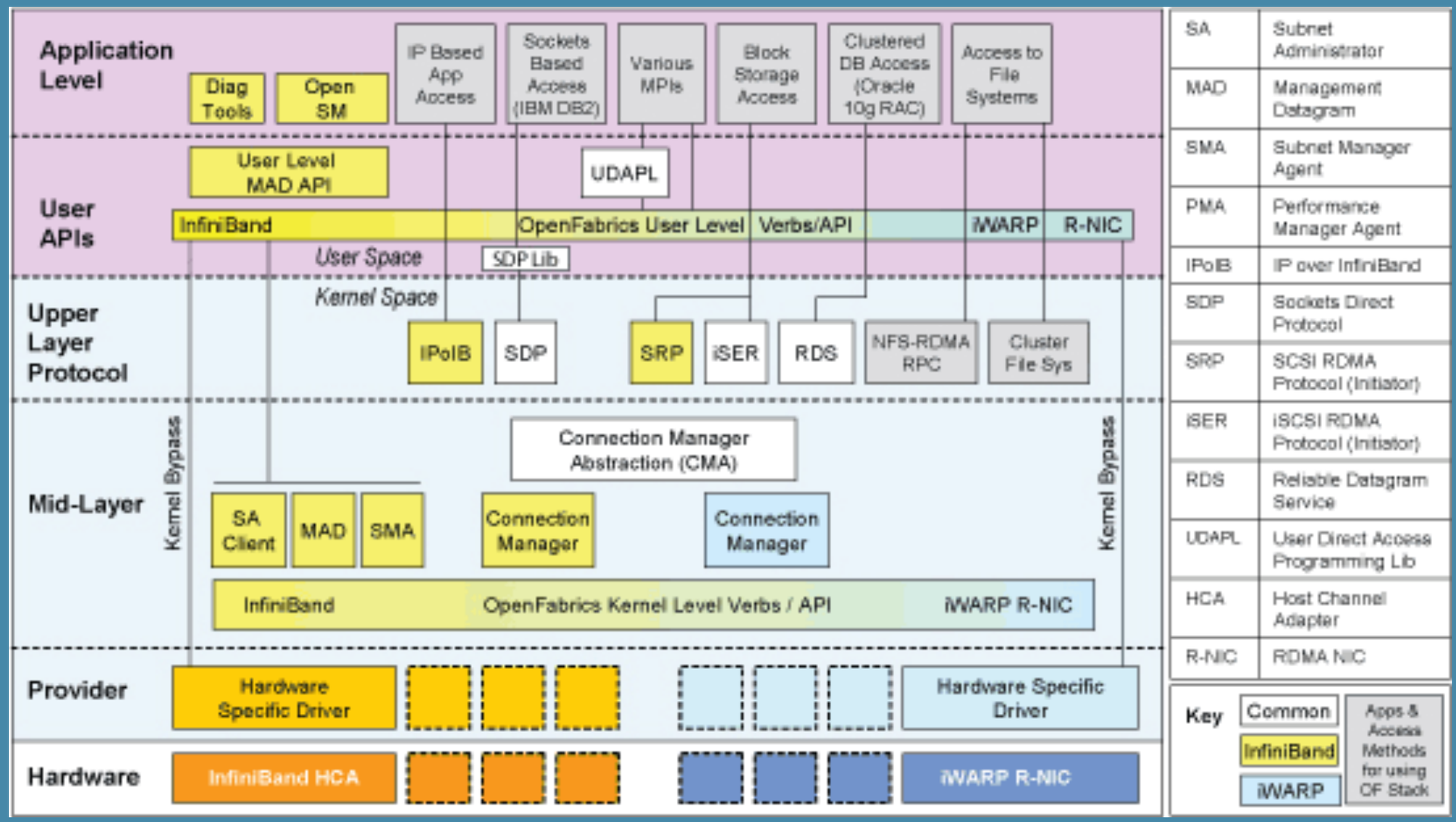
What InfiniBand Offers

- Lowest Layer for scalable IO interconnect
- High-Performance (SDR 10Gb, DDR 20Gb, QDR 40Gb)
- Low Latency
- Reliable Switch Fabric
- Offers higher layers of functionality
- Application Clustering
- Fast Inter Process Communications
- Storage Area Networks

InfiniBand Layers

- InfiniBand Adapters (Mellanox ConnectX Family)
- Drivers (mlx4_en-mlnx-1.6.1.2-471530.zip)
- Mellanox OFED 1.8.2 Package for ESXi 5.x
  - Provides ConnectX family low-level drivers
  - Provides kernel modules for InfiniBand (ib_ipoib & ib_srp)
- OpenSM Package (ib-opensm-3.3.16-64.x86_64.vib)
- Config partitions.conf for HCA
- Configure vmnic_ib in vSphere 5.5.0

OpenFabrics Enterprise Distribution (OFED)



SA	Subnet Administrator
MAD	Management Datagram
SMA	Subnet Manager Agent
PMA	Performance Manager Agent
IPoIB	IP over InfiniBand
SDP	Sockets Direct Protocol
SRP	SCSI RDMA Protocol (Initiator)
iSER	iSCSI RDMA Protocol (Initiator)
RDS	Reliable Datagram Service
UDAPL	User Direct Access Programming Lib
HCA	Host Channel Adapter
R-NIC	RDMA NIC

Key: Common; InfiniBand; iWARP; Apps & Access Methods for using OF Stack

Installing InfiniBand on vSphere 5.5

- Mellanox drivers are in vSphere 5.5, and they support 40GbE inbox when using ConnectX-3 and SwitchX products.
- If not same HCA you need to uninstall the Mellanox drivers
- esxcli software vib remove -n=net-mlx4-en -n=net-mlx4-core
- Reboot ESXi host
- Install the Mellanox drivers net-mlx4-1.6.1.2
- Install the Mellanox OFED 1.8.2
- Install the OpenSM 3.3.15-x86_64 or 3.3.16-x86_64
- Reboot ESXi host
- Stop OpenSM "/etc/init.d/opensmd stop"
- Disable the OpenSM with "chkconfig opensmd off" (only one needed if no HW SM)

```
/tmp # unzip mlx4_en-mlnx-1.6.1.2-offline_bundle-471530.zip
Archive:  mlx4_en-mlnx-1.6.1.2-offline_bundle-471530.zip
  inflating: index.xml
  inflating: vendor-index.xml
  inflating: metadata.zip
  inflating: vib20/net-mlx4-en/Mellanox_bootbank_net-mlx4-en_1.6.1.2-1OEM.500.0.0.406165.vib
/tmp # esxcli software acceptance set --level=CommunitySupported
Host acceptance level changed to 'CommunitySupported'.
/tmp # esxcli software vib install -d /tmp/mlx4_en-mlnx-1.6.1.2-offline_bundle-471530.zip
Installation Result
   Message: The update completed successfully, but the system needs to be rebooted for the changes to be effective.
   Reboot Required: true
   VIBs Installed: Mellanox_bootbank_net-mlx4-en_1.6.1.2-1OEM.500.0.0.406165
   VIBs Removed:
   VIBs Skipped:
/tmp # esxcli software vib install -d /tmp/MLNX-OFED-ESX-1.8.2.0.zip --no-sig-check
Installation Result
   Message: The update completed successfully, but the system needs to be rebooted for the changes to be effective.
   Reboot Required: true
   VIBs Installed: Mellanox_bootbank_net-ib-cm_1.8.2.0-1OEM.500.0.0.472560, Mellanox_bootbank_net-ib-core_1.8.2.0-1OEM.500.0.0.472560, Mellanox_bootbank_net-ib-ipoib_1.8.2.0-1OEM.500.0.0.472560, Mellanox_bootbank_net-ib-mad_1.8.2.0-1OEM.500.0.0.472560, Mellanox_bootbank_net-ib-sa_1.8.2.0-1OEM.500.0.0.472560, Mellanox_bootbank_net-ib-umad_1.8.2.0-1OEM.500.0.0.472560, Mellanox_bootbank_net-mlx4-core_1.8.2.0-1OEM.500.0.0.472560, Mellanox_bootbank_net-mlx4-ib_1.8.2.0-1OEM.500.0.0.472560, Mellanox_bootbank_scsi-ib-srp_1.8.2.0-1OEM.500.0.0.472560
   VIBs Removed:
   VIBs Skipped:
/tmp # esxcli software vib install -v /tmp/ib-opensm-3.3.16-64.x86_64.vib --no-sig-check
Installation Result
   Message: The update completed successfully, but the system needs to be rebooted for the changes to be effective.
   Reboot Required: true
   VIBs Installed: Intel_bootbank_ib-opensm_3.3.16-64
   VIBs Removed:
   VIBs Skipped:
/tmp # reboot
/tmp #
```

Configure MTU and OpenSM

```
/tmp # cat partitions.conf
Default=0x7fff,ipoib,mtu=5:ALL=full;
/tmp # esxcli system module parameters set -m=mlx4_core -p=mtu_4k=1
/tmp # cp partitions.conf /scratch/opensm/0x0002c9030001c71
0x0002c9030001c717/  0x0002c9030001c718/
/tmp # cp partitions.conf /scratch/opensm/0x0002c9030001c717/
/tmp # cp partitions.conf /scratch/opensm/0x0002c9030001c718/
/tmp #
```

Partitions.conf contains Protocol identifiers, like IPoIB.

Physical adapters

esx11.ebk.lab | Actions

Summary | Monitor | **Manage** | Related Objects

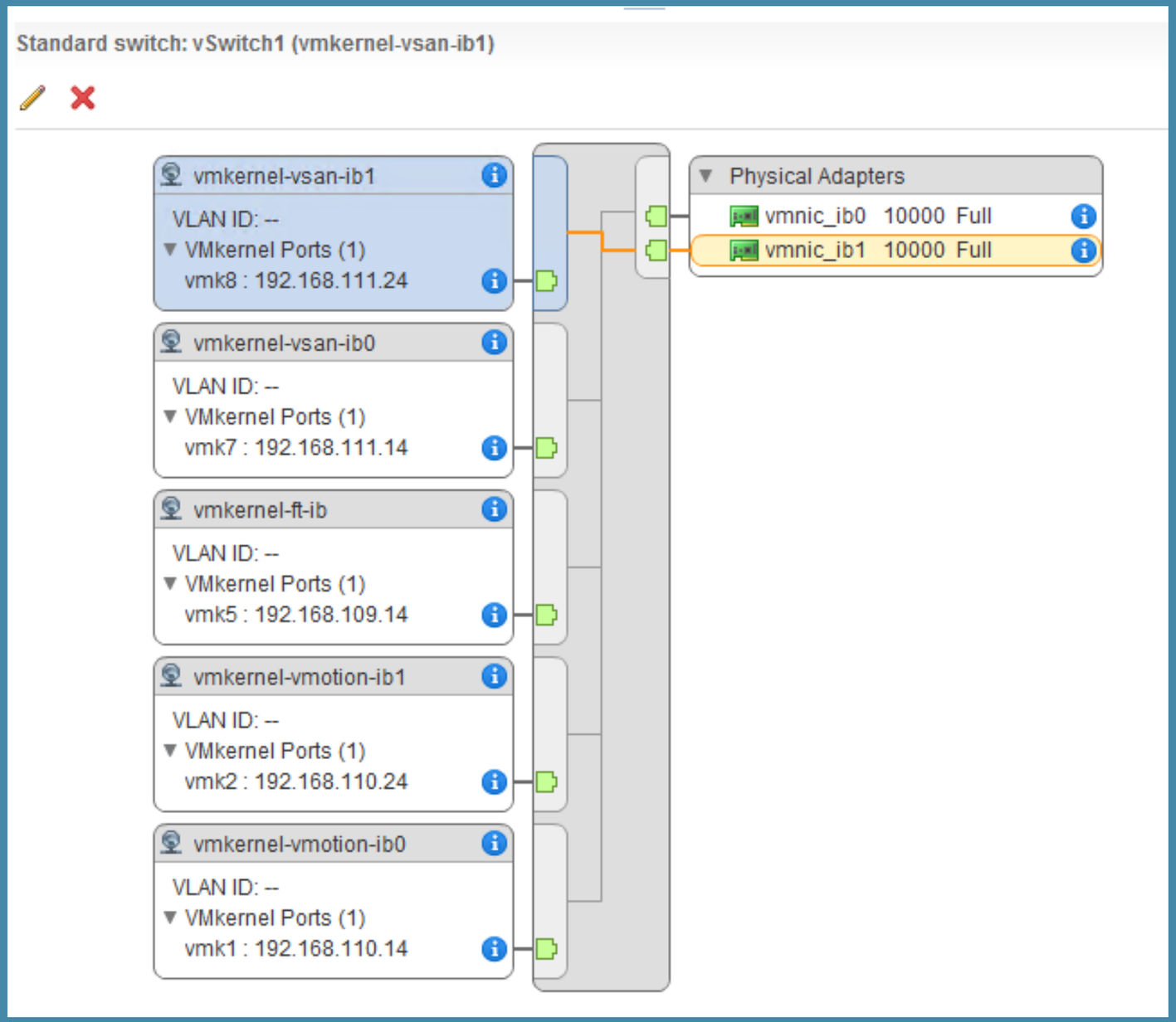
Settings | Networking | Storage | Alarm Definitions | Tags | Permissions | Application Services | Hyperic Agents

- Virtual switches
- VMkernel adapters
- **Physical adapters**
- TCP/IP configuration
- Advanced

Physical adapters

Device	Actual Speed	Configured Speed	Switch	MAC Address
Intel Corporation 82576 Gigabit Network Connection				
vmnic1	1000 Mb	Auto negotiate	vSwitch0	50:57:a8:af:f8:f6
vmnic2	1000 Mb	Auto negotiate	vSwitch0	50:57:a8:af:f8:f7
Mellanox Technologies MT25418 [ConnectX VPI - 10GigE / IB DDR, PCIe 2.0 2.5GT/s]				
vmnic_ib0	20000 Mb	20000 Mb	vSwitch1	00:02:c9:01:8a:b1
vmnic_ib1	20000 Mb	20000 Mb	vSwitch1	00:02:c9:01:8a:b2

InfiniBand IPoIB backbone for VSAN

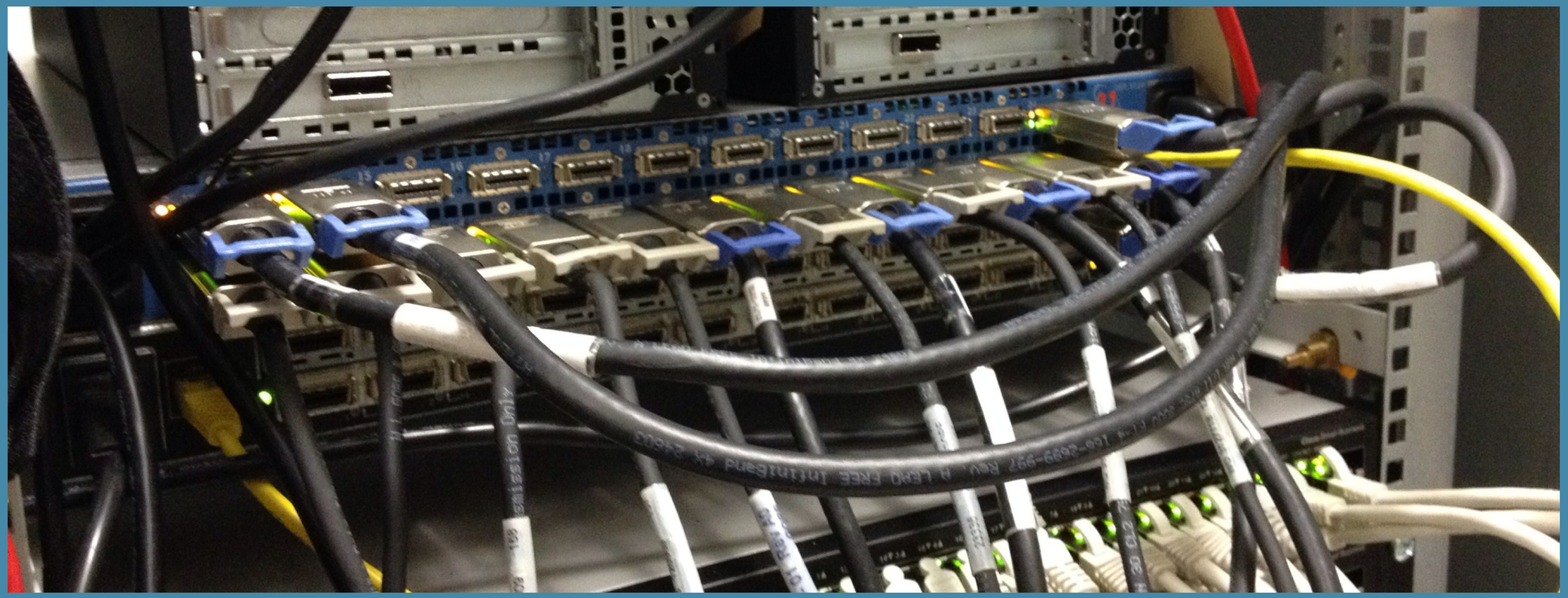


My hosted lab

- Voltaire 9024D (DDR) 24x 20GBps (without SubnetManager)
- Silverstorm 9024-CU24-ST2 (with SubnetManager)

My hosted lab

- Voltaire 9024D (DDR) 24x 20GBps (without SubnetManager)
- Silverstorm 9024-CU24-ST2 (with SubnetManager)



My hosted lab (Compute & Storage)

- 3x Cisco UCS C200M2 VSAN Storage Nodes
- 2X Cisco UCS C210M2 VSAN Compute Nodes & FVP Nodes

InfiniBand in the Lab

Fast & Cheap

Thanks to Raphael Schitz,William Lam,
Vladan Seget, Gregory Roche

Erik Bussink

@ErikBussink

www.bussink.ch

www.VSAN.info