T2_US_Nebraska Site Report

HEPiX Spring 2018 Workshop
May 14th, 2018

Garhan Attebury

Holland Computing Center

- Provides research computing for all University of Nebraska entities

- ~35k cores in four HPC clusters and an OpenStack ‘cloud’

- **USCMS Tier2 site**
  T2_US_Nebraska
  12k 2+GB job slots
  6.68PB raw HDFS storage
  100Gb ESnet/LHCONE/Internet2
  Dual stack IPv4/IPv6

- CMS Tier3 for UNL, KU, KSU

- Heavy involvement with the OSG

T2_US_Nebraska Resources

- **Physical boxen:** 344 workers + 12 transfer nodes + ~10 miscellaneous
- **CE resources:** ~12,000 job slots for ~110k HS06
- **SE resources:** 6.68PB across 210 HDFS datanodes (on a good day)

- Bonus resources from Anvil (8C + 24GB RAM VMs) …
  (really only a single VM for months now)
- Bonus resources from HCC HPC clusters (Crane, Tusker, Sandhills)
- Bonus nodes retired from HCC (40 * 64 = +2,560 cores, counted in CE total above)
- Three CMS focused admins (Carl, Garhan, John)
  and the ~~instigators~~ other OSG/CMS folk (Brian, Derek, Marian)

Last year : 28% storage, 36% HS06 out of warranty
This year : 59% storage, 54% HS06 out of warranty

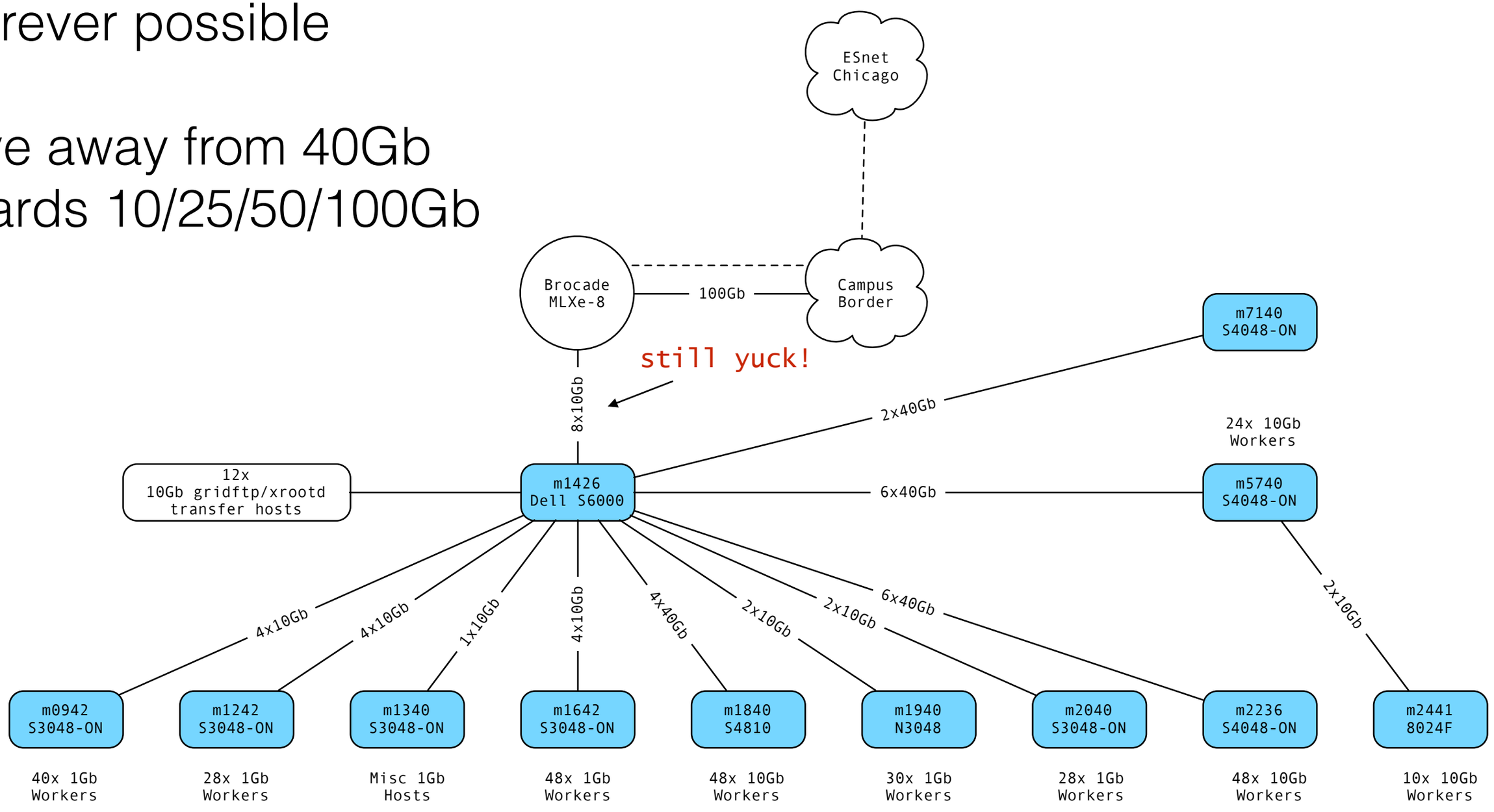
Qty	Type	CPU	RAM	Net	Cores / HT	HDFS	Purchase	Warranty
24	Dell R710	Xeon E5520 2.27GHz	24GB	1Gb	8 / 16	10.3TB	Mar 2011	out
40	Dell R510	Xeon E5530 2.4GHz	24GB	1Gb	8 / 16	21.7TB	Aug 2011	out
20	Dell R510	Xeon E5530 2.4GHz	24GB	1Gb	8 / 16	21.7TB	Jun 2012	out
36	Dell R410	Xeon X5650 2.67GHz	64GB	1Gb	12 / 24	9.8TB	Dec 2012	out
48	Dell R720xd	Xeon E5-2660v2 2.2GHz	128GB	10Gb	16 / 32	43.5TB	Nov 2013	out
28	Huawei	Xeon E5-2650v2 2.6GHz	128GB	1Gb	16 / 32		Mar 2014	out
16	Dell R730xd	Xeon E5-2640v3 2.6GHz	128GB	10Gb	16 / 32	43.5TB	Nov 2014	Nov 2019
40	Dell C6320	Xeon E5-2650v3 2.3GHz	128GB	10Gb	20 / 40		Sep 2015	Sep 2020
24	Dell R730xd	Xeon E5-2660v4 2.0GHz	256GB	10Gb	28 / 56	87.1TB	Nov 2016	Nov 2021
28	Dell R440	Xeon Gold 6126 2.60GHz	192GB	10Gb	24 / 48		Nov 2017	Nov 2022
40	Dell R815	Opteron 6272	256GB	1Gb	64			out
344					**7,280 / 12,000**	**6,776 TB**		

- Over 50% of storage and compute out of warranty … *but do we care?* Benefit of being in an academic HPC center is ‘free’ parts
    - This was true last year, and still true today!
    - Still finding 2TB drives underneath couch cushions, (and even some 4TB ones when we look hard enough)
- Previous infrastructure purchases (PDUs, racks, cabling) have been somewhat absorbed by the campus computing center, not likely going to be the case going forward.
- Bonus CPUs still a thing, currently ~40x Dell R810 w/64 cores each (somewhat horrible on the watt-per-HS06 scale, but… free!)

T2_US_Nebraska Networking 2018

Dual stack IPv4/IPv6
wherever possible

Move away from 40Gb
towards 10/25/50/100Gb



Software side

CentOS 7.4 is primary OS choice (some 7.5 now)

CE: Condor 8.6.9 + HTCondorCE + Docker with EL6 containers (+ singularity)

SE: HDFS 2.6 + combined gridftp/xrootd transfer hosts, LVS via Keepalived

Management: Cobbler 2.8.1 + Puppet 5.5.1

Why containers? Legacy reasons… mostly

Our Tier2 plays guinea pig for a lot of the new Grid/OSG software, occasionally bites us, but we usually circle back to “stable” releases

Network Research Activities

- SDN xio-callout for GridFTP-HDFS for 1+ year
- NDN + CVMFS
- Combine flow / transfer data with machine learning to:
    - Improve LVS balancing using application-aware information combined with machine learning based predictive analytics
    - Automate QoS and queue management decisions using predictive analytics
    - Develop mechanisms to predict per-user or per-project resource requirements and pre-allocate resources

~~CAPSLOCK~~ Colons are cruise control for cool -.-

- Almost everything dual stacked for a few years now with success… but what about pure IPv6 only?

- Necessary?
  Not *really*, but when has that ever been the point?

- Goal: make v6 only hosts do what dual stack ones can

IPv6 only: Round 1
The low hanging fruit

- Provisioning…
- Pure IPv6 PXE support exists (in theory) in newer hardware
- Hybrid approaches exist like bootstrapping via iPXE
- Accept failure, provision with v4 as you might still need that anyway

IPv6 only: Round 2
The marginally higher hanging fruit

- External repos without IPv6

  pkg.duosecurity.com (used to have v6 right?)
  yum.puppetlabs.com (same here, why are we going backwards?)
  repo.opensciencegrid.org (seriously people?! … ohh wait)

- Fixable with local mirrors / easy button hiera knobs in puppet … but building technical debt

IPv6 only: Round 3
Software compatibility

- A few years ago the list of things not “speaking” IPv6 was high … much better picture these days

- Few components we use needed config tweaks / upgrading

  Ganglia / check_mk / frontier-squid

- “Weird” things like SSSD ldap_uri parsing (which is 100% fine) or autofs segfaulting because puppet can’t look up HDFS namenode

- HDFS might never support IPv6
  Had to work around various checks ensuring HDFS and FUSE function

IPv6 only: Round 4
Never forget the condor knobs

- Two critical knobs in the 8.6.x series at least:

```
PREFER_IPV4 = False
IPV4_ENABLE = False
```

- It … just works?

IPv6 only: Round 5
Enter the docker

Handler for POST /v1.26/containers/create returned error: No such image: unlhcc/osg-wn-el6:latest
Error getting v2 registry: Get https://hcc-docker-registry.unl.edu/v2/: dial tcp 129.93.175.38:443: connect: network is unreachable
Attempting next endpoint for pull after error: Get https://registry-1.docker.io/v2/: dial tcp 34.200.90.16:443: connect: network is unreachable

- Again with the world not drinking the v6 koolaid …
  ohh wait, that's us again :(
- Could run local registry, which we do, just not correctly
- Load image manually… debt continues to build
- Side note: using NDP proxying for containers as testbed switch is only a pretend L3 switch, not a proper one

IPv6 only: Round 5
Light at the end of the tunnel

May 13 09:43:46 red-c1005.unl.edu cvmfs2: (cms.cern.ch) switching proxy from http://129.93.239.137:3128 to http://131.225.205.134:3126
May 13 09:43:46 red-c1005.unl.edu cvmfs2: (cms.cern.ch) switching proxy from http://131.225.205.134:3126 to http://131.225.205.133:3126
May 13 09:43:46 red-c1005.unl.edu cvmfs2: (cms.cern.ch) switching proxy from http://131.225.205.133:3126 to http://128.142.33.31:3126
May 13 09:43:46 red-c1005.unl.edu cvmfs2: (cms.cern.ch) switching proxy from http://128.142.33.31:3126 to http://128.142.168.202:3126
May 13 09:43:46 red-c1005.unl.edu cvmfs2: (cms.cern.ch) failed to download repository manifest (6 - proxy connection problem)

When connecting to a proxy, by default it will try on the IPv4 address unless the proxy only has IPv6 addresses configured. The CVMFS_IPFAMILY_PREFER=[4|6] parameter can be used to select the preferred IP protocol for dual-stack proxies.

IPv6 only: Round 6
Prognosis: not very six-ish

```
root    27815  0.1  0.1 1342040 40124 ?     Ssl  May12   0:55 /usr/bin/dockerd-current --add-runtime docker-runc=/usr/libexec/docker/docker-runc-current --default-runtime=dock
root    27826  0.0  0.0 917584 18696 ?      Ssl  May12   0:30  \_ /usr/bin/docker-containerd-current -l unix:///var/run/docker/libcontainerd/docker-containerd.sock --metrics-i
root      344  0.0  0.0 412932  4140 ?      Sl   12:56   0:00      \_ /usr/bin/docker-containerd-shim-current 9f56e063af748145071ab7be2a6f7f4f1e0d55b03e1ec4203ad0d6ac8d092035
cmsprod   362  0.9  0.0  24332  2064 ?      Ss   12:56   0:01          \_ /bin/bash ./condor_exec.exe -v std -name gfactory_instance -entry CMS_T2_US_Nebraska_Red_gw1_whole_cm
cmsprod  7151  0.3  0.0  23940  1748 ?      S    12:58   0:00            \_ /bin/bash /var/lib/condor/execute/dir_320/glide_zIZdB0/main/condor_startup.sh glidein_config
cmsprod  8112  0.0  0.0  55332  5784 ?      S    12:58   0:00              \_ /var/lib/condor/execute/dir_320/glide_zIZdB0/main/condor/sbin/condor_master -f -pidfile /var/
```

This is looking promising…

```
05/13/18 17:58:18 (pid:7388) attempt to connect to <188.184.83.197:9685> failed: Network is unreachable (connect errno = 101).  Will keep trying for 300 total seconds (299 to go).
05/13/18 18:03:17 (pid:7388) attempt to connect to <188.184.83.197:9685> failed: Network is unreachable (connect errno = 101).
05/13/18 18:03:17 (pid:7388) CCBListener: connection to CCB server vocms0806.cern.ch:9685 failed; will try to reconnect in 60 seconds.
05/13/18 18:03:20 (pid:7388) attempt to connect to <131.225.205.232:9685> failed: Network is unreachable (connect errno = 101).  Will keep trying for 300 total seconds (297 to go).
```

… or not

vocms0806.cern.ch = IPv4 only
cmssrv258.fnal.gov = IPv4 only
gfactory-1.t2.ucsd.edu = IPv4 only

IPv6 only: Final Score

- Functioning worker node?
  Yes, so long as you don't actually want to run CMS jobs

- Most IPv6 issues are fixable with a few config tweaks or upstream/external support

- Globus (gfal-copy / globus-url-copy) work fine
  XRootD (xrdcp) works fine

- Just factory support remaining? USCMS only issue?

Questions?

More on IPv6 later today from the HEPiX IPV6 working group

and IPv6 deployment experience from the GridKa Tier-1 at KIT