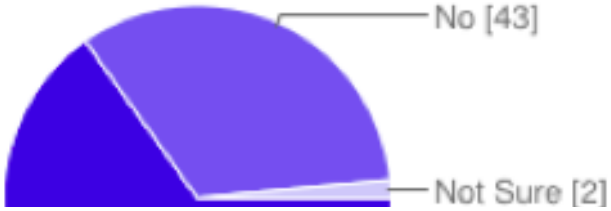
Analysis support issues, Virtual analysis facilities (ie Tier 3s in the cloud)

Doug Benjamin
(Duke University)
on behalf of
Sergey Panitkin, Val Hendrix, Henrik Ohman, Peter Onyisi

How do Atlas physics work today?

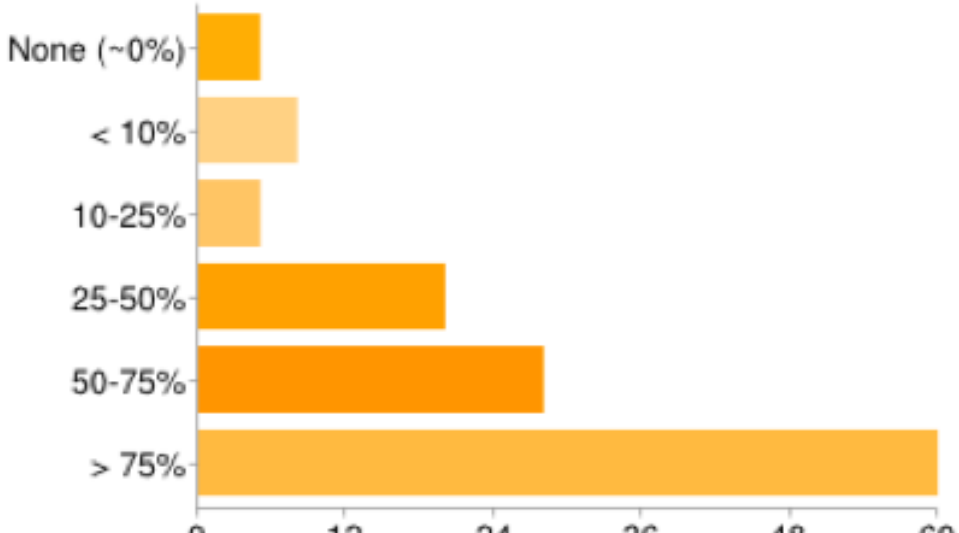
- User surveys were conducted (US ATLAS 145 responses and ATLAS 267)

7) Do you run part of your analysis jobs on the grid?



Yes	**84**	65%
No	**43**	33%
Not Sure	**2**	2%

14) What fraction of your analysis jobs are performed on local resources (not part of Tier 1 or Tier 2 sites)?

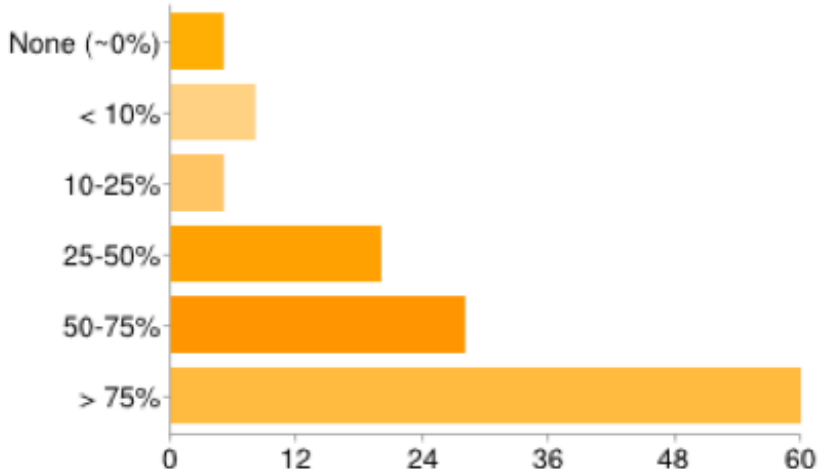


None (~0%)	**5**	4%	%
< 10%	**8**	6%	%
10-25%	**5**	4%	
25-50%	**20**	16%	
50-75%	**28**	22%	
> 75%	**60**	47%	

US ATLAS

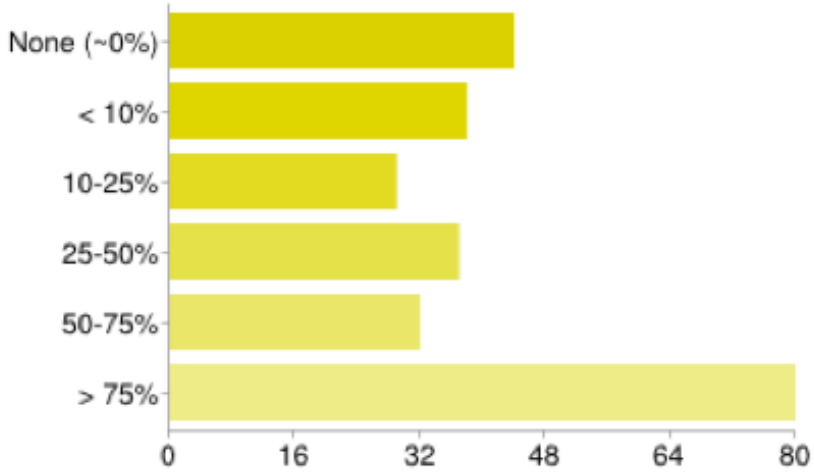
Difference in cultures ATLAS vs US ATLAS (local usage)

14) What fraction of your analysis jobs are performed on local resources (not part of Tier 1 or Tier 2 sites)?



None (~0%)	5	4%
< 10%	8	6%
10-25%	5	4%
25-50%	20	16%
50-75%	28	22%
> 75%	60	47%

10) What fraction of your analysis jobs are performed on local resources (Tier 1 or Tier 2 sites)?



None (~0%)	44	17%
< 10%	38	14%
10-25%	29	11%
25-50%	37	14%
50-75%	32	12%
> 75%	80	30%

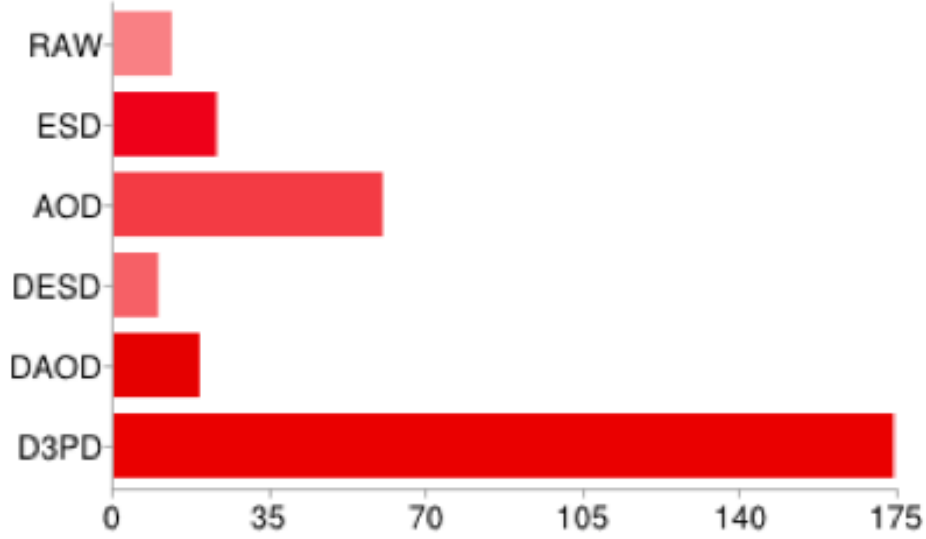
US Analyzers – rely more heavily on local resources than collegues

What data do people now use?

RAW	5	5%
ESD	10	10%
AOD	33	32%
DESD	13	13%
DAOD	8	8%
D3PD	71	70%

People may select more than one checkbox, so percentages may add up to

3) If you or your analysis team create your own private ntuples, which data format do you use for input?

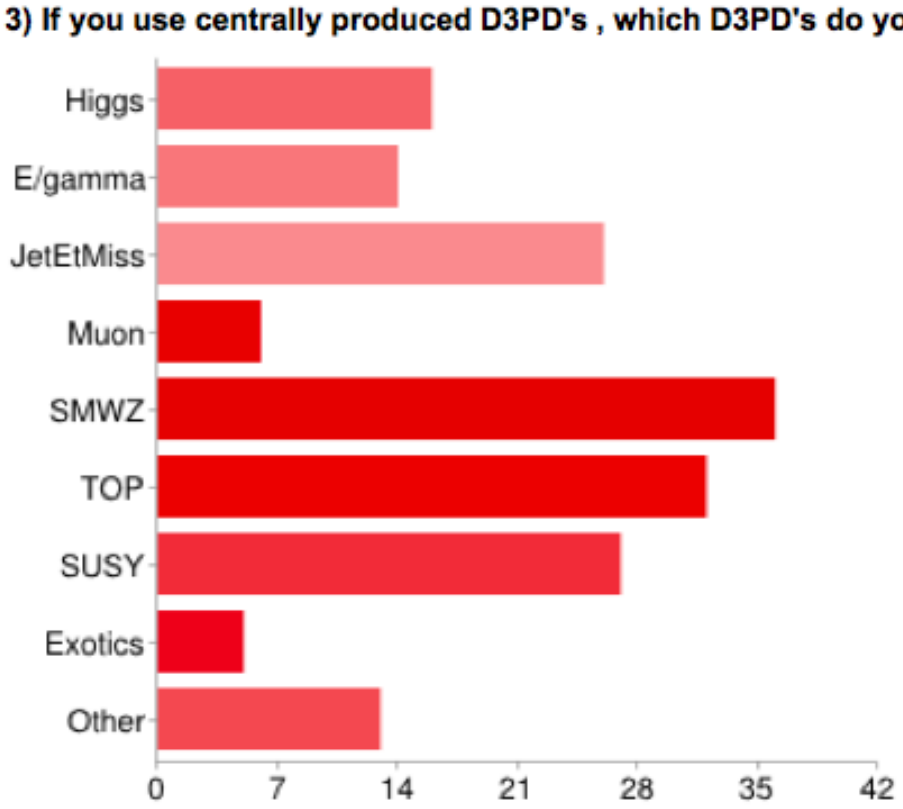


RAW	13	6%
ESD	23	10%
AOD	60	26%
DESD	10	4%
DAOD	19	8%
D3PD	174	77%

People may select more than one checkbox, so percentages may add up to more than 100%.

US and ATLAS analyzers not that different

What types of D3PD's are used?



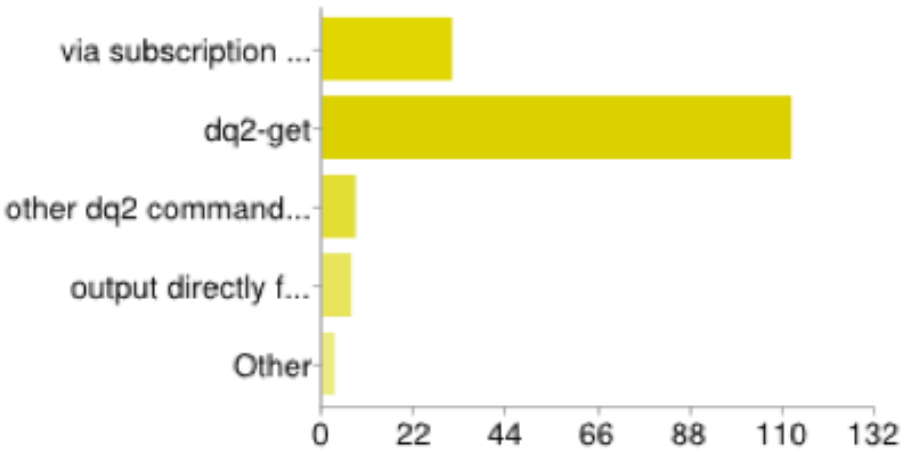
Higgs	**16**	14%
E/gamma	**14**	12%
JetEtMiss	**26**	23%
Muon	**6**	5%
SMWZ	**36**	32%
TOP	**32**	28%
SUSY	**27**	24%
Exotics	**5**	4%
Other	**13**	12%

People may select more than one checkbox, so percentages may add up to more than 100%.

US ATLAS plot - ATLAS wide plot not so different

How do people get their data?

- dq2-get very popular amount analyzers
  - Seen in surveys and dq2 traces



via subscription (DATRI)	**31**	26%
dq2-get	**112**	93%
other dq2 command line tools	**8**	7%
output directly from ganga/panda submisions	**7**	6%
Other	**3**	3%

Moving toward the future

- 2014/2015 will be important time for ATLAS and US ATLAS
  - Machine will be at full energy
  - ATLAS wants to have a trigger rate to tape ~ 1 kHZ (vs ~ 400 Hz now)
  - We need to produce results as fast as before (maybe faster)
  - ARRA funded computing will be 5 years old and beginning to age out
  - US funding of LHC computing likely flat at best.
- Need to evolve the ATLAS computing model
  - Need at least factor of 5 improvement (more data and need to go faster)
- Need to evolve how people do their work
  - Fewer local resources will be available in the future
  - US ATLAS users need to move towards central resources
  - But…. Need to avoid a lost in functionality and performance

Future of Tier 3 sites

- Why do research groups like local computing?
  - Groups like the clusters that they have because now they are easy to maintain (We succeeded!!!)
  - They give them instant access to their resources
  - They give unlimited quota to use what they want for what ever they want
- We need to have a multi-faceted approach to the Tier 3 evolution
- Need to allow and encourage sites with the funding to continue to refresh their existing clusters.
  - We do not want to give the impression in any way that we would no welcome addition DOE/NFS funding to be used in this way
  - What about cast off compute nodes from Tier 1/Tier 2 sites?
- Need to use beyond pledged resources to the benefit of all users and groups of a given country
  - If we can not get sufficient funding to fund the refresh of the existing Tier 3's
  - Beyond Pledged resources are a key component to the solution.

New technologies on horizon

- Federated storage
  - With caching Tier 3's can really reduce data management
- WAN data access
  - User code must be improved to reduce latencies
  - Decouples Storage and CPU
- Analysis Queues using beyond pledged resources
  - Can be used to make use of beyond pledges resources for data analysis not just MC production
- Cloud computing
  - Users are becoming more comfortable for cloud computing (gmail, icloud)
  - Virtual Tier 3's
- All are coupled
- Done right, they should allow us to do more with the resources that we will have

Virtual Analysis clusters(Tier 3's)

- R&D effort by Val Hendrix, Sergey Panitkin, DB and qualification task for student Henrik Ohman
  - Part time effort for all of us, is a issue toward progress
  - Peter Onyisi has just joined University of Texas as a professor and is starting
- We been working on a variety of clouds (by necessity and not necessarily by choice)
  - Public clouds – EC2 (micro spot instances), google (while still free)
  - Research clouds – Future grid (we have an approve project)
  - We are scrounging resouces where we can (not always the most efficient)
- Sergey and I have both independently working data access/handling using Xrootd and the federation
- We could use a stable Private cloud testbed.

Virtual analysis clusters (cont.)

- Panda is used for work load management in most cases (Proof is only exception)
  - Perhaps we can use Proof on demand
- Open issues
  - Configuration and contextualization - We are collaborating with CERN and BNL on puppet - outlook looks promising
  - What is the scale of the virtual clusters?
    - Personal analysis cluster
    - Research group analysis cluster (institution level)
    - Multi institution analysis cluster (many people , several groups to country wide)
  - Proxies for the Panda Pilots (personal or robotic)
  - What are the latency incurred by this dynamic system? Spin up and tear down time?
  - What leave to data reuse do we want (ie caching)?
  - How is the data handling done in and out of the clusters
    - Initial activity used Federation storage as source
  - What is the performance penalty of virtualization

Google Compute Engine

- During the initial free trial period – Sergey got a project approved
  - He was able to extend it through at least into December
  - Review next week on the progress and useage
  - Would like to extend it
- Current status
  - Variety of VM's mostly based on CENTOS 6 x86_64
  - Have a prototype SL 5.8 image (with lots of issues)
  - Have 10 TB persistent storage, upto 1000 instances or 1000 cores
- Proof cluster established
  - 8 core VM's w/ 2 1.7 TB ephermial storage
  - over 60 worker nodes more that 500 cores
  - Centos 6 image used
  - Used by colleagues at BNL primarily

Google Compute Engine (cont)

- Analysis cluster
  - Goal to run Panda jobs producing D3PD's
  - Storage will be provided by BNL spacetokens
    - Local site mover will copy files to VM's for processing
    - Output directed back to BNL
    - Have a robotic certificate (with needed production voms extensions)
  - Tested at Centos 6 VM
    - Could not use the image to clone into other VM and connect via ssh
    - Did contain OSG worknode middleware and required libraries to pass node validation tests.
    - Will need to make new version to fix non function ssh
    - Need to Validate D3PD production on SL 6 platform (not AOD production)
  - CVMFS establish and tested in cluster
  - Condor installed but not fully tested
  - APFv2 needs to be installed and configured
    - Could use help with queue definitions in sched config
  - Within the Last 7 days received a prototype SL 5.8 image from Google
    - Added the needed libraries and configurations but….
    - Had to install and hack code used to create cloned images
    - SSH access does not work --- might have to

Proof on Google Compute Engine

- Sergy Panitkin -

- T3 type Proof/Xrootd cluster on Google Compute Engine
- 64 nodes with ~500 cores
- 180TB of ephemeral storage
- The cluster is intended for ATLAS D3PD based physics analysis.
- Studying data transfer from ATLAS using dq2, xrdcp and also direct read.
- Plan to study Proof workload distribution system scalability at a very large number of nodes.

GCE/Future Grid
Henrik Ohman

- Node configuration
  - Working with puppet for node configurations (cvmfs, xrootd, condor, panda, etc.)
  - Considering benefits and drawbacks of masterful vs. masterless setups - what would be easiest for the researcher who wants to setup his own cluster in the cloud?
- Cluster orchestration
  - Investigating whether puppet can also be used for scaling the size of the cluster
  - http://puppetlabs.com/solutions/google-compute-engine/ (node_gce) *
  - http://puppetlabs.com/blog/building-application-stacks-with-puppet/ (gce_compute)
- This work will be useable on GCE/Future Grid (Cloud Stack) and Amazon Web Services

Amazon Cloud (AWS)

Val Hendrix,Henrik Ohman, DB

- Beate's LBNL group received grant to run physics analysis on Amazon Cloud
- Purpose –

"the analysis using Amazon resources will start from skimmed/slimmed D3PDs and go from there. The would eventually like to do the analysis from DAODs, but I think that would be too expensive in terms of CPU and disk for this particular project." -- Mike Hance.

- limited sum of Grant money and they would like to find another place for them to do their analysis once there Amazon money runs out.

Future Grid
Peter Onyisi

- The UT-Austin ATLAS group is interested in having a "personal cluster" solution deployed on university-hosted cloud resources (the TACC Alamo FutureGrid site).
  - Have our own CentOS-derived disk image that delivers basic functionality but will migrate to more "mainline" images soon.
- We are looking at a very interactive use case (more towards PROOF than PanDA)
  - the really heavy lifting will use T1/T2 resources.
  - Good integration with local storage resources is essential.
  - Remote access to D3PD via federated xrootd is also in the plan.

Conclusions

- New technologies are maturing that can be used and have the potential of being transformative
- So additional resources from the facilities would go a long way toward helping
- Virtual Tier 3 are making slow process
  - New labor has been added