Efforts to Build a T3gs at Illinois

Mark Neubauer, Dave Lesny
University of Illinois at Urbana-Champaign

US ATLAS Distributed Facilities Meeting
LIGO Livingston Observatory

March 3, 2009

Tier-3 Task Force and Site Structure

Key recommendation of Tier-3 Task Force (T3TF):

(my paraphrasing)

- US ATLAS computing infrastructure should be deeper than the Tier1+2's and accommodate a set of 4 possible Tier-3 architectures: T3gs, T3g, T3w, and T3af.
  - The goal is to utilize significant, additional computing resources to produce a system which is flexible and nimble enough to weather the realities of ATLAS analysis once the data starts flowing from CERN.
  - The Tier-2's should be considered a finite and precious resource that must be used efficiently with as much end-user analysis as feasible done on the local Tier-3's

What is a T3gs?

T3gs: Tier-3 w/ Grid Services

- Cluster w/ software deployed for pAthena jobs, DQ2 services, and possibly LFC
- Expected to have "significant" CPU and storage resources:
    - ≈ several hundred CPU cores
    - ≈ several tens of TB
- Expected to a level of infrastructure (power, cooling), networking, and administrative support commensurate with available resources

Functionally-equivalent to Tier-2 from a services perspective

Distinct from Tier-2 in important ways:

- Size (not likely to be as large at Tier-2 in terms of available resources)
- Local policy control (resources funded at institutional-level)

Why bother trying to deploy a T3gs?

Value Added From a T3gs

1) Leverage University and Lab-based computing to increase overall capacity for production tasks like Data Format Production (e.g. D^1PD ➔ D^2PD) and Monte Carlo production

   - Given the specific T3 nature of T3gs, this is expected to be during intensive periods of critical need (e.g data reprocessing, MC before conference seasons)

   - T3gs with their T2-like configuration and capabilities in best position to pitch in when needed

2) Offload from Tier-2's computationally intensive jobs that do not need high-bandwidth access to data (e.g. ME calculations, MVAs, pseudo-expt generation for systematic evaluations). Generically true for any Tier-3 site (T3gs could also offload data-intensive jobs given its nature)

See the T3TF Document for some quantitative arguments regarding 1) and 2)

3) Spread around the wealth (of ATLAS computing knowledge)

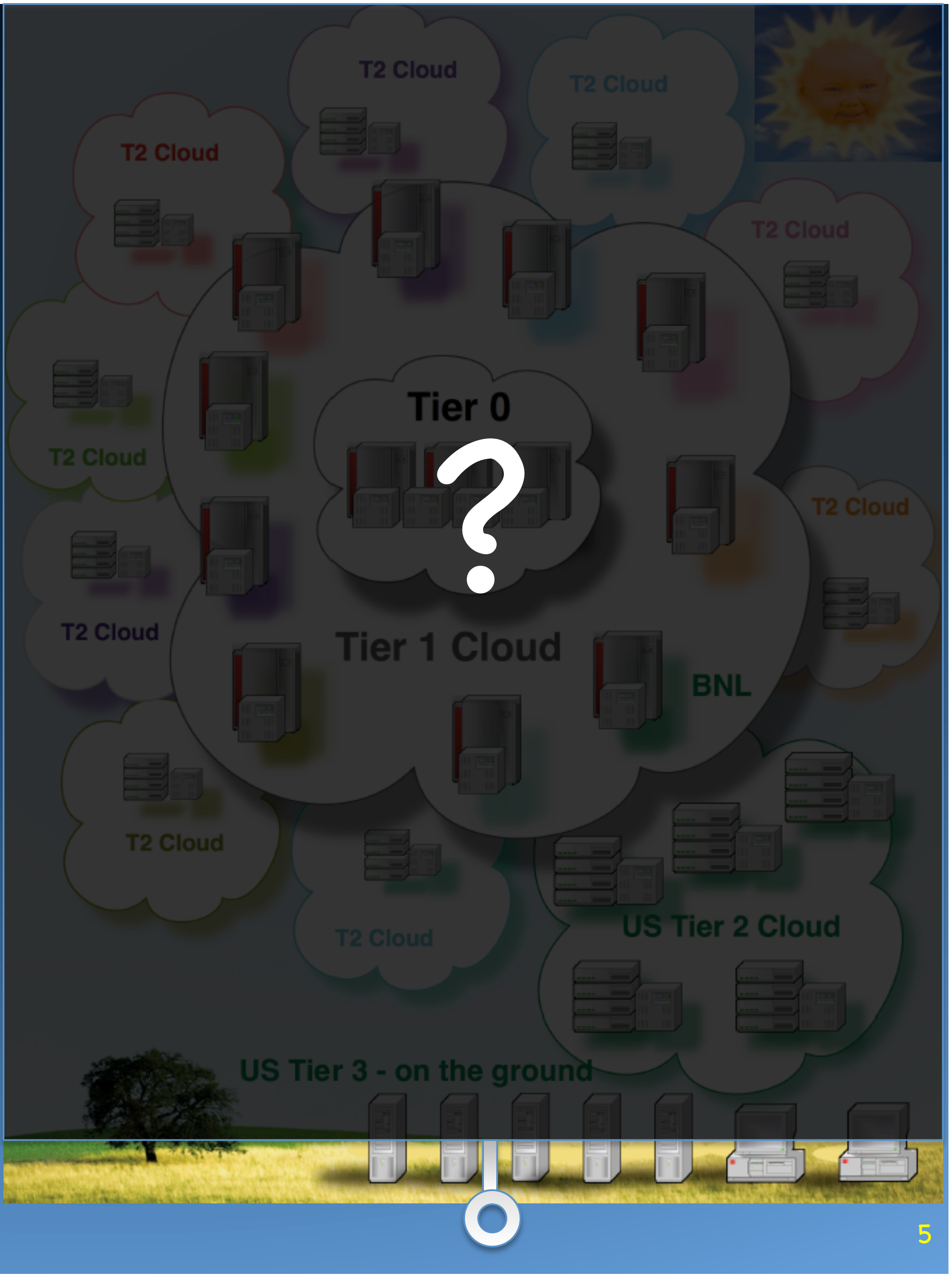
My candid impression in trying to get our T3gs (T2-like) site together:

- Much of the documentation resides with a few experts (T2 site admins, etc). Some is contained within a disparate set of Twiki pages, NNs and disseminated when need arises through email, HN, etc.

-The experts have been very willing to help out when asked!

- To an outsider, this can appear to be a very large, very black box (I know that most don't need to know how this sausage is made..)

Main point: there is some value for outsiders like us to try to replicate a T2-like site, to help shed some light on how all of this works (widen knowledge base for future T3 installations)



The IllinoisHEP Grid Site

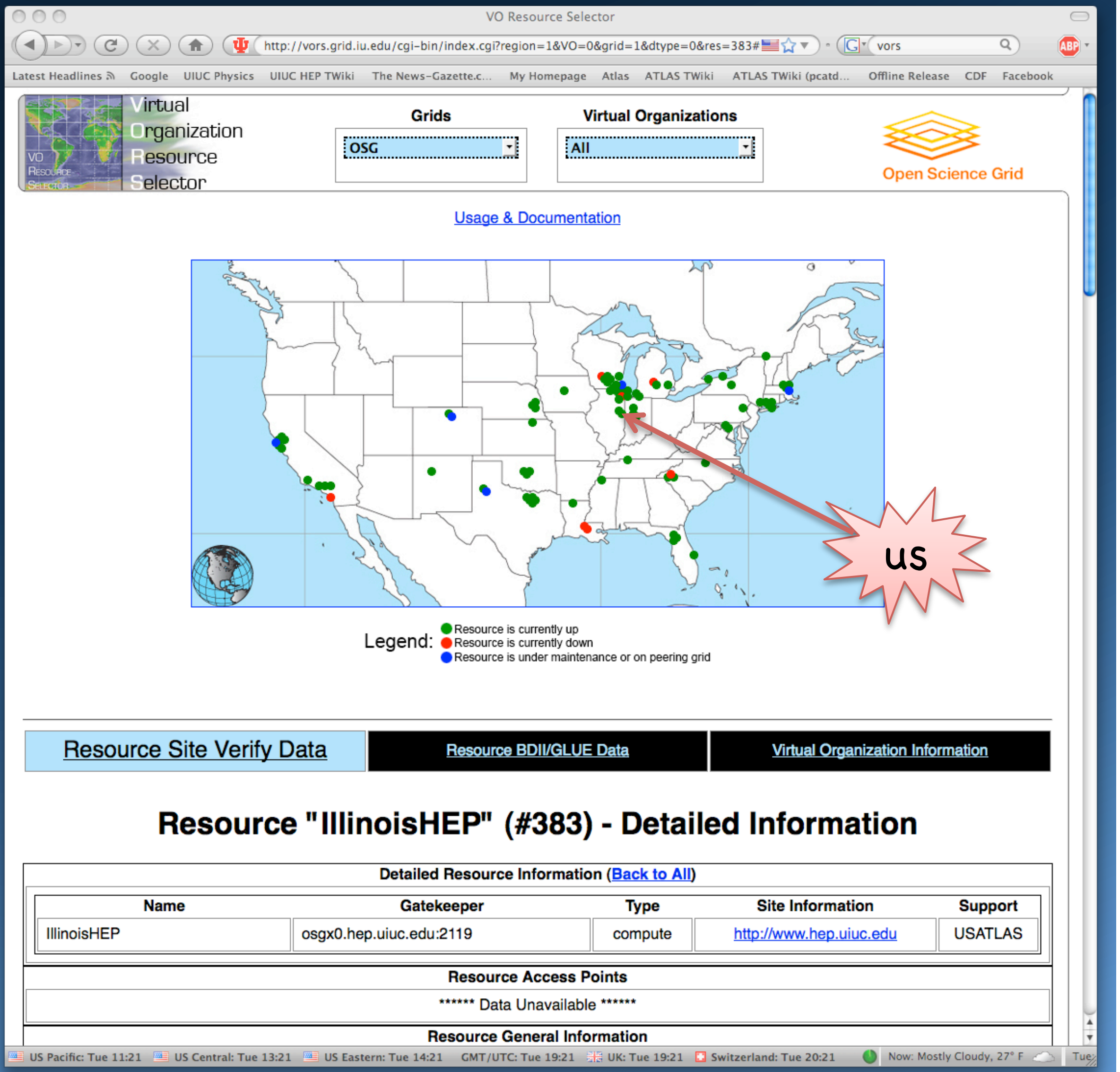
We've been participating OSG site for ≈ 1 yr

We've been plugged into the Panda job mgmt on a similar time scale

- We've participated (not on purpose!) in M* reprocessing
- Done FDR data analysis on our site (on purpose)

Up until a month or so ago, we've been peer-configured to utilize the IU SE

- We've basically been running as a T3g before we came up with the name!



Legend:
- 🟢 Resource is currently up
- 🔴 Resource is currently down
- 🔵 Resource is under maintenance or on peering grid

Resource "IllinoisHEP" (#383) - Detailed Information

Detailed Resource Information (Back to All)

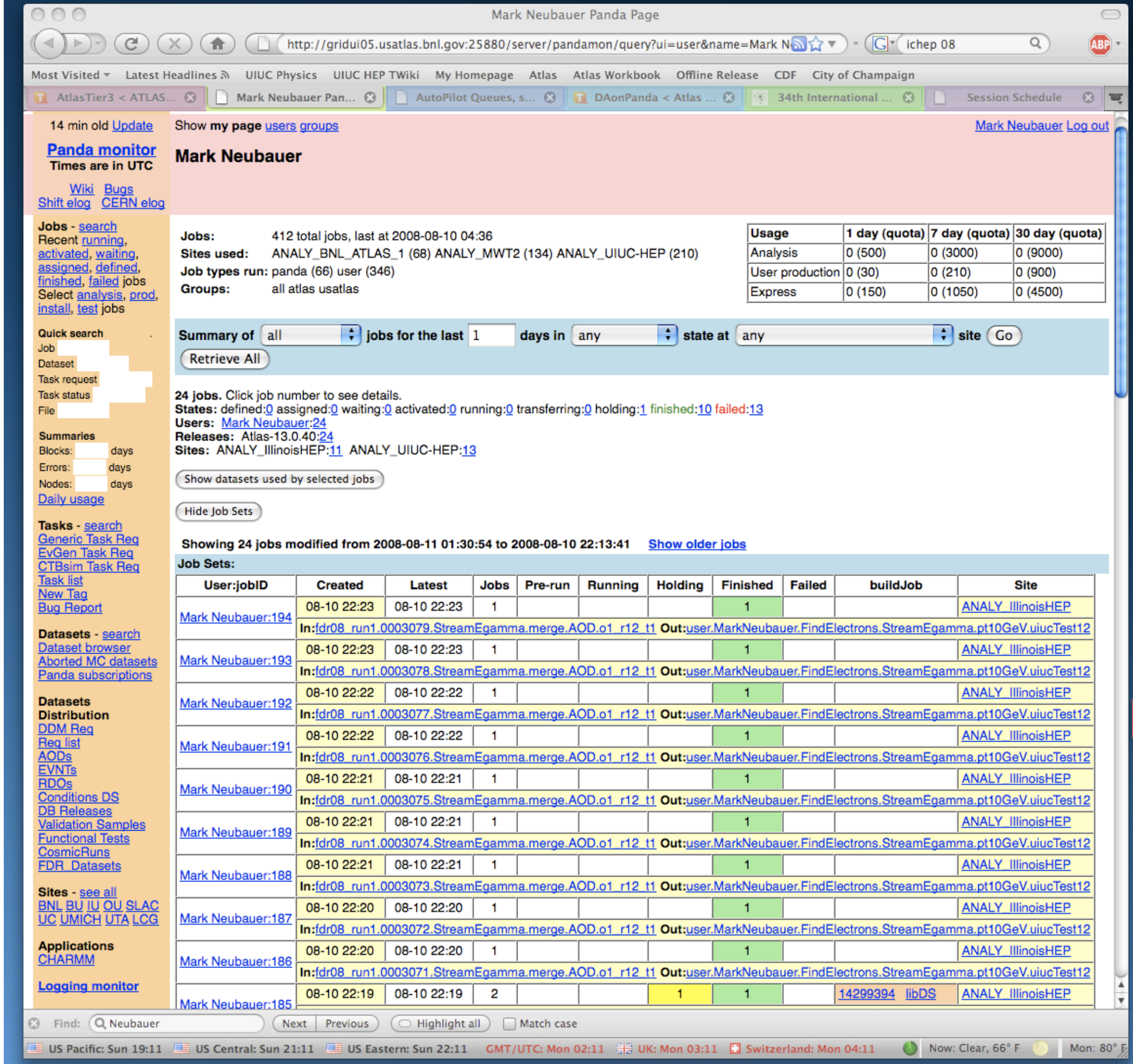
Name	Gatekeeper	Type	Site Information	Support
IllinoisHEP	osgx0.hep.uiuc.edu:2119	compute	http://www.hep.uiuc.edu	USATLAS

Resource Access Points

****** Data Unavailable ******

Resource General Information

IllinoisHEP in Action


Mark Neubauer Panda Page
http://gridui05.usatlas.bnl.gov:25880/server/pandamon/query?ui=user&name=Mark N
ichep 08
Most Visited · Latest Headlines · UIUC Physics · UIUC HEP TWiki · My Homepage · Atlas · Atlas Workbook · Offline Release · CDF · City of Champaign
AtlasTier3 < ATLAS... · Mark Neubauer Pan... · AutoPilot Queues, s... · DAonPanda < Atlas ... · 34th International ... · Session Schedule

14 min old Update

Panda monitor
Times are in UTC

Wiki Bugs
Shift elog CERN elog

Jobs - search
Recent running, activated, waiting, assigned, defined, finished, failed jobs
Select analysis, prod, install, test jobs

Quick search
Job
Dataset
Task request
Task status
File

Summaries
Blocks: days
Errors: days
Nodes: days
Daily usage

Tasks - search
Generic Task Req
EvGen Task Req
CTBsim Task Req
Task list
New Tag
Bug Report

Datasets - search
Dataset browser
Aborted MC datasets
Panda subscriptions

Datasets Distribution
DDM Req
Req list
AODs
EVNTs
RDOs
Conditions DS
DB Releases
Validation Samples
Functional Tests
CosmicRuns
FDR_Datasets

Sites - see all
BNL BU IU OU SLAC UC UMICH UTA LCG

Applications
CHARMM

Logging monitor

Show **my page** users groups — Mark Neubauer Log out

Mark Neubauer

Jobs: 412 total jobs, last at 2008-08-10 04:36
Sites used: ANALY_BNL_ATLAS_1 (68) ANALY_MWT2 (134) ANALY_UIUC-HEP (210)
Job types run: panda (66) user (346)
Groups: all atlas usatlas

Usage	1 day (quota)	7 day (quota)	30 day (quota)
Analysis	0 (500)	0 (3000)	0 (9000)
User production	0 (30)	0 (210)	0 (900)
Express	0 (150)	0 (1050)	0 (4500)

Summary of all **jobs for the last** 1 **days in** any **state at** any **site** Go
Retrieve All

24 jobs. Click job number to see details.
States: defined:0 assigned:0 waiting:0 activated:0 running:0 transferring:0 holding:1 finished:10 failed:13
Users: Mark Neubauer:24
Releases: Atlas-13.0.40:24
Sites: ANALY_IllinoisHEP:11 ANALY_UIUC-HEP:13

Show datasets used by selected jobs

Hide Job Sets

Showing 24 jobs modified from 2008-08-11 01:30:54 to 2008-08-10 22:13:41 Show older jobs

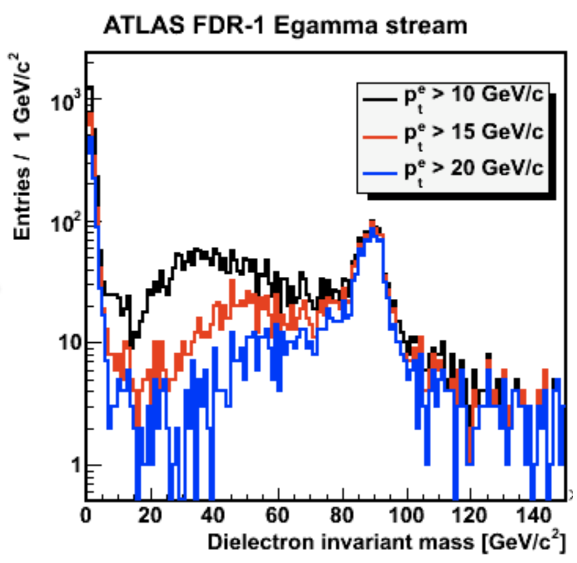
Job Sets:

User:jobID	Created	Latest	Jobs	Pre-run	Running	Holding	Finished	Failed	buildJob	Site
Mark Neubauer:194	08-10 22:23	08-10 22:23	1				1			ANALY_IllinoisHEP
	In:fdr08_run1.0003079.StreamEgamma.merge.AOD.o1_r12_t1 **Out:**user.MarkNeubauer.FindElectrons.StreamEgamma.pt10GeV.uiucTest12									
Mark Neubauer:193	08-10 22:23	08-10 22:23	1				1			ANALY_IllinoisHEP
	In:fdr08_run1.0003078.StreamEgamma.merge.AOD.o1_r12_t1 **Out:**user.MarkNeubauer.FindElectrons.StreamEgamma.pt10GeV.uiucTest12									
Mark Neubauer:192	08-10 22:22	08-10 22:22	1				1			ANALY_IllinoisHEP
	In:fdr08_run1.0003077.StreamEgamma.merge.AOD.o1_r12_t1 **Out:**user.MarkNeubauer.FindElectrons.StreamEgamma.pt10GeV.uiucTest12									
Mark Neubauer:191	08-10 22:22	08-10 22:22	1				1			ANALY_IllinoisHEP
	In:fdr08_run1.0003076.StreamEgamma.merge.AOD.o1_r12_t1 **Out:**user.MarkNeubauer.FindElectrons.StreamEgamma.pt10GeV.uiucTest12									
Mark Neubauer:190	08-10 22:21	08-10 22:21	1				1			ANALY_IllinoisHEP
	In:fdr08_run1.0003075.StreamEgamma.merge.AOD.o1_r12_t1 **Out:**user.MarkNeubauer.FindElectrons.StreamEgamma.pt10GeV.uiucTest12									
Mark Neubauer:189	08-10 22:21	08-10 22:21	1				1			ANALY_IllinoisHEP
	In:fdr08_run1.0003074.StreamEgamma.merge.AOD.o1_r12_t1 **Out:**user.MarkNeubauer.FindElectrons.StreamEgamma.pt10GeV.uiucTest12									
Mark Neubauer:188	08-10 22:21	08-10 22:21	1				1			ANALY_IllinoisHEP
	In:fdr08_run1.0003073.StreamEgamma.merge.AOD.o1_r12_t1 **Out:**user.MarkNeubauer.FindElectrons.StreamEgamma.pt10GeV.uiucTest12									
Mark Neubauer:187	08-10 22:20	08-10 22:20	1				1			ANALY_IllinoisHEP
	In:fdr08_run1.0003072.StreamEgamma.merge.AOD.o1_r12_t1 **Out:**user.MarkNeubauer.FindElectrons.StreamEgamma.pt10GeV.uiucTest12									
Mark Neubauer:186	08-10 22:20	08-10 22:20	1				1			ANALY_IllinoisHEP
	In:fdr08_run1.0003071.StreamEgamma.merge.AOD.o1_r12_t1 **Out:**user.MarkNeubauer.FindElectrons.StreamEgamma.pt10GeV.uiucTest12									
Mark Neubauer:185	08-10 22:19	08-10 22:19	2			1	1		14299394 libDS	ANALY_IllinoisHEP

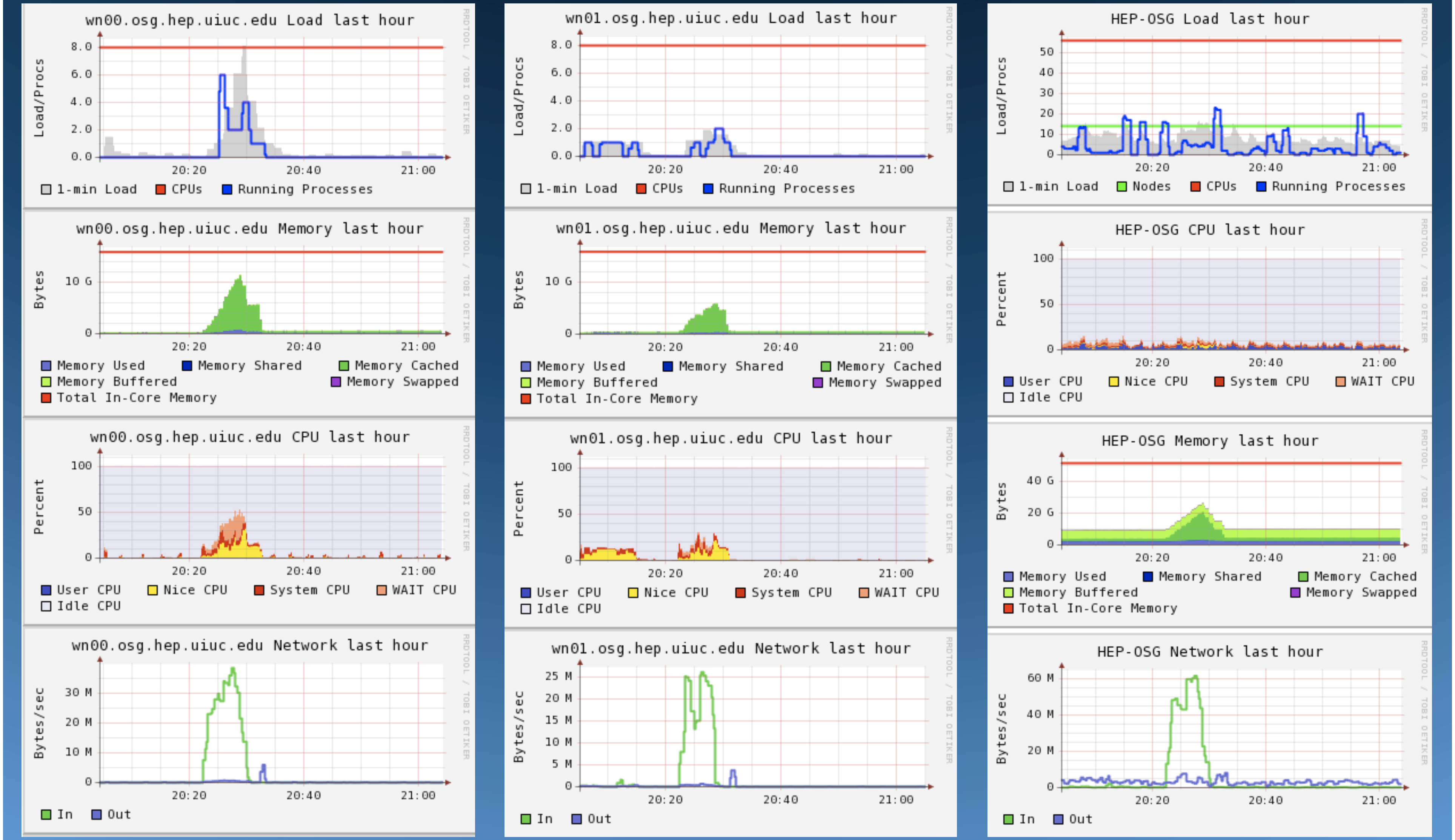
Find: Neubauer
Next
Previous
Highlight all
□ Match case
US Pacific: Sun 19:11 · US Central: Sun 21:11 · US Eastern: Sun 22:11 · GMT/UTC: Mon 02:11 · UK: Mon 03:11 · Switzerland: Mon 04:11 · Now: Clear, 66° F · Mon: 80° F

Fully integrated into the pATHENA ATLAS job control infrastucture.

Analysis job running over the full FDR dataset to the IllinoisHEP Tier-3 site



IllinoisHEP in Action (Ganglia Monitoring)



Infrastructure & Support at Illinois

There is a great deal of infrastructure on campus for scientific HPC

- National Center for Supercomputing Applications (NCSA)
  - Enterprise-quality building in terms of space, power, cooling, networking
  - Blue Waters in 2011: World's 1st sustained Petaflop computer (likely of limited use for our particular applications)
- 10 Gbps connectivity to ICCN, Internet-2, … w/ access from our Physics Building (Loomis Lab) and NCSA via a "Campus Research Network" designed to avoid bottlenecks in the campus firewalls for data-intensive research (we're the poster child for this)
- Loomis Lab has sufficient space, power, cooling, networking for a few full racks of compute nodes and storage devices
- Senior Research Physicist (Dave Lesny): The one who put all of this together
- I have some past experience with scientific computing in putting together and operating the CDF CAF

Deployment Strategy

Note: We are not a large site at this point in time

- 16 CPU cores and 20 TB of disk

Our focus has not been on deploying a large amount of hardware, but rather deploying the full set of services required to make our site a functional T3gs

- Some of our "utility" hardware needs to be upgraded/replaced
- We've deployed extensive monitoring (via Ganglia) to discover bottlenecks
- Leave scaling of CPU cores and storage until later

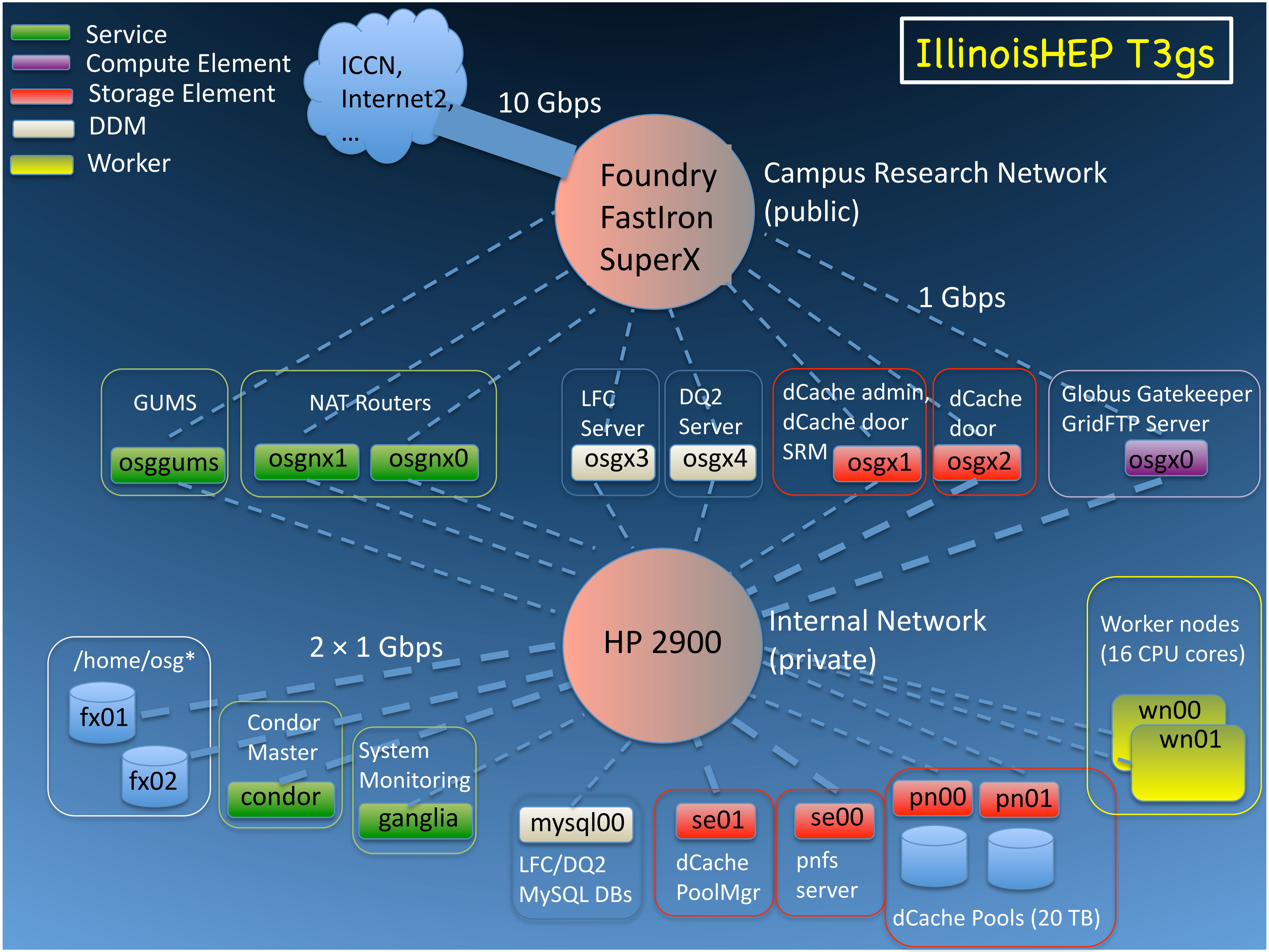
We've chosen to deploy our IllinoisHEP site in Loomis Lab rather than NCSA in order to have "full" control over what we do in the development stage

Once our T3gs is in production and scaled to a reasonable level, we would like to consider replicating this site to NCSA within a partnership that we are trying to foster (Neubauer NCSA faculty fellow)

In other words, we would like to house a future "substantial" procurement in NCSA

- This depends upon a number of factors like a willing partnership with NCSA, demonstrated utility of the IllinoisT3gs, and external $$

IllinoisHEP T3gs



- 🟩 Service
- 🟪 Compute Element
- 🟥 Storage Element
- ⬜ DDM
- 🟨 Worker

ICCN, Internet2, …

10 Gbps

Foundry FastIron SuperX

Campus Research Network (public)

1 Gbps

GUMS: osggums

NAT Routers: osgnx1, osgnx0

LFC Server: osgx3

DQ2 Server: osgx4

dCache admin, dCache door SRM: osgx1

dCache door: osgx2

Globus Gatekeeper GridFTP Server: osgx0

HP 2900

Internal Network (private)

2 × 1 Gbps

/home/osg*: fx01, fx02

Condor Master: condor

System Monitoring: ganglia

mysql00: LFC/DQ2 MySQL DBs

se01: dCache PoolMgr

se00: pnfs server

pn00, pn01: dCache Pools (20 TB)

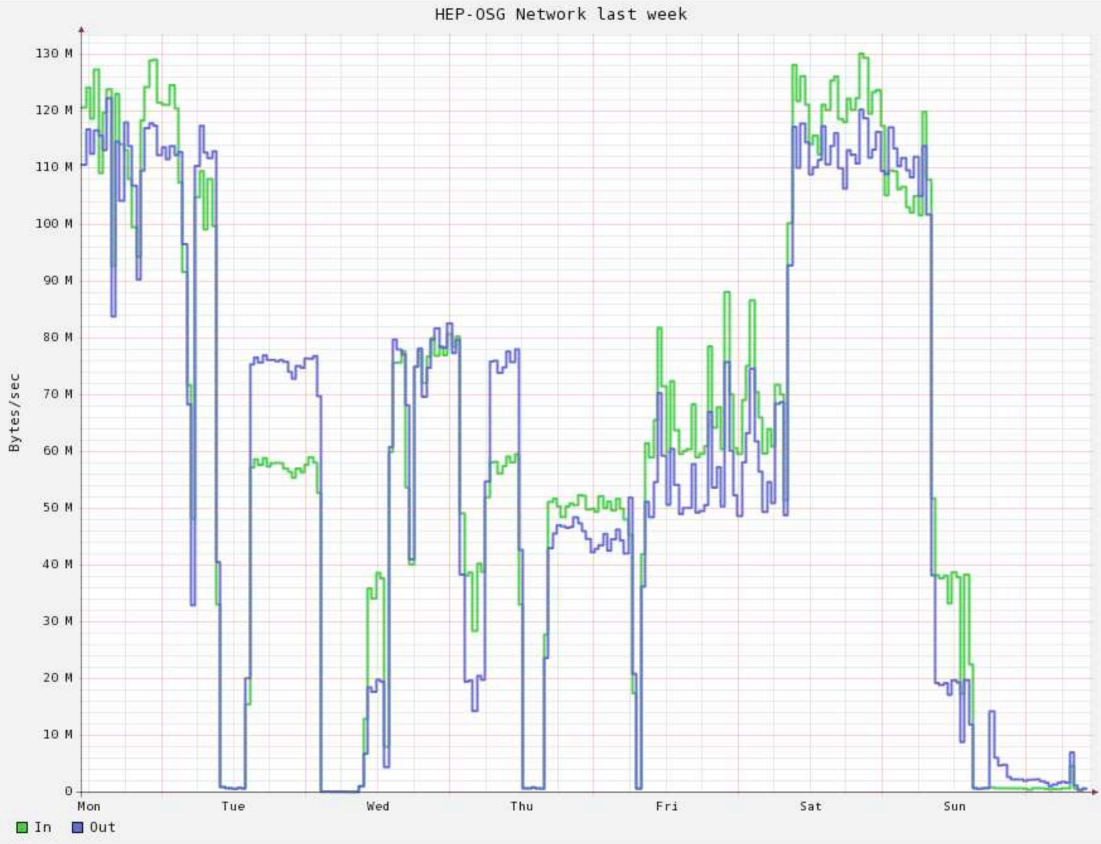
Worker nodes (16 CPU cores): wn00, wn01

Some Baby Pictures

SE and LFC Tests

Dave Lesny has been performing stress tests of our Storage Element and LFC by using the SRM interfaces to write/read files and register the files into our LFC

Hiro has also performed tests of our DDM configuration by doing simple FTS transfers between BNL and IllinoisHEP



Current IllinoisHEP T3gs Status/Plans

As far as we know, we are fully configured as a T3gs site with the required services (Panda-aware, DQ2 site services, LFC, etc)

There are a (small?) number of configuration questions still to be ironed out like our DQ2 site name (IllinoisHEP or UIUC) and whether or not the Grid can handle more combinations of UI, IU, UC, ...

At this point, I think we are ready for more high-level tests (e.g. DQ2 subscriptions and full chain w/ Panda jobs)

There are a number of policy questions that need to be answered about how the T3gs and the other sites that the T3TF recommend should be utilized within US ATLAS. Hopefully some will get answered at this workshop

It would be fantastic for US ATLAS if we had Tier-3 sites such that their resources could be dynamically allocated to accommodate local analysis needs and pitch in to help during times of intensive need for processing power within ATLAS.

Happy sqrt() day to all!