



ANL T3 Meeting Report

R. Yoshida

Argonne National Laboratory

US ATLAS Tier-2/Tier-3 Meeting,

Arlington, 11 Nov 2009

T3 Meeting at ANL 29,30 Oct



➤ Aims of the ANL T3 meeting:

- ❑ Bring together people who will be involved with T3's at their home institutes.
- ❑ Inform people about the current state of T3's.
- ❑ Discuss and agree on the goals and strategies for T3's.
- ❑ Come up with a definite set of plans and timetables
- ❑ Get people who are interested in working on T3's; and start work.
- ❑ Talks are at <http://indico.cern.ch/conferenceDisplay.py?confId=71202>



52 people registered
About 1/2 in person
About 1/2 on EVO

**(EVO was “ok” according to
post-meeting survey)**

US ATLAS Tier-3



- The aim of Tier-3 coordination (Doug and me) is to maximize the effectiveness of US ATLAS Tier-3's for physics analysis. We are charged with **coordinating** the efforts of the institutes to bring up Tier-3's, operate them and integrate them into the US ATLAS computing system.
- **Tier-3 computing** and **Tier-3 resources** here means analysis computing resources which are guaranteed to be available to the institute members to do their analysis. (Not shared resources like T2 analysis queues)
- First order of business was to determine the current status of T3 computing at the US ATLAS institutes.
- While some compilation of institute resources were available, we felt we needed more in-depth information: thus the telephone interviews with every US ATLAS institute we conducted.
- This information we found forms one of the basis of T3 planning.
- **So what did we find? →**

Some General Numbers



- The total number of potential analyzers in US ATLAS: ~500.
 - ❑ This is a head count, not FTE
 - ❑ In this context “analyzer” means someone who intends to do analysis computing of some kind.

- The number of analyzers per institute for 2010 ranges from 0 to 27.

- Institutes typically plan to have 30-40% of its analyzers at CERN.
 - ❑ Most institutes plan to use the T3 resources at their institutes—even for those stationed at CERN.

- The average T3 resources per analyzer for 2010 in US ATLAS is about 5-10 “cores”
 - ❑ Cores is used as a convenient measure of resources. It should be noted that the analyzer will have access to commensurate amount of disk space
 - ❑ These numbers are very rough. We made no attempt to convert different type processors to some “spec” units.
 - ❑ This assumes that new funding (mostly ARRA) is effectively used.

Existing Tier-3's



- **T3's associated with T2 or T1:**
 - ❑ There are 8 institutes whose T3 resources are closely coupled to a T2 or a T1 operations at the same institutes. The analyzers at these institutes have a number of slots dedicated for T3 computing.

- **T3gs:**
 - ❑ T3gs is a T3 which are full Grid sites capable of accepting Grid jobs from outside. There are 4 institutes which expect to operate T3gs's in 2010.

- **T3g:**
 - ❑ T3g is a T3 which is connected to the Grid to receive data but does not accept jobs from outside. There are 12 institutes who operate this type of T3, currently.
 - ❑ The setup of these T3's vary enormously. 4 of these institutes take advantage of the University of Departmental computing service or cluster. The others are standalone.
 - ❑ Most of these plan an expansion in 2010 (based on ARRA funds)

Institutes without Tier-3's



- There are 20 institutes without T3 capabilities.
- 12 Institutes plan to build a T3 cluster from scratch in 2010. Most of these are funded by ARRA. The average funding is ~\$30-40k.
- The other 8 institutes do not currently have funds for T3's in 2010. Four of these may receive funding from the current NSF application.

36 T3's expected total

Analysis Facilities



- We surveyed the interest in the proposed Analysis Facilities at BNL and SLAC.
 - ❑ Five institutes (not including BNL and SLAC) said they are interested in committing, or have already committed, resources (~\$10k-\$50k) to BNL or SLAC.
 - ❑ When asked whether the use of AF was of interest independent of contributions, seven institutes (in addition to the five above) said yes.

Other needs, interests



- Six institutes are already running or expressed interest in Proof.
- Seven institutes indicated that they needed to work at the ESD or Raw data level and thus needed access to Conditions DB. (Solution exists. Frontier/Squid)

Tier-3 “volunteers”



- Finally, many expressed willingness to contribute to the overall T3 effort.

Plan of the meeting



Thursday 29 October 2009

[top](#)

09:30	Welcome to ANL ASC (10')  Slides  )	Rik Yoshida
09:40	US ATLAS Tier-3 Status (30')	Rik Yoshida
10:10	Tier-3 Commissioning and Integration plans (30')	Doug Benjamin (<i>Duke University</i>)
10:40	break (15')	
10:55	US ATLAS Tier-3 Support from OSG (30')	Dan Fraser
11:25	Timescale and Milestone (proposals) (15')	
11:40	Discussion (20')	

Lay out the status and our current plans on how to proceed on broad terms.

Discussion:

- Is this the right direction?**
- Are we missing something?**

Plan of the meeting



Thu. afternoon

13:15	Tier-3g Architecture (20')	Doug Benjamin (<i>Duke University</i>)
13:35	XROOTD for T3 Batch System (30')	Andrew Hanushevsky (<i>STANFORD LINEAR ACCELERATOR CENTER</i>)
14:05	Tier-3 Storage Element Status (30')	Hironori Ito
14:35	break (15')	
14:50	Storage discussion (30')	
15:20	Experience and Benchmarking with Virtual Machines (30')	Harris Kagan (<i>Ohio State University</i>)
18:30	Dinner at Argonne Guest House (2h00')	

General technical outline the proposed T3g (Doug)

Using XROOTD for T3 (Andy)

T3 Storage Element (Hiro)

Discussion:

- Is the storage concept right?
- Is the integration strategy of S.E. to US ATLAS computing right?

Virtual Machines (Harris) Maybe the future direction for T3's

Plan of the meeting



Friday 30 October 2009

[top](#)↑

09:00	Condor and ArCond as a T3 Batch System (40')	Sergei Chekanov (<i>HEP division (ANL)</i>)
09:40	Condor, Proof and PQ2 (30')	Neng Xu (<i>Department of Physics - University of Wisconsin</i>)
10:10	break (15')	
10:25	Proof and Xrootd for T3 (15')	Sergey Panitkin (<i>Department of Physics - Brookhaven National Laboratory (BNL)</i>)
10:40	Using SFrame and Proof (30')	Attila Krasznahorkay (<i>New York University</i>)
11:10	XrootD, XrootDFS and BeStMan-Gateway (30')	Wei Yang
11:40	Discussion (15')	

In order to process large amounts of data (TB size), we need to parallelize: How do we do this without bringing in the heavy machinery of panda, dcache etc. to T3?

**Condor and ArCond (Sergei) How this is being done at ANL
Developments at Wisconsin (Neng)
Proof and XrootD experience at BNL (Sergey)
Using Proof at NYU (Attila)
How XrootD and T3 S.E (BeStMan-Gateway) works (Wei)**

Plan of the meeting



Friday afternoon

13:15	Cluster Management (30')	John Brunelle (<i>Harvard University</i>)
13:45	Cluster Configuration Management (30')	Saul Youssef
14:15	Timescales and Milestones (rediscussion) (30')	

Tools for managing clusters (John and Saul)

Discussion: Do we have more ideas after having heard the technical talks?

General Conclusions



- T3 sites (8) closely associated with T2 and T1 tends to be well-supported due to the sharing of infrastructure.
- T3gs sites (4) run the same services as a T2 and has nearly the same support load.
- Most other sites are of such size and available manpower that T3g is probably the appropriate solution.
- One desirable thing T3g lacks currently is the ability to subscribe to the data from the Grid. (Four T3 SRM's up now. Being tested) This will be the primary coupling to T2/T1.

General Conclusions



- Concentrate on T3g for now.
- Need a complete T3g solution for (12) institutes starting from scratch. A standard solution is desirable.
- The same solution should be largely useful also for the existing T3g sites.

Technical Conclusions



- Details in Doug's talk
- General T3g scheme:
 - ❑ Interactive cluster
 - ❑ Batch system. Baseline: Condor. Accommodate other batch sys.
 - Data storage: distributed or separate
 - XrootD
 - Front end for Condor submission: "ArCond"
 - ❑ "Storage Element": SRM BeStMan Gateway (or bare gridFTP?)
- Future Directions:
 - ❑ Virtual Machines
 - ❑ Web File system: CVMFS
 - ❑ PROOF

Rough Timescale



- ARRA funds are expected to be available beginning next year. (January at the earliest)
- Need a set of instructions for building a T3g which is complete on the same timescale.
 - But we think T3g with appropriate characteristics can be designed on that time scale.
 - 80% of what we need already exists in principle (needs a lot of packaging to be easy to install and use)
 - Scalable to different budgets.
 - Need to work together to pull this off in this time scale
 - Dell ATLAS matrix with attractive pricing means we have a reasonable chance of unifying the hardware to a large extent as well.

Since the meeting



- DB and RY have worked with OSG, VTD, Condor, Glow teams at Madison to hammer out the cluster commissioning procedure.
- New T3g setup Twiki pages being put together.
- Beginning to set up a T3 “integration cluster” at ANL.
- 16 people offered to help with T3 project
- Had organizational meeting last Friday (more from Doug)
- Progress meeting in 1.5 weeks.

Backup



Tier-3-why it's needed



Tier 2 resources

- ▶ 50%,
centrally managed for simulation
- ▶ 50%
for national analyses
- ▶ How much full simulation?
30% → 20% → 10%

US Pledge to wLCG	2007	2008	2009	2010	2011
CPU (kSI2k)	2,560	4,844	7,337	12,765	18,194
Disk (TB)	1,000	3,136	5,822	11,637	16,509
Tape (TB)	603	1,715	3,277	6,286	9,820

6000
for analysis

Athena runs
at 5-10 Hz
per cpu

Sample	Generation	Simulation	Digitization	Reconstruction
Minimum Bias	0.0267	551.	19.6	8.06
$t\bar{t}$ Production	0.226	1990	29.1	47.4
Jets	0.0457	2640	29.2	78.4
Photon and jets	0.0431	2850	25.3	44.7
$W^\pm \rightarrow e^\pm \nu_e$	0.0788	1150	23.5	8.07
$W^\pm \rightarrow \mu^\pm \nu_\mu$	0.0768	1030	23.1	13.6
Heavy ion	2.08	56,000	267	-

Table 18. in kSI2k-s, without pileup

K. Assamagan, et al., ATLAS Monte Carlo Project, 2009.

Tier-3 why it's needed



Tier 2 simulation for one year

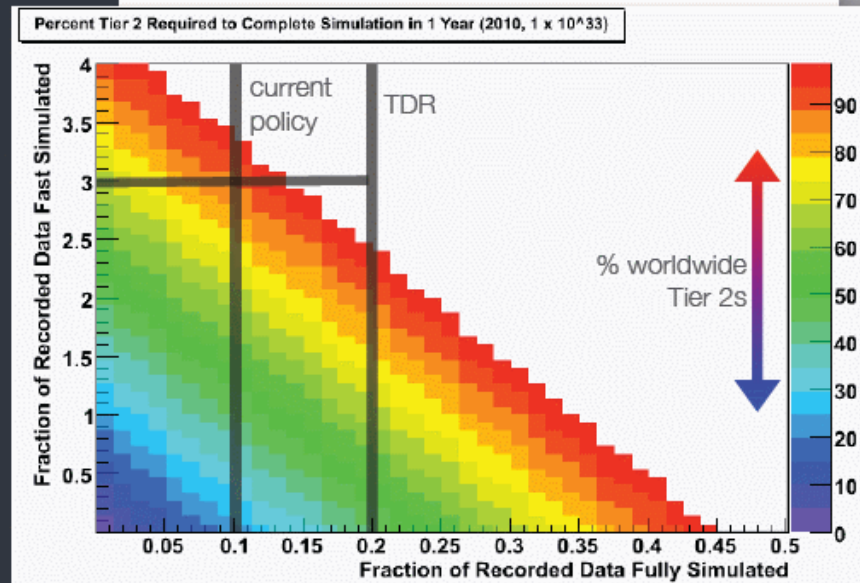
► horizontal axis:

fraction fully simulated

► vertical axis:

fraction fast-simulated

(ATLFAST-II...from Assamagan)



Is 10%
full sim.
and 100%
fast sim
enough?

We're
supposed
to be here