Integrating HEP Workflows on NSF HPCs: Blue Waters, Stampede & Jetstream

Mark Neubauer

University of Illinois at Urbana-Champaign

OSG All Hands Meeting - UCSD/SDSC
March 6-8, 2017

With contributions also from:

- Ben Hooberman, Dave Lesny, Matt Zhang, Dewen Zhong (*Illinois*)
- Rob Gardner (*Chicago*)
- Peter Onyisi (*Texas*)
- Amir Farbin (*UT Arlington*)

Blue Waters Supercomputer

- NSF-funded Supercomputer @ U. Illinois and NCSA
- Cray XE6/XK7 hybrid machine composed of
  - AMD 6276 "Interlagos" processors
  - NVIDIA GK110 "Kepler" accelerators (XK nodes)
  - Cray Gemini 3D Torus interconnect
- Compute
  - 23k Cray XE nodes
    - 362k Bulldozer cores; 1.4 PB memory
  - 4.2k Cray XK nodes
    - 34k Bulldozer cores + 4.2k Kepler accelerators
    - 135 TB CPU / 25 TB GPU memory
- Storage, I/O
  - Online: 26 PB, aggregate I/O > 1 TB/sec
  - Near-line: 380 PB,1.2 GB cache front-end, 58 GB/s aggregate to tape

National Petascale Computing Facility

Blue Waters via OSG (Atlas)Connect

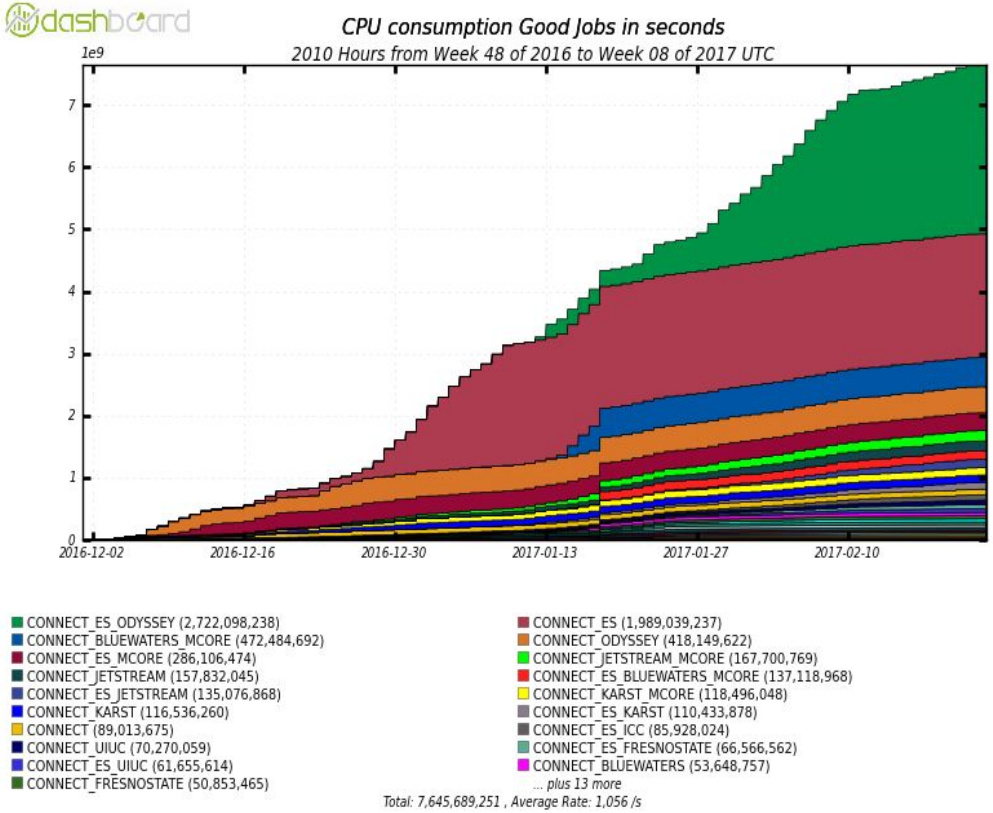
Gardner, Lesny, Neubauer

- Approach: Blue Waters as a target site for OSG CONNECT
  - Run all Atlas jobs via PanDA as any other Grid Site
  - No special changes at PanDA side
  - Use stock Atlas Pilot and access to data
  - Use Blue Waters as a normal user (i.e. no special access required)
- Leverage off of current OSG CONNECT and MWT2 services
  - Remote Cluster Connect Facility (RCCF)
  - Remote Cluster Connect AutoPyFactory
  - Stratum-R CVMFS replication service
  - MWT2 Storage Element
  - MWT2 Squids, GUMS, MWT2 CVMFS Repository

OSG CONNECT PanDA queues

Gardner, Lesny, Neubauer

- Separate set of PanDA Queues for each CONNECT target site
- Currently nine active target sites
- Last 3 months provided over 2M CPU-Hours



CONNECT-BW Development Timeline

Gardner, Lesny, Neubauer

- Allocation available to Illinois faculty outside of XSEDE
  - Mark Neubauer, Illinois faculty and MWT2 PI
  - Proposal submitted in March 2016 requesting 1M node hours
  - 50k+25K node hours granted for development in May 2016
  - Allocation starting August 1, 2016 and ending March 31, 2017
- Initial investigation on scope of work began June 2016
- Blue Waters accounts created August 1, 2016
  - Using OTP RSA SecurID, create proxies valid for 11 days
  - Use GSISSH to log into BW Login nodes
  - Learn the layout of BW, PBS scheduler, aprun, etc

Blue Waters Devel Timeline (cont.)

Gardner, Lesny, Neubauer

- Modify Connect to support GSISSH, BW PBS, etc
  - First CONNECT Glidein ran on August 31, 2016

```
##############################################################################
#####                                                                    #####
#####          Job is running within a Remote Cluster Connect Factory    #####
#####                                                                    #####
##### Date: Wed Aug 31 09:25:04 CDT 2016                                 #####
##### RCCF: bluewaters                                                   #####
##### User: ddl                                                          #####
##### Host: nid00176                                                     #####
#####                                                                    #####
##############################################################################

nid00176

##############################################################################
################################ Job Complete ################################
##############################################################################
```

Blue Waters Devel Timeline (cont.)

Gardner, Lesny, Neubauer

- Initial Shifter/docker work revealed bugs in Shifter
  - Problem with network access
  - Cray did not resolve the problem until December 2016
- Lack of fuse created CVMFS access problems
  - Stratum-R GO replication became operational in December 2016
- First Shifter/Docker/CVMFS job ran December 15, 2016
- Local Site Mover (LSM) from MWT2 adopted for BW
  - Used pCache to reduce the number of files in stagein
  - Records transfer data to Elastic Search

Blue Waters Devel Timeline (cont.)

Gardner, Lesny, Neubauer

- First PanDA pilot ran January 4, 2017
- First Hammer Cloud job ran January 6, 2017
- Initial PanDA jobs failed
  - Some issues with GO replication (fixed)
  - Some missing RPMs in Docker Image (fixed)
- Began a production run from January 13-17, 2017
  - Standard and Event Service production jobs
  - Used up to 48 glide-ins
    - 16 nodes / glidein, 16 cores / node
  - 12K cores used all 50K hrs in 1 week

→ 25k more

CONNECTing Blue Waters to PanDA

Gardner, Lesny, Neubauer

- Platform
  - Blue Waters OS is Cray Linux Environment (CLE)/SUSE Linux 3.x
  - Container support via Shifter / Docker
  - Built Image on Docker Hub:
    - SL6 with all need utilities
    - HEP OSlibs meta rpm
    - Many other compatibility rpms not listed in HEP OSlibs
- ATLAS releases and other repositories
  - "Stratum-R": full disk replica of ATLAS releases sync'd daily
  - Replication to local disk at Blue Waters via Globus Online
  - Symlink in Docker image points /cvmfs to replicas on disk

Blue Waters Storage

Gardner, Lesny, Neubauer

- Very Large Lustre Scratch disk, 14P total
  - Quota: 500TB and 100M files
  - Holds Stratum-R files, pCache and scratch space for all jobs
  - Needed to get quota raised to 200M files (using 150M)
  - Needed whitelisting of space to avoid 30 day cleanup
  - Globus Online endpoint: ncsa#Bluewaters

Blue Waters CVMFS Replicas

Gardner, Lesny, Neubauer

- CVMFS client is not available (fuse not allowed)
  - Solution: replicate all needed repositories with Stratum-R
- Stratum-R: Provides local disk copies of all repositories
  - Servers at UChicago, Illinois, Indiana
  - Serves files via “rsync” (over x509 protected stunnel) and Globus Online
- Use Globus Online to copy/sync repositories to BW
  - 11 day proxy created with OTP on GO to access the BW endpoint
  - GO does not replicate symlinks or permissions
  - After GO transfer completes, a symlink creation pushed via GSISSH
  - Automatically scheduled via crons from the Connect Facility

Blue Waters Glideins

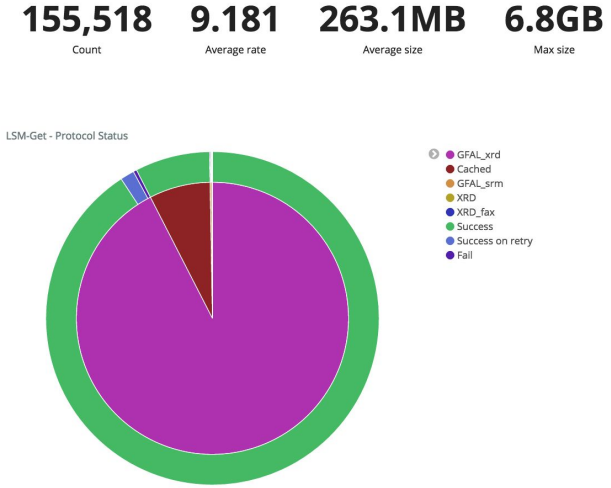
Gardner, Lesny, Neubauer

- Local Scheduler: PBS
  - Requires multiple nodes reservation per job: Currently requesting 16
  - Each node 32 cores, 64 GB, no swap => use only 16 cores to avoid OOM
- GSISSH based Glidein (Connect Factory)
  - Authorization: One Time Password creates proxy good for 11 days
  - Glidein requests 16 nodes and runs one HTCondor overlay per node
  - Requests Shifter usage with a Docker Image from Docker Hub
  - HTC overlay creates 16 partitionable slots with 16 cores per slot
  - Connect AutoPyFactory injects pilots into these slots which run on BW
  - Glidein life is 48 hours and will run consecutive Atlas jobs in the slots
  - Need a mix of standard and Event Service jobs to minimise idle cores

Blue Waters Data Transfer

Gardner, Lesny, Neubauer

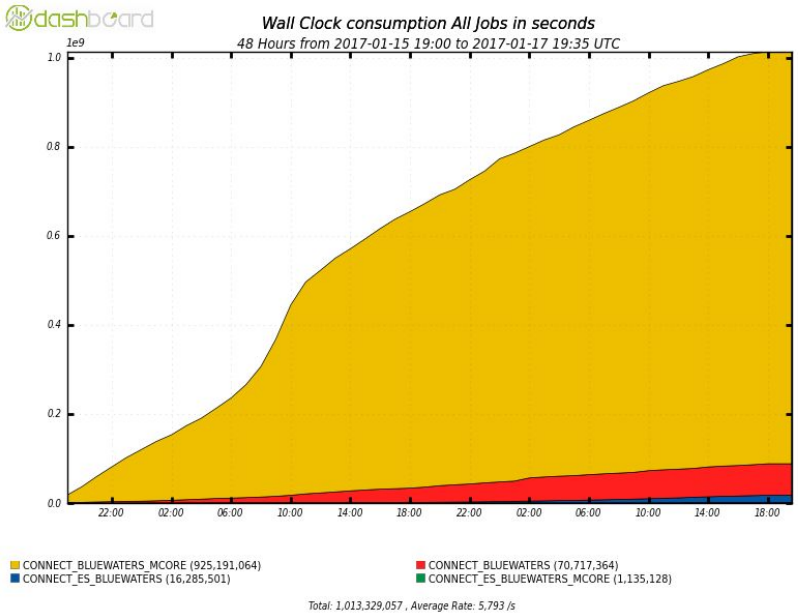
- BW nodes have limited access to WAN
  - Number of ports available to outside is restriction
  - Ports needed for HTC overlay and stagein/out of data
- "Local Site Mover" (`lsm-get`, `lsm-put`)
  - Using MWT2 SE as storage endpoint
  - Transfer utility is gfal-copy, root://, srm:// or Xrootd; retries with simple backoff and protocols change on failure; pCache (WN cache) used by lsm-get to help reduce stagein of duplicate files
  - I/O metrics logged to Elastic Search

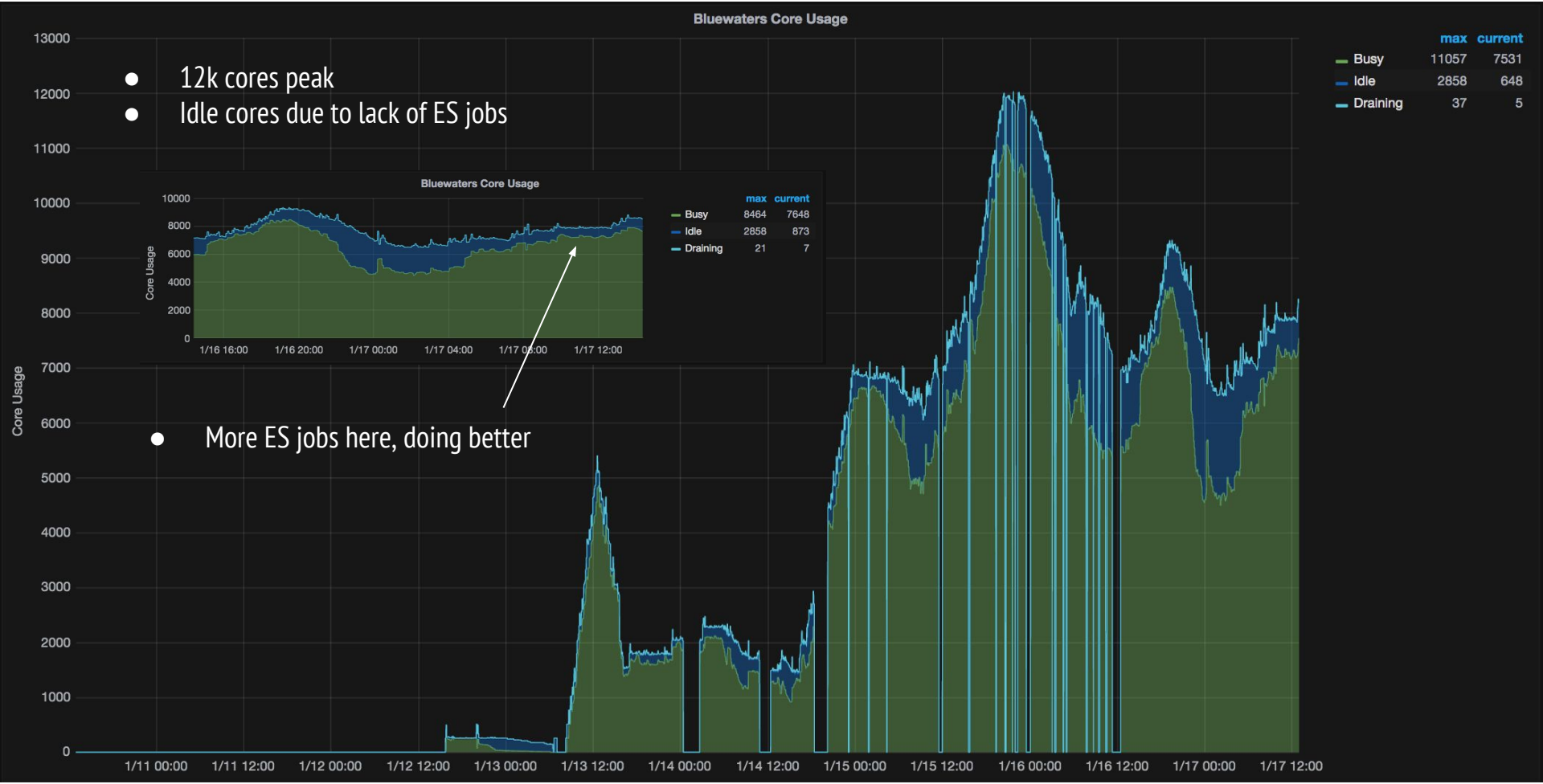


Panda Queues setup

Gardner, Lesny, Neubauer

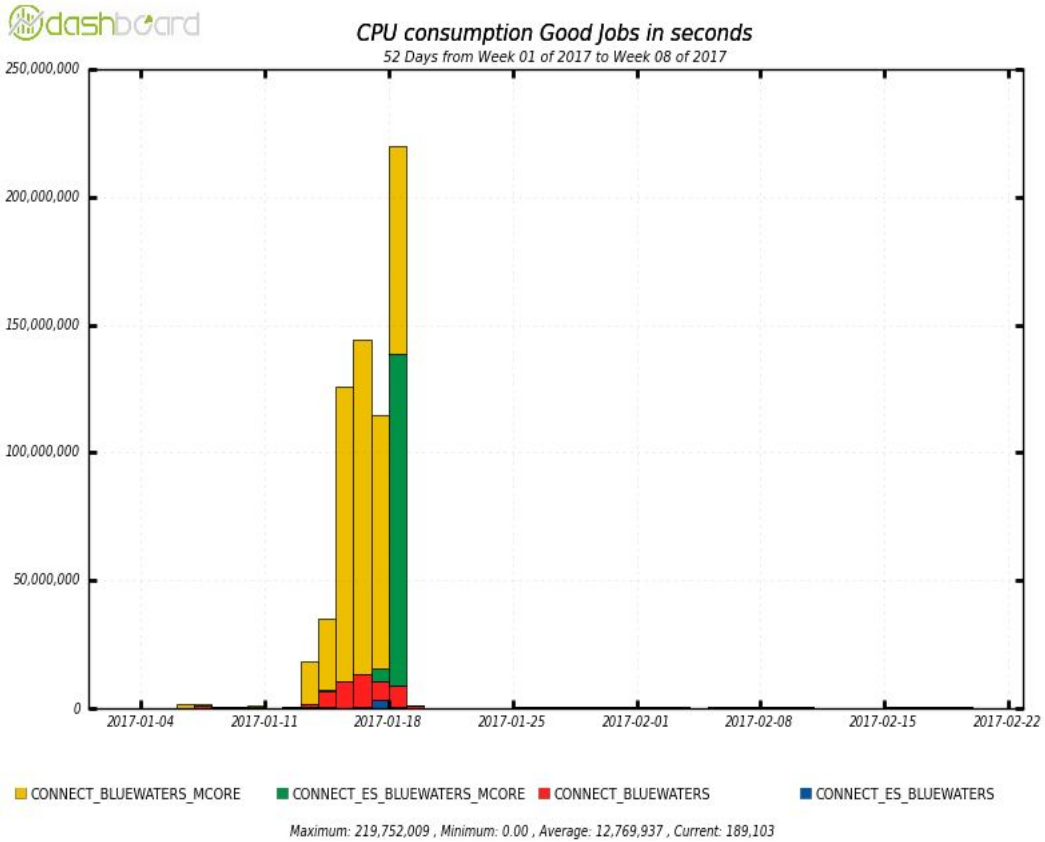
- 4 Panda (**general**) Production Queues
  - CONNECT_BLUEWATERS
  - CONNECT_BLUEWATERS_MCORE
  - CONNECT_ES_BLUEWATERS
  - CONNECT_ES_BLUEWATERS_MCORE
  - **No restriction on tasks or releases**
- Each queue configured for BW
  - LSM transfer
  - Standard: 36H guaranteed
  - ES: 4H guaranteed up to 36H max
  - 4H jobs fill in scheduling holes

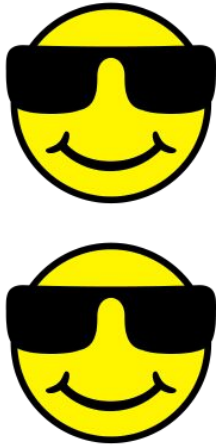



Bluewaters Core Usage
Core Usage
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1/11 00:00
1/11 12:00
1/12 00:00
1/12 12:00
1/13 00:00
1/13 12:00
1/14 00:00
1/14 12:00
1/15 00:00
1/15 12:00
1/16 00:00
1/16 12:00
1/17 00:00
1/17 12:00
max current
Busy 11057 7531
Idle 2858 648
Draining 37 5
12k cores peak
Idle cores due to lack of ES jobs
More ES jobs here, doing better
Bluewaters Core Usage
Core Usage
10000
8000
6000
4000
2000
0
1/16 16:00
1/16 20:00
1/17 00:00
1/17 04:00
1/17 08:00
1/17 12:00
max current
Busy 8464 7648
Idle 2858 873
Draining 21 7

PanDA CPU provided by Blue Waters

Gardner, Lesny, Neubauer





Stampede via CONNECT

Lesny, Onyisi

- Currently no allocation on Stampede
- Stampede is just another target site for CONNECT
- CONNECT submits SSH Glideins via special head node
  - Each glidein requests whole nodes of 16 cores
  - Maximum of 50 glideins (800 cores total)
- No CVMFS or fuse so Stratum-R replicates repositories
  - Uses rsync over stunnel with x509 keys
  - All nodes have symlink from /cvmfs to replication
  - ACE Image is used to provide all Atlas dependencies
  - PanDA production have run successfully on Stampede

Jetstream

Lesny, Onyisi

- XSEDE resource: cloud primarily targeted at few-VM interactive users, but has potential for HTC as well

- UT-Austin/Chicago/Illinois granted dedicated allocation of 2.1M core-hours (240 core-years)
  - Exploring potential for backfill as well
  - Interactive use → often not at capacity
- Nodes are “boring”
  - Full WAN access, choice of OS, CVMFS is “easy”…
  - But we need to configure many independent machines

Jetstream Operations

Lesny, Onyisi

- Use Jetstream-standard CentOS 7 image + Docker to run a CentOS 6 instance
  - CVMFS has caused some trouble
  - Currently using “Portable CVMFS” brought in with pilots
  - Plan to move to exporting CVMFS from the parent VM
- Cloud-cfg + docker image brings up a VM node which automatically joins UT-Austin condor cluster
  - jobs dispatched transparently from ATLAS Connect
- Data files accessed from WAN

Jetstream Observations

Lesny, Onyisi

- Mismatch of VM “flavors” with needs
  - Particularly disk per core (memory per core is OK)
  - Working to get more ephemeral storage
- Data-heavy jobs (e.g. ATLAS skimming) get ~ 50% CPU efficiency. CPU-heavy jobs (simulation) near 100%.
  - Would xrootd cache help?
- VMs sometimes die for no clear reason
  - But overall works well

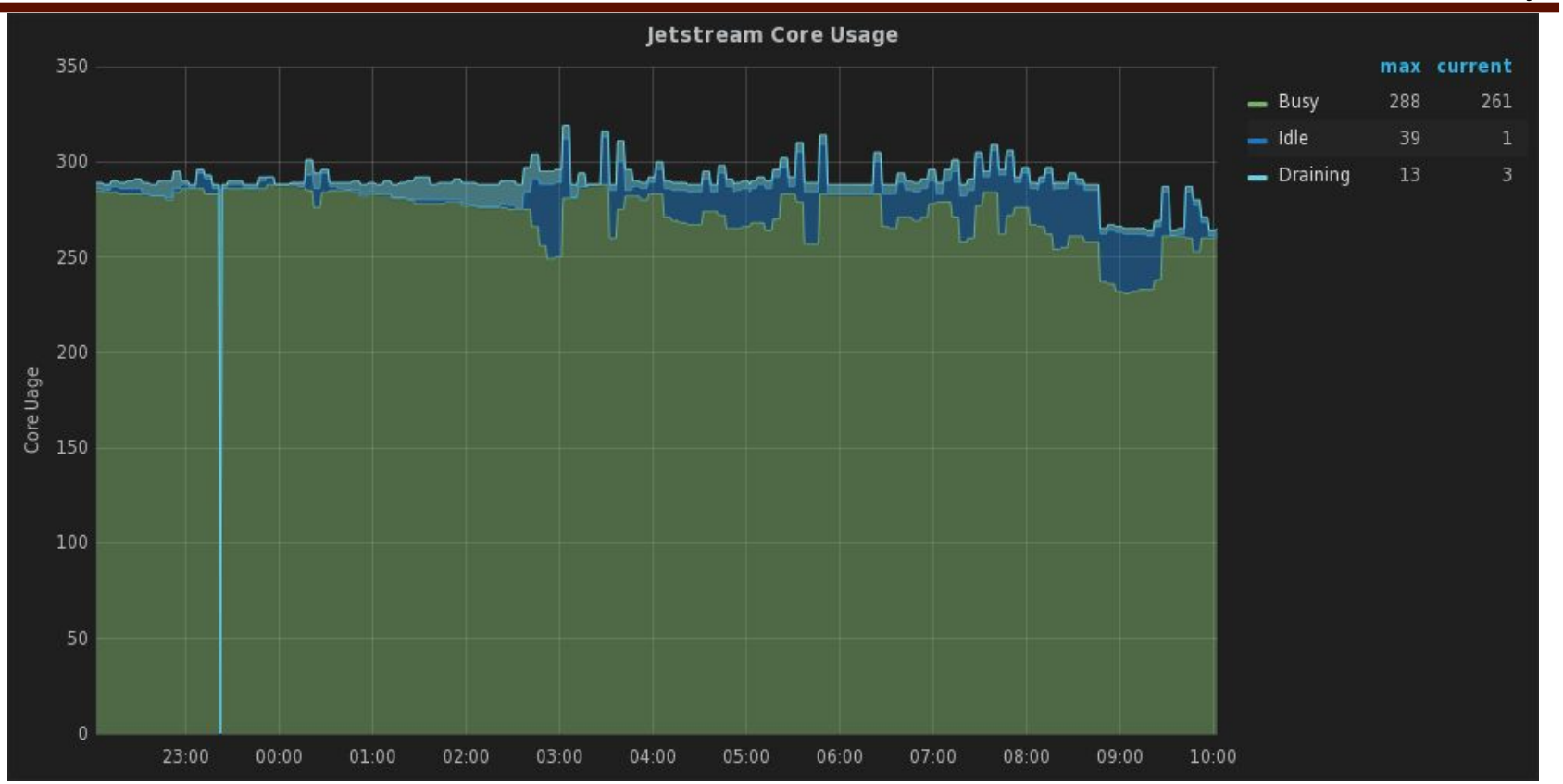
JetStream via CONNECT

Lesny, Onyisi

- Jetstream is just another target site for CONNECT
  - VMs reside in a Condor pool with SCHEDD on utatlas tier3 login node
- CONNECT submits SSH Glideins into this pool
  - Each glidein requests the whole VM (24 cores, 48GB memory)
  - Allows Connect to do its own scheduling, matchmaking, classads
  - PortableCVMFS brought into the VM (which has fuse)
  - Docker image has all other Atlas dependencies
- PanDA access via CONNECT AutoPyFactory
  - CONNECT_JETSTREAM, CONNECT_JETSTREAM_MCORE
  - CONNECT_ES_JETSTREAM, CONNECT_ES_JETSTREAM_MCORE

JetStream Cores via CONNECT

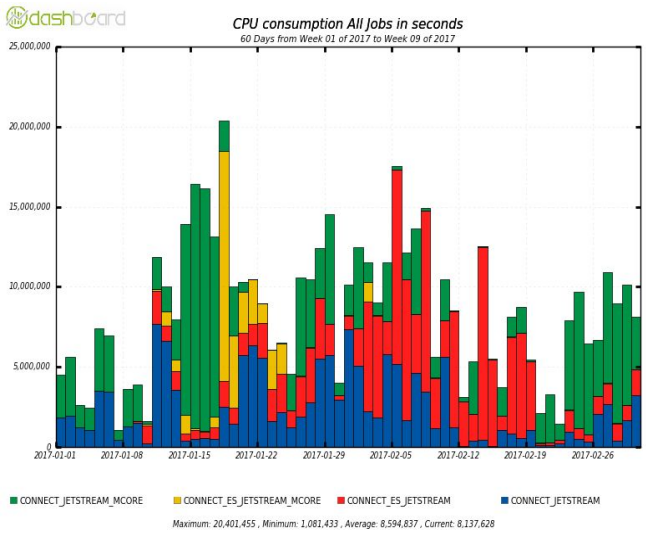
Lesny, Onyisi

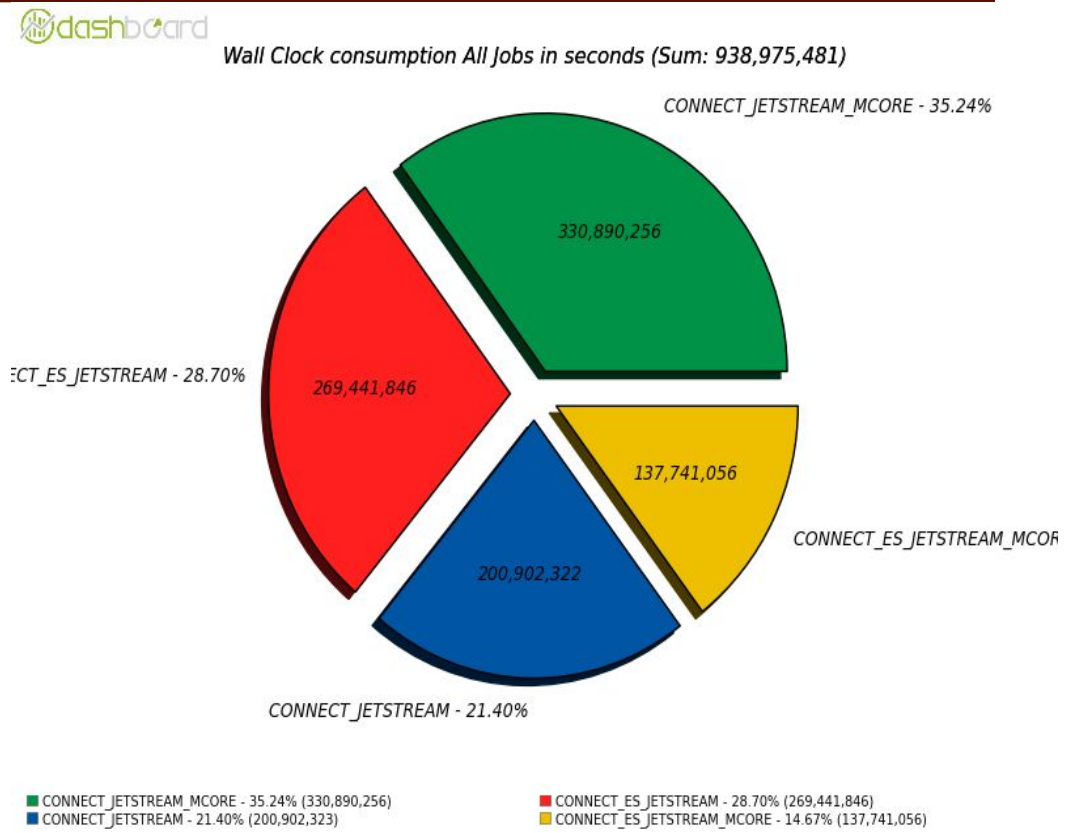


JetStream PanDA (January 1, 2017 to today)

Lesny, Onyisi

- Total: 261K cpus hours
- Using 12 24-core VMs
- Evenly split over all Qs





Machine Learning Applications for HPCs

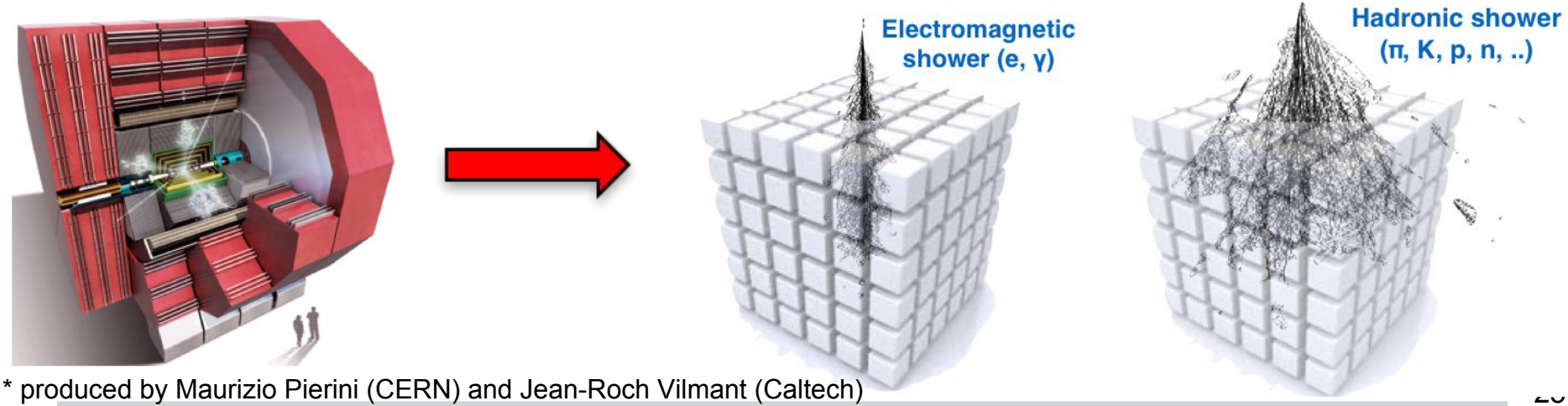
Farbin, Hooberman, Neubauer, Peirini, Vilmant, Zhang, Zhong

- Many HPCs have very large GPU resources (e.g. BW)
- Support for concurrency via GPUs built into ML tools
- HPCs provide large I/O resources for training
  - HPCs fertile ground for DNN training
- Ideas we are pursuing for DNN applications on HPCs:
  - Particle ID (Farbin, Hooberman, *et al.*) [see next slides]
  - Object ID with multiple particle (e.g. H taggers) (Neubauer, *et al*)
  - Matrix Element Method using DNNs (Neubauer, *et al*)

ML on HPC Example: Particle ID

Farbin, Hooberman, Pierini, Vilmant, Zhang

- **Goal**: Train DNNs for $e/\gamma/\pi^0/\pi^\pm$ discrimination
- Proof-of-principle w/ samples* based on CLIC LCD detector
  - Simple calorimeter geometry with uniform cell sizes for development
- Extend to more complex geometries (e.g. CMS/ATLAS)



* produced by Maurizio Pierini (CERN) and Jean-Roch Vilmant (Caltech)

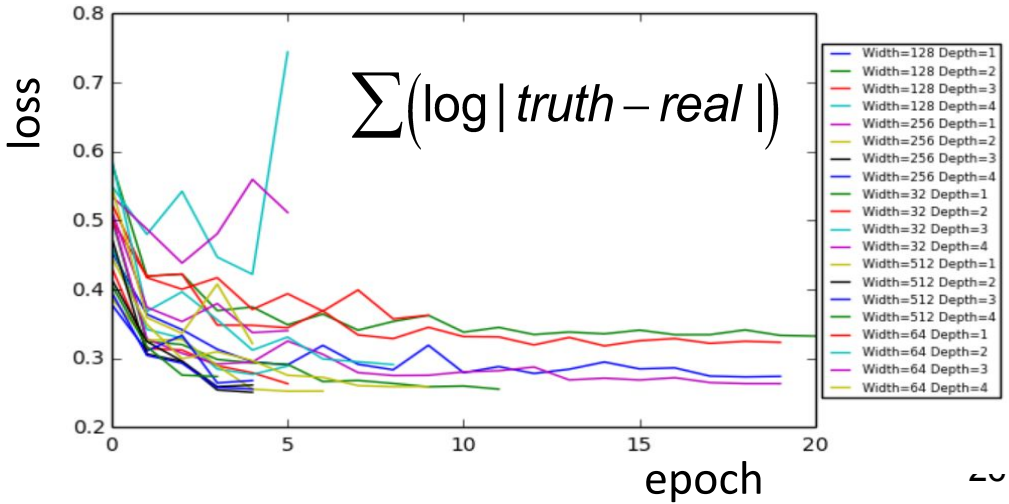
ML on HPC Example: Particle ID

Farbin, Hooberman, Zhang

- Train NN ensembles w/ different widths & depths ("models")
- Use 4 output nodes for 4 particle types ($e/\gamma/\pi^0/\pi^\pm$)
- Over epochs, find models that minimize loss & training time

make ROC curves for each particle type against the others and look at "Area Under Curve"

Model Name	Width	Depth	Epochs	$e^\pm$ AUC	π^0 AUC	$\pi^\pm$ AUC	γ AUC
Width=32 Depth=1	32	1	25	0.9857	0.9552	0.9977	0.9559
Width=32 Depth=2	32	2	13	0.9841	0.9474	0.9985	0.9510
Width=32 Depth=3	32	3	9	0.9627	0.8205	0.9965	0.9086
Width=32 Depth=4	32	4	9	0.9621	0.9015	0.9980	0.9037
Width=64 Depth=1	64	1	23	0.9875	0.9562	0.9985	0.9605
Width=64 Depth=2	64	2	12	0.9893	0.9646	0.9988	0.9656
Width=64 Depth=3	64	3	9	0.9833	0.9561	0.9984	0.9595
Width=64 Depth=4	64	4	7	0.9854	0.9577	0.9985	0.9585
Width=128 Depth=1	128	1	23	0.9922	0.9676	0.9991	0.9663
Width=128 Depth=2	128	2	15	0.9934	0.9696	0.9993	0.9703
Width=128 Depth=3	128	3	8	0.9930	0.9678	0.9991	0.9690
Width=128 Depth=4	128	4	9	0.9904	0.9624	0.9990	0.9639
Width=256 Depth=1	256	1	23	0.9931	0.9699	0.9990	0.9691
Width=256 Depth=2	256	2	10	0.9942	0.9712	0.9991	0.9704
Width=256 Depth=3	256	3	7	0.9941	0.9683	0.9992	0.9688
Width=256 Depth=4	256	4	7	0.9933	0.9663	0.9992	0.9673
Width=512 Depth=1	512	1	13	0.9934	0.9701	0.9989	0.9694
Width=512 Depth=2	512	2	7	0.9948	0.9699	0.9991	0.9700
Width=512 Depth=3	512	3	7	0.9952	0.9682	0.9992	0.9685
Width=512 Depth=4	512	4	6	0.9939	0.9653	0.9989	0.9657



Summary & Future Work on Blue Waters

- Submit a second proposal by March 15, 2017
  - Technical feasibility established with “pilot” (75k node-hrs) allocation
  - Request 1M node hours (should get ~1 yr run with 12k cores)
- Change Glide-ins to use backfill and preemption
  - Backfill are small node allocations requesting under 6 hours
  - Preemption allows jobs to be killed at any time
  - Uses ½ of the node hours
  - Perfect for Event Service jobs
- Continue development of ML applications on HPCs