New Directions and HTCondor @ BNL

USATLAS TIER-3 & NEW COMPUTING DIRECTIVES

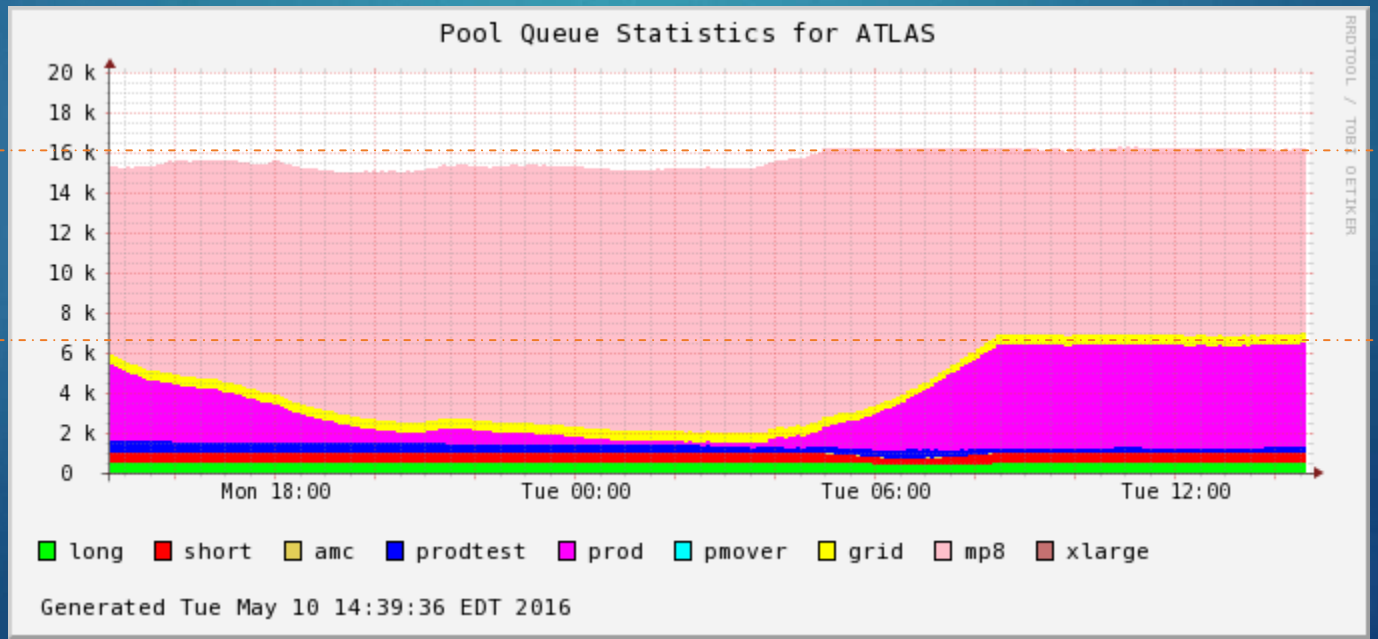
William Strecker-Kellogg
RHIC/ATLAS Computing Facility (RACF)
Brookhaven National Lab
May 2016

RACF Overview

- RHIC—Collider at BNL
  - STAR+PHENIX detectors
  - ~15000 slots each
  - ~6Pb central Disk (GPFS/NFS)
  - ~16Pb Distributed Disk (dCache/xrootd)
  - ~60Pb HPSS Tape
- USATLAS T1
  - ~15kCores
  - ~11Pb Replicated dCache Disk
  - ~20Pb HPSS Tape
- 3 Large HTCondor clusters
  - Share resources via flocking
  - Several Smaller clusters
  - Recent 8.2→8.4 update

ATLAS Load Balancing

- Working steadily since last year's talk, and HEPIX Fall 2015 talk
- Allows occupancy to remain at 95% or above despite dynamically changing workload with no human intervention
- Prevents starvation due to competition with larger jobs
  - Only inefficiency is due to (de)fragmentation



V8.4 Bug Fixing

- Minor issue with classad not appearing in the negotation context
  - RemoteGroupResourcesInUse
  - In context of group-based preemption
  - Fixed quicky by Greg
- Ticket #5593

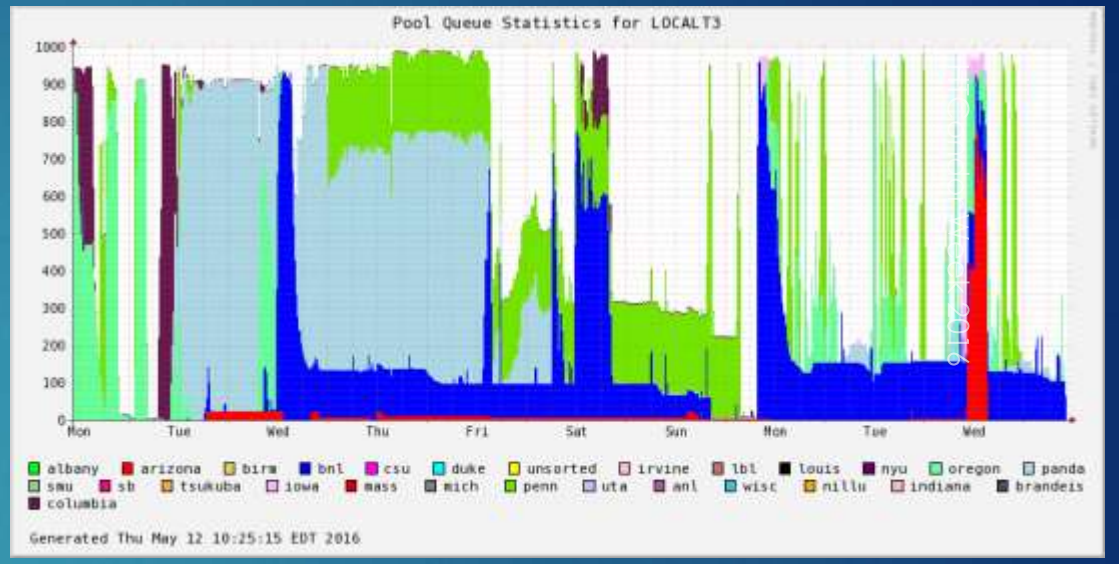
- Major issue with Schedd halting for up to an hour
  - No disk/network IO of any kind
  - No syscalls
  - GDB shows a mess (STL)
- Ticket #5648

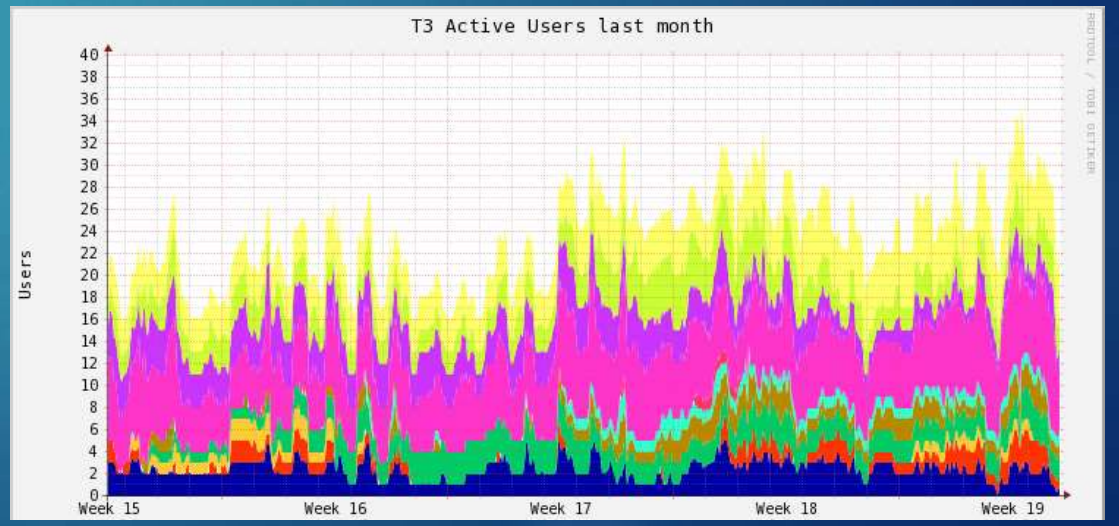
V8.4 Schedd Bug

- Schedd spending up to an hour recomputing internal array using autocluster→jobid and the reverse mapping
- Jobs would die after their shadows couldn't talk to their schedd that went dark
- After day of debugging, code fix was implemented
  - Built & tested at BNL, max time reduced from 1h to 2s
  - Thank you to Todd & TJ
- Still suspect something off in our environment (500k ac/day)
  - Not a problem anymore!

USATLAS Tier-3 @ BNL

- Consolidates previously scattered Tier-3 facilities
- Shared resource ~1000 cores
- Many user-groups represented
  - Local submission
  - Hierarchical Group Quotas
    - Group-Membership authorization problem
    - Surplus sharing
  - Group-based fair-share preemption
- After slow start, increased usage in past few months





Group Membership

1 users		SUNY Albany	albany	group_atlas.albany
1 users		Argonne National Laboratory	anl	group_atlas.anl
3 users		University of Arizona	arizona	group_atlas.arizona
1 users		University of Birmingham	birm	group_atlas.birm
34 users		Brookhaven National Laboratory	bnl	group_atlas.bnl
1 users		Brandeis University	brandeis	group_atlas.brandeis
1 users		Columbia University	columbia	group_atlas.columbia
1 users		California State	csu	group_atlas.csu
4 users		Duke University	duke	group_atlas.duke
1 users		Generic or Unknown Institute	general	group_atlas.general
1 users		Indiana University	indiana	group_atlas.indiana
1 users		UC Irvine	irvine	group_atlas.irvine
1 users		Lawrence Berkeley National Lab	lbl	group_atlas.lbl
1 users		Louisville	louis	group_atlas.louis
		Northern Illinois University	nillu	group_atlas.nillu
3 users		New York University	nyu	group_atlas.nyu
4 users		University of Oregon	oregon	group_atlas.oregon

- Extended group-quota editing web UI
  - Added user-institute-group mappings
  - Cron generates a config fragment that asserts Owner → Group in START expression
  - Require group at submission
- Currently 74 users and 26 groups use the T3
- Surplus sharing & group-respecting preemption
  - RemoteGroupResourcesInUse

Preemptable Partitionable Slots

- Users are frequently asking to run high-memory jobs…
- …which motivates: User-prio preemption with partitionable slots
  - Absolutely need to support group-constrained fair-share with preemption
  - Currently Pslot Preemption operates on entire slot
    - Entire group's quota can fit on one 40-core machine
      - A function of the small scale of having 30+ groups sharing 1000 cores
  - No way to respect group quotas as schedd splits slots

- Not Currently Possible!

New Directions

WHAT WE TALK ABOUT WHEN WE TALK ABOUT

H[TP]C

New Science & Evolving Needs

Traditional Model

- RACF as a microcosm of HEP/NP computing
  - Many other facilities of a similar (within an order of magnitude) scale and capability
  - Embarrassingly parallel workloads
- Data storage as large or larger a problem than computing
- Batch is simple—provisioning/matchmaking vs. scheduling
- Large, persistent, well-staffed experiments

New Model

- New Users with traditionally HPC-based workloads
  - National Synchrotron Lightsource II
  - Center for Functional Nanomaterials
- Revolving userbase
  - No institutional "repository" of computing knowledge
  - Not used to large-scale computing
- Software support not well-defined
  - Large pool of poorly supported open source or free/abandon-ware
  - Commercial or GUI-Interactive

Institutional Cluster

- New cluster of 108 nodes, each with 2xGPUs
  - Plans to increase next year by a factor of 2, then perhaps more
- Infiniband interconnect in fat-tree topology
- SLURM is being evaluated
  - Seems to be the growing choice for new HPC clusters
    - Active Development
    - DOE experience
    - Open Source
    - Large userbase (6 of top 10 of TOP500)
  - Test sample workloads in proposed queue configurations
  - Set up Shifter for docker intergration
- Will be run as a "traditional" HPC cluster
  - MPI support an important consideration

New Science & Evolving Needs

- Lab-management support for consolidation of computing
- Computational Science Initiative (CSI) at BNL "leverages" experience at RACF in support of other non-HEP/NP science domains
- CSI is also organizing software support to fill a gap in current BNL computational services

…

- The $10,000 question:

Can we leverage existing infrastructure?

Running at HTC Facility

Zero-Order Requirements

- Embarrassingly Parallel
  - (small) Input → (one) Process → (small) Output
  - No communication of intermediate results
- X86_64
  - Other hardware not standard in the community
- Data accessible
  - May seem obvious, but need adequate bandwidth to get data to the compute and back
  - Something to think about of moving from single desktop

Running at HTC Facility

First-Order Requirements

- Linux (RedHat)
  - Virtualization is an extra complexity, Windows expensive
  - Containers / Docker allows simple cross-linux compatibility
- Free Software
  - Instance-limited licenses are hard to control across many pools
  - Cost of licenses becomes prohibitive with exponential computing growth
- "Friendly" resource profile
  - Code runs not just within the machine, but within the general limits its neighboring jobs use

HTC/HPC Divide

- Not a false dichotomy, but surely an increasingly blurry line
  - Several users in our experience fit in middle-ground (albeit with considerable help from the RACF to fit workload into an exclusive HTC environment)
    1. Biology Group: 800 cores for 5 months simple dedicated scheduler
    2. Wisconsin group at CFN: successfully ran opportunistically on RHIC resources
- Key factors
  - How much state transfer and with what IO patterns?
  - Size
    - 10 years to now: 2 racks collapse into 1 machine
    - How many problems fit inside one machine today?

Scheduling

- HTCondor recently can submit to SLURM via grid universe
- Different sharing models
  1. Condor-as-SLURM-job (glidein-style)
  2. Coexist, mutually exclude via policy
  3. Flocking/Routing (needs work for our users)
- Ideal: transparent for users who know the requirements of their workload

The End

THANK YOU FOR YOUR TIME!

QUESTIONS? COMMENTS?