Holland Computing Center - Site Report

Brian Bockelman

About Our Facility

- The Holland Computing Center provides for the research computing needs of the state of Nebraska.

- We provide the training, operations, software support, and hardware necessary for a wide range of sciences on the Lincoln, Omaha, and Medical Center Campuses.

- We run three general-purpose campus clusters and a dedicated CMS T2. We serve about 100 different research groups on campus.

- Offices in Lincoln and Omaha.

Our Facilities

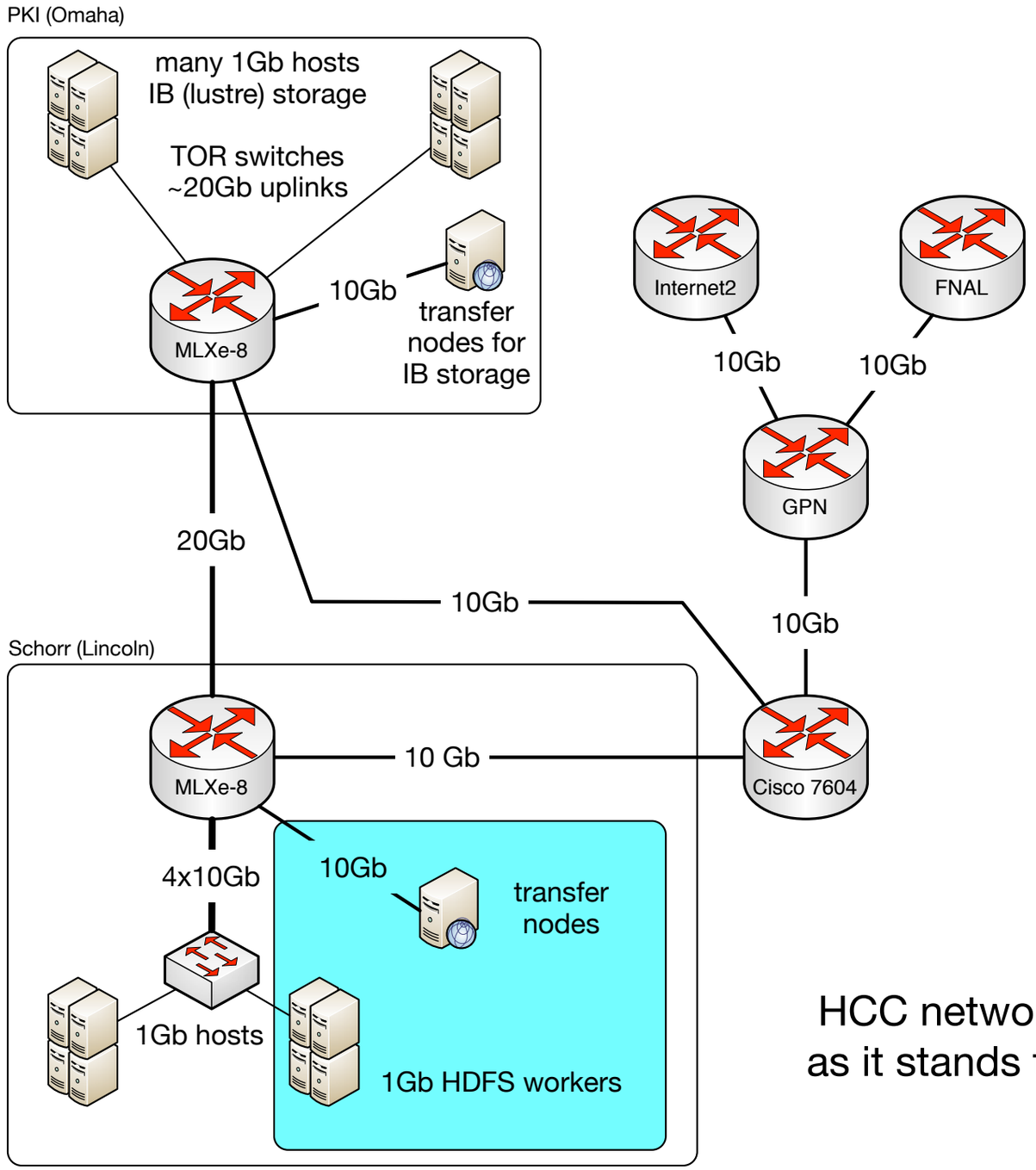
Cluster	Overview	Processors	RAM	Connection	Storage
Crane	452 node Production-mode LINUX cluster	Intel Xeon E5-2670 2.60GHz 2 CPU/16 cores per node	64GB RAM per node	QDR Infiniband	~1452 TB shared Lustre storatge ~1.8 TB local scratch per node
Tusker	106 node Production-mode LINUX cluster	Opteron 6272 2.1GHz, 4 CPU/64 cores per node	256 GB RAM per node 2 Nodes with 512GB per node	QDR Infiniband	~500 TB shared Lustre storage ~500GB local scratch
Sandhills	140 Node Production-mode LINUX cluster (condominium model)	1536 Opteron 6128 1408 Opteron 6376 368 Opteron 2354 24 Xeon E5620	44 nodes @ 128GB 2 nodes @ 256GB 44 nodes @16GB 44 nodes @ 192GB	QDR Infiniband Gigabit Ethernet	175TB shared Lustre storage ~1.5TB per node
Red	337 node Production-mode LINUX cluster	Various Xeon and Opteron processors models 5,888 cores maximum, actual number of job slots depends on RAM usage	1.5-4GB RAM per job slot	Gigabit and 10Gb Ethernet	~3.3PB of raw storage space

Red (the CMS Tier-2 cluster) receives incremental, yearly upgrades. Our other clusters are tendered as a single purchase.

Networking Infrastructure

- An exciting year for network infrastructure!
- We have done an extensive set of hardware upgrades:
    - Matching Brocade MLXe routers were deployed at our two facilities.
    - This significantly increased our bandwidth - prior core switch chases had limited number of 10Gbps ports. Each MLXe has 24 10Gbps ports lit.
    - The MLXe platform also gave us new capabilities - we no longer rely on campus for Layer-3 . Significantly simplifies internal configuration - number of broadcast domains was multiplying quickly and static routes were barely cutting it.

HCC Networking - April



HCC networking
as it stands today

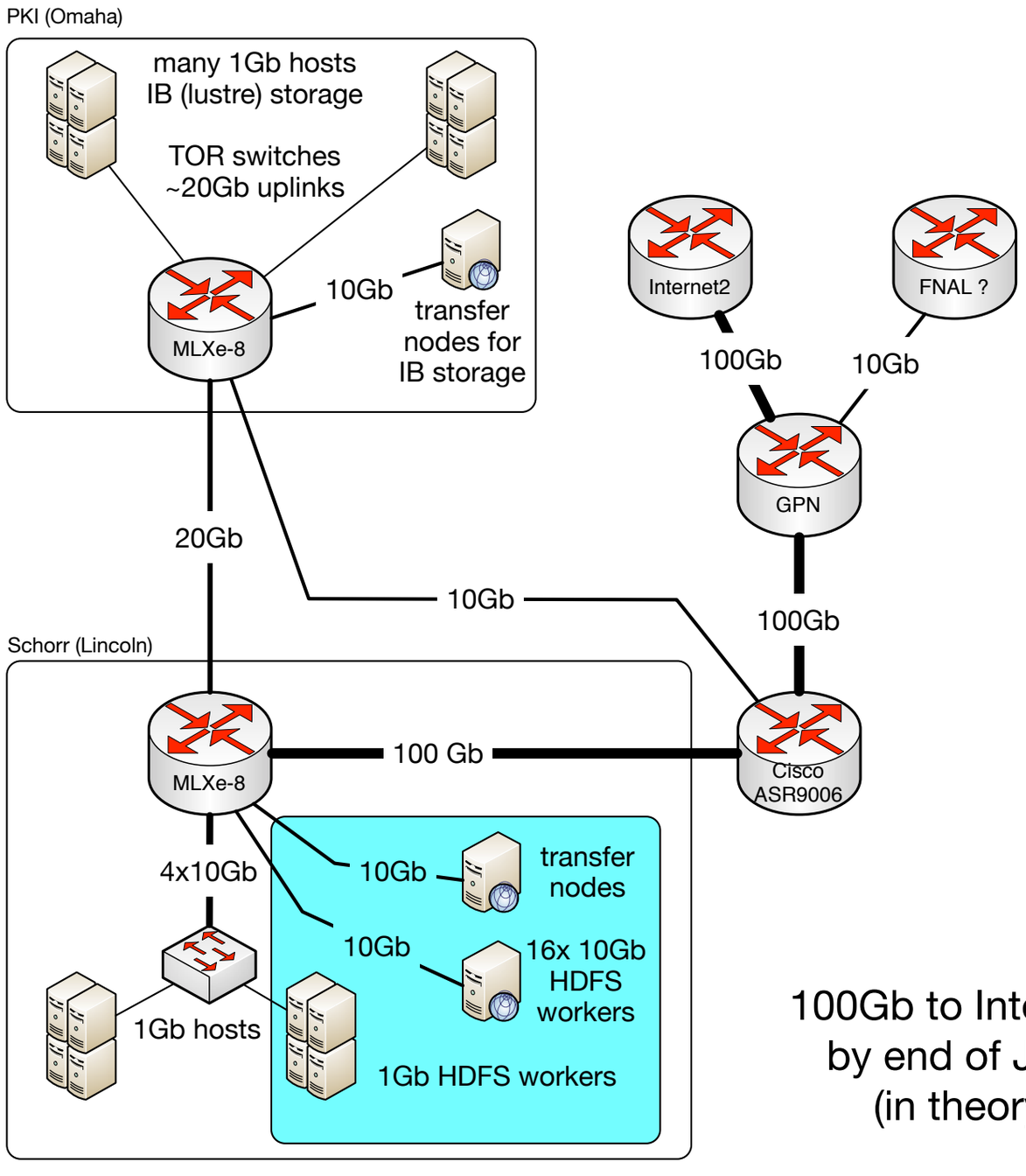
Networking Infrastructure - 100Gbps

- The MLXe also provides us with a 100Gbps capability; indeed, last month got a shiny new 100Gbps line card.

- All campus 100Gbps equipment for the connection to Internet2 is on order; hopeful install will happen during June 2014.

- It will take us a year or so to re-engineer our internal network to source 100Gbps of traffic; we expect to do around 50Gbps “day one”.

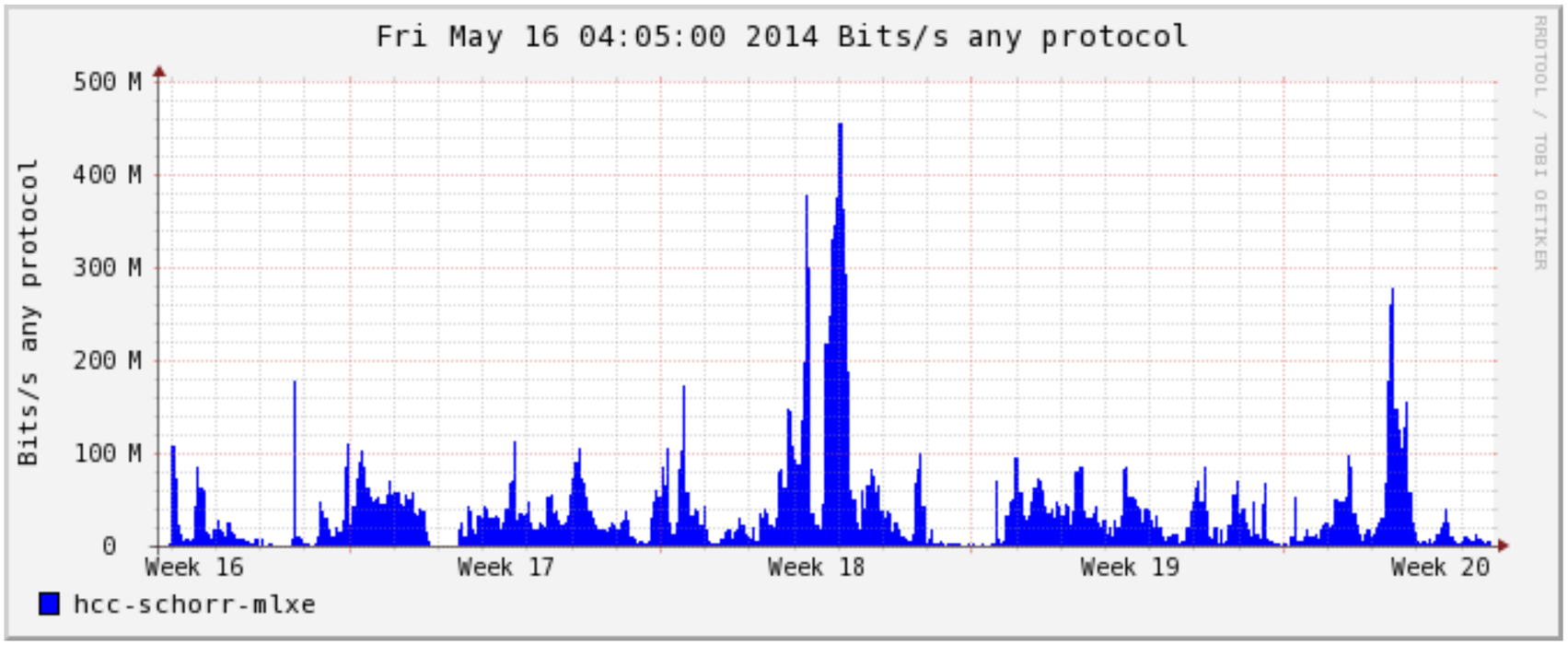
HCC Networking - June



100Gb to Internet2
by end of June
(in theory)

Networking Capabilities - IPv6

- All nodes in the CMS cluster are routable over IPv6; production traffic peaks in the hundreds of Mbps, not ten Gbps.
- Production GridFTP, SRM, Xrootd, GRAM services are IPv6-enabled.
- In the next two months:
  - DHCPv6 / AAAA records for all worker nodes.
  - Cut over HTCondor and Xrootd federation host.



Networking Capabilities - SDN

- Again, the MLXe platform opens new doors:
  - We run the research subnet using OpenFlow.
  - The production subnet has a handful of ports enabled in hybrid mode.
    - Current targeted application is prioritizing GridFTP traffic.
- Hoping to pass a VLAN from I2 to our MLXe; this way, our facility will run its own border routers. By controlling the border, we can use I2's AL2S service to setup dedicated circuits between us and the other T2s.

Hardware Deployment

- For CMS, 2013's computing purchase was 28 Huawei 1U servers with dual Intel Xeon E5-2650v2 (2.6GHz, 16HT core) and 128GB RAM. Works out to be a bit under 11 kHS06.

- Storage purchase was 288 4TB drives into pre-existing empty hard drive bays. About 0.5PB new into HDFS.

- We expect a similar outlay for 2014.

Increasing Throughput

- Our worker nodes with the most HDFS space are being upgraded with 10Gbps NICs.
- We are upgrading our 10 transfer servers (combination GridFTP / Xrootd / Squid) to each have 10Gbps NICs.
- The bulk of our filesystem will still sit behind a 4x10Gbps LAG; we are predicting we will saturate the WAN at only 50Gbps until new switch upgrades are performed.
- For problematic international links, we will start running UDT-based GridFTP transfers.

Planning Ahead - SL7

- We are starting to deploy initial test VMs with the RHEL7 RC image.

- Hope is to have worker nodes switched over by Fall 2014.

  - As with the SL6 upgrade, we will run WLCG jobs in a chroot until they are ready for SL7 native.

  - Meanwhile, any kernel and management tool improvements (systemd!) will be available to the node admins.

Conclusions / Thoughts on Run II

- Compared to 2013, the site middleware upgrades are modest for 2014.
  - No major OSG or HDFS upgrades. Continued minor updates.
  - HTCondor 8.3.0 - but we're quite good at HTCondor upgrades!
- After the networking upgrades, the rest of 2014 will be about bulletproofing and passing scale tests.
  - We have aggressive targets, especially for Xrootd exports.
- Run II? It'll be great!