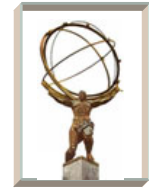


US Tier 2 issues



Jim Shank

Boston University

U.S. ATLAS Tier 2 Meeting

Harvard University

17-18 August, 2006

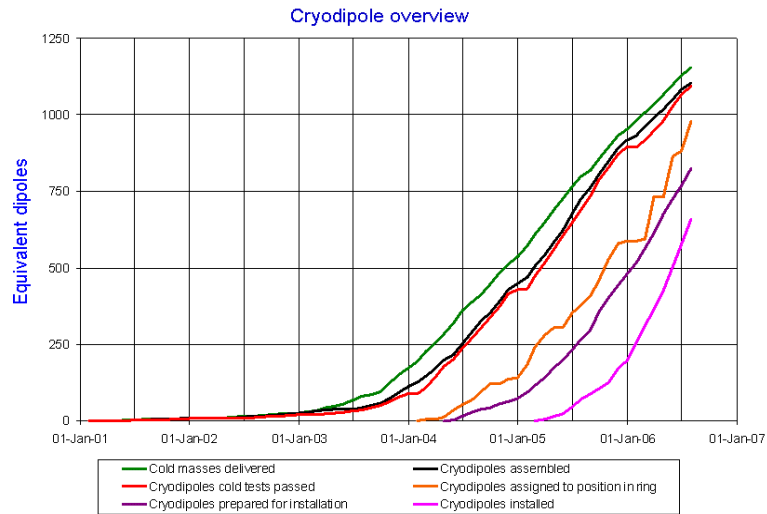
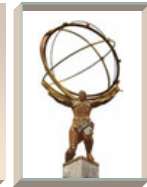
Overview



- LHC Schedule news and implications
- News from the DOE/NSF Review of 15 August
- Goals for this workshop
 - T2 Planning Wiki
 - Action Items from last T2 meeting in Chicago

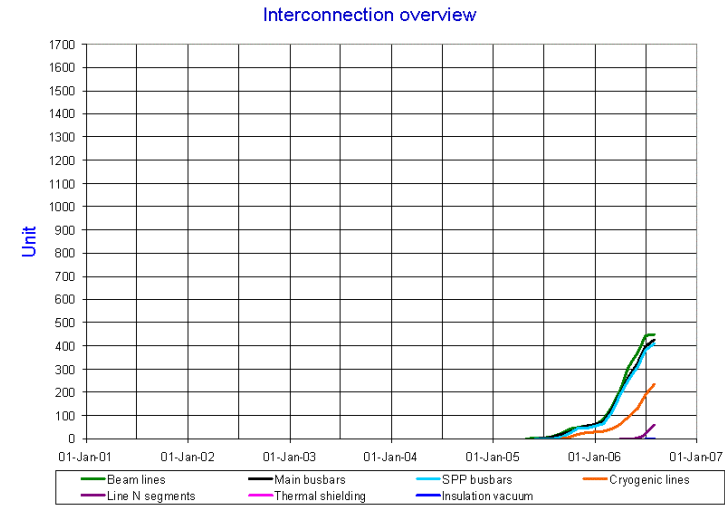
Welcome to our 2 new Tier 2 Sites!

LHC Schedule



Updated 31 Jul 2006

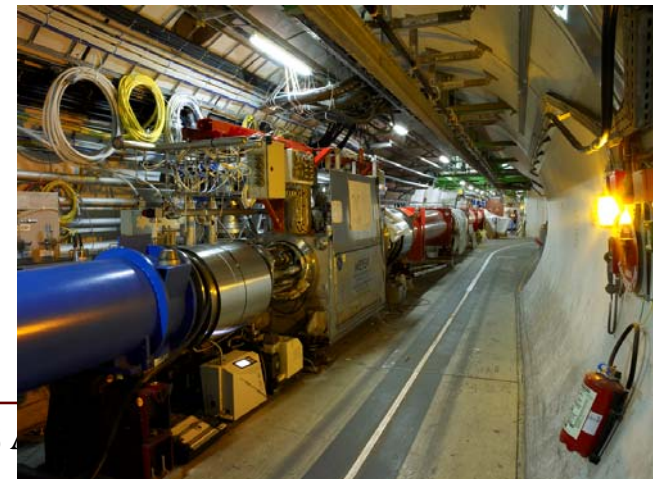
Data provided by D. Tommasini AT-MAS, L. Bottura AT-MTM



Updated 31 Jul 2006

Data provided by I Ph. Tock AT-CRI

LHC tunnel



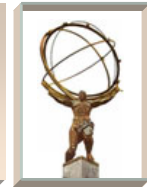
- Magnet installation going well – 1,250 total
 - Passed the halfway mark
- Interconnection is a challenge
 - Install does not mean interconnected
- Very aggressive schedule for 2007

LHC Schedule



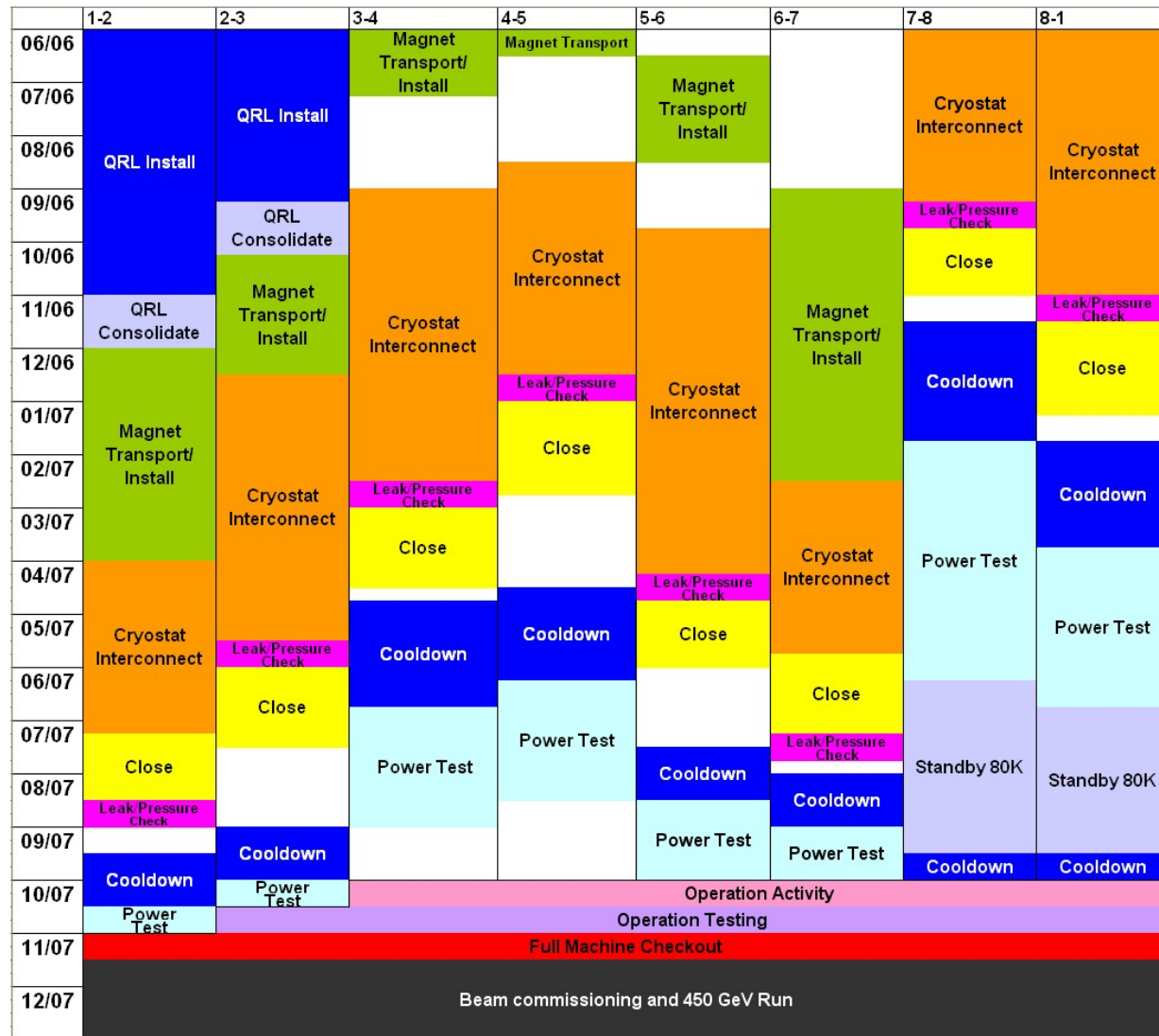
- **LHC schedule delay announced – 380 days to go!**
 - Now – Spring 2007 – LHC installation end game
 - last magnet: delivered (10/06), tested (12/06), installed (3/07)
 - August 2007 -- beam pipe closes
 - nominal two month delay from previous schedule
 - Sept 2007 -- few weeks of controlled access to cavern
 - Nov 2007 – Two month LHC commissioning run at injection energy (450 GeV on 450 GeV, no ramp, no squeeze) till end 2007. Sectors 7-8, 8-1 will be commissioned at 7 TeV, others not
 - Early 2008 -- few month shutdown during which remaining LHC sectors commissioned w/o beam at full 7 TeV energy
 - Mid 2008 -- 14 TeV running
 - goal is substantial (1-4 fb⁻¹??) by end 2008
- **In September, CERN should clarify the goals for 2008-9 since this will affect computing resources**

LHC Schedule



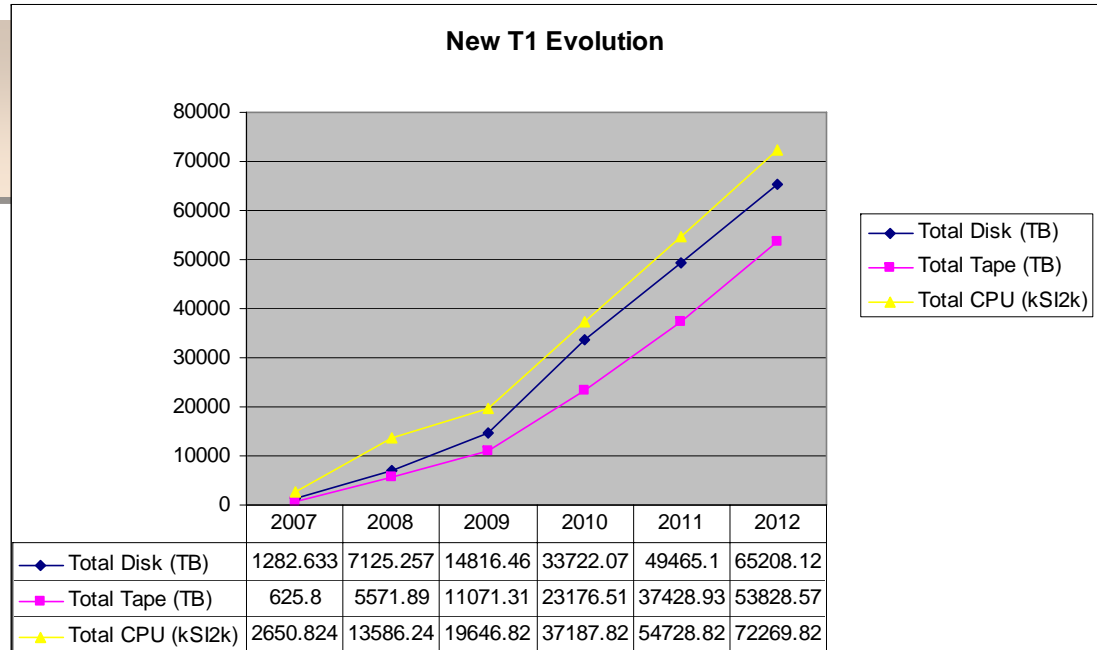
- My simplified graphical view based on 7/7/06 detailed update in

<http://sylvainw.home.cern.ch/sylvainw/planning-follow-up/Schedule.pdf>

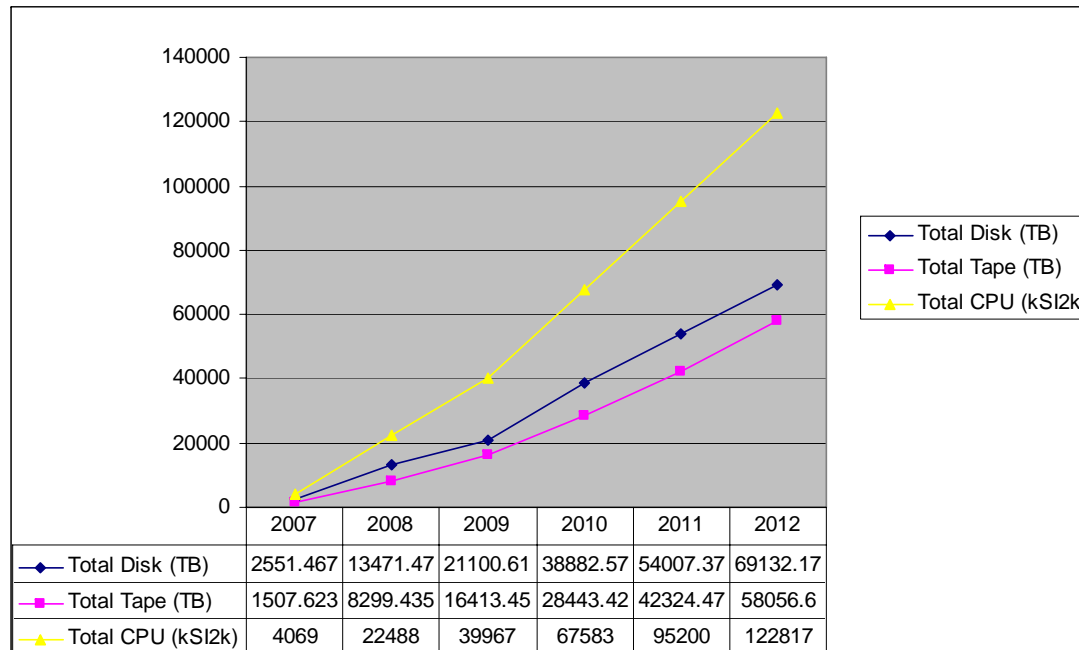


Tier 1 Evolution

Very Preliminary!



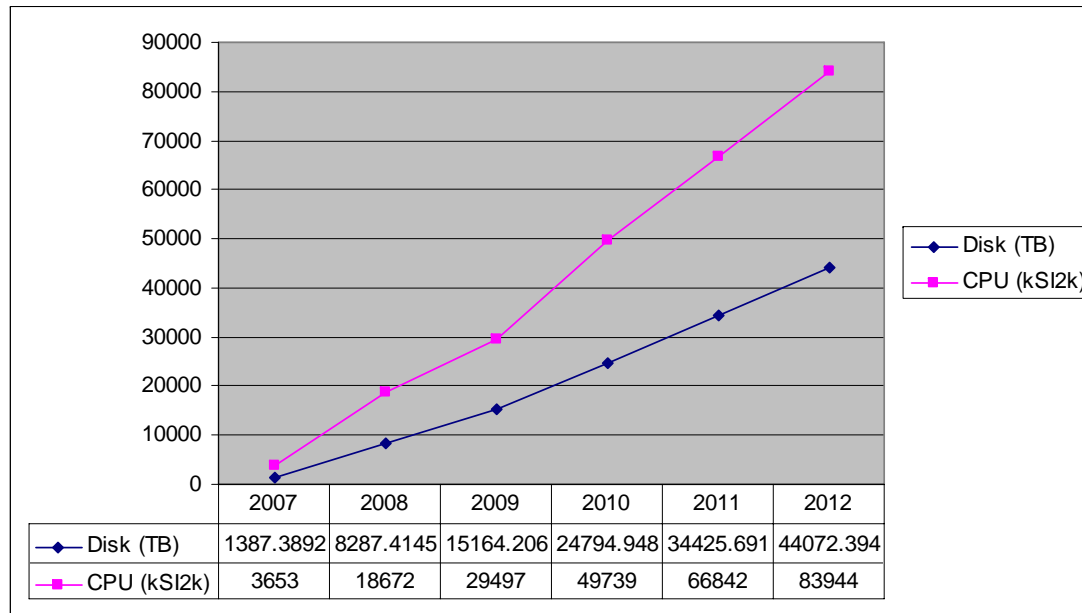
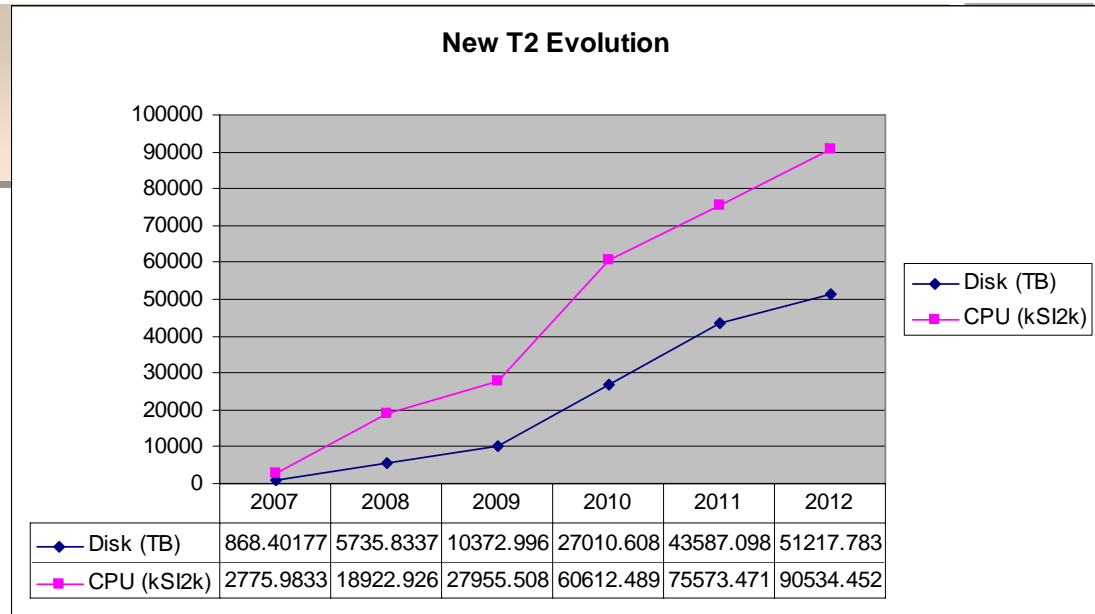
C-TDR values



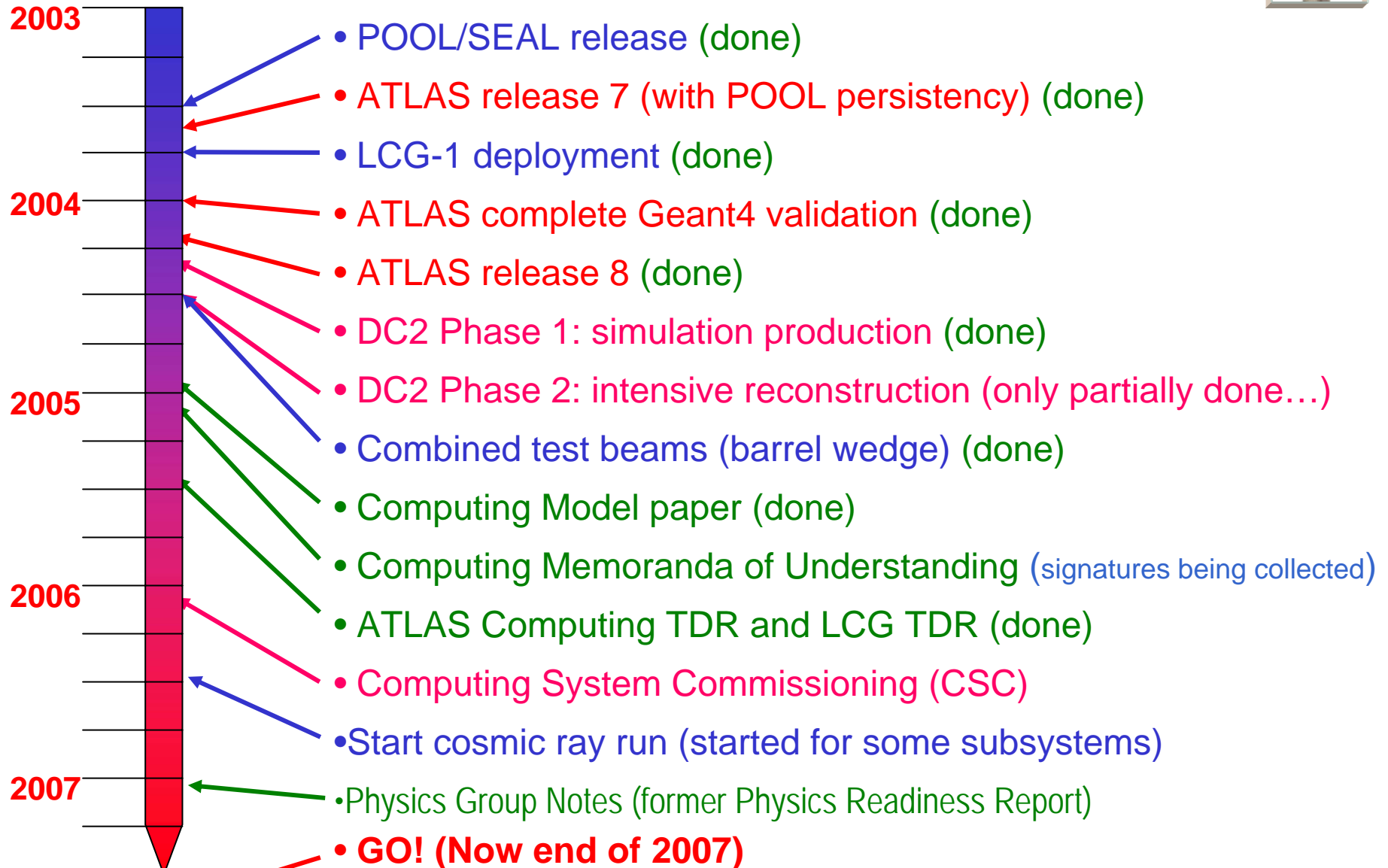
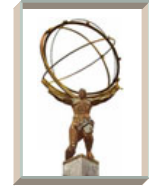
Tier 2 Evolution

Very Preliminary!

C-TDR values



ATLAS Computing Timeline

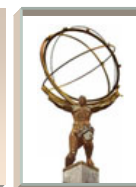


Resource Allocation Committee



- Newly formed committee to coordinate U.S. resources (CPU/disk)
 - Prioritize: Physics simulation, analysis, calibration
- Committee members:
 - Physics Adviser (I. Hinchliffe)
 - Chair of the ASG (S. Willocq)
 - Facilities Manager (B. Gibbard)
 - Tier 2 (R. Gardner)
 - The US Production manager (K. De)
 - The EPM (J. Shank)
 - The Software Manager (S. Rajagopalan)
 - Physics users (B. Mellado)
 - Calibration users (B. Zhou)
- First meeting 25 Jan, 2006
 - <http://www.usatlas.bnl.gov/twiki/bin/view/AtlasSoftware/ResourceAllocationCommittee>

Resource allocation policy enforcement...



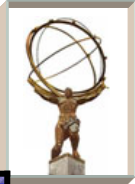
BNL USATLAS Tier 1 Condor Policy

The following table reflects the Condor policy for the BNL USATLAS Tier 1 Linux Farm compute nodes to take effect after July 5, 2006

	CAS 2 CPU_Speed == 2	CAS 3 CPU_Speed == 3	CAS 4 CPU_Speed == 4
Highest Priority	Short Jobs/Dial Jobs/Software Testing +RACF_Group = "short"	Production Jobs (available to OSG)	Distributed Analysis Jobs
	USATLAS Grid and Local Jobs (available to OSG) +RACF_Group = "usatlas"	USATLAS Grid and Local Jobs (available to OSG) +RACF_Group = "usatlas"	USATLAS Grid and Local Jobs (available to OSG) +RACF_Group = "usatlas"
	ATLAS Grid Jobs (Available to OSG and LCG)	ATLAS Grid Jobs (available to OSG and LCG)	ATLAS Grid Jobs (available to OSG and LCG)
Lowest Priority	Non-ATLAS Grid Jobs/General Queue (available to OSG)	Non-ATLAS Grid Jobs/General Queue (available to OSG)	Non-ATLAS Grid Jobs/General Queue (available to OSG)
Node Allocation	51 nodes, 102 CPUs (Dell 3.4 GHz) 46 nodes, 92 CPUs (PC 3.06 GHz)	130 nodes, 260 CPUs (Dell 3.4 GHz)	41 nodes, 82 CPUs (Dell 3.4 GHz)

- Total of 536 CPUs allowing up to 536 jobs running simultaneously.
- The Dell 3.4 Ghz machines have a SPECint2000 rating of 1345. The Penguin 3.06 Ghz machines have a SPECint2000 rating of 945.
- A higher priority job suspends a lower priority job until it finishes. Only two jobs may be running at a time, jobs are either suspended or running depending on their priority.
- Short jobs are limited to a runtime of 90 minutes.
- Each of the four job categories can run two jobs, one for each CPU.
- A Fairshare policy is in effect where a user may have his/her hold on a resource preempted if he/she has a low user priority.

Projected T2 Hardware Growth (dedicated to ATLAS)



Tier 2 Center	2005	2006	2007	2008	2009
Boston/Harvard					
CPU (kSi2k)	210	350	730	1,090	1,600
Disk (TB)	40	170	370	480	630
Southwest					
CPU (kSi2k)	500	900	1,500	1,700	2,100
Disk (TB)	60	200	380	540	700
Midwest					
CPU (kSi2k)	360	510	900	1,100	1,300
Disk (TB)	50	130	260	465	790

- Assumes Moore's law doubling of CPU (3 yrs) and disk capacity (1.5 yrs) at constant cost
- Assumes replacement of hardware every 3 years

Tier 3 Centers



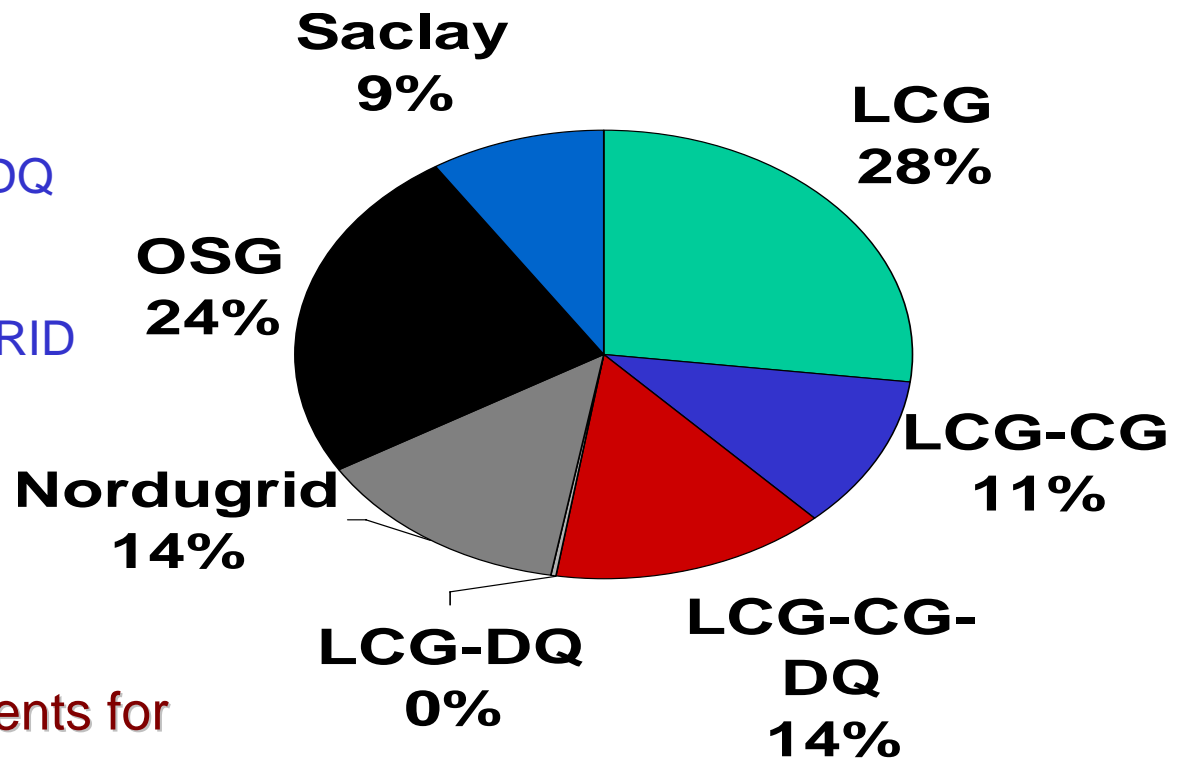
- **U.S. ATLAS formed a T3 task force in April 2006**
 - Composition: Gustaaf Brooijmans, Rob Gardner, Bruce Gibbard, Tom LeCompte, Shawn McKee, Razvan Propescu, Jim Shank.
- **Produced whitepaper on role of T3 in U.S. ATLAS**
- **Summary from whitepaper:**
 - Some local compute resources, beyond Tier-1 and Tier-2, are required to do physics analysis in ATLAS.
 - These resources are termed Tier-3 and could be as small as a modern desktop computer on each physicist's desk, or as large as a Linux farm, perhaps operated as part of a shared facility from an institution's own resources.
 - Resources outside of the U.S. ATLAS Research Program are sometimes available for Tier-3 centers. A small amount of HEP Core Program money can sometimes leverage a large amount of other funding for Tier-3 centers. Decisions on when it is useful to spend Core money in this way will have to be considered on a case by case basis.
 - Support for Tier-3 centers can be accommodated in the U.S. Research Program provided the Tier-3 centers are part of the Open Science Grid and that they provide access those resources with appropriate priority settings to US ATLAS via the VO authentication, authorization and accounting infrastructure.

CSC11 Production Summary



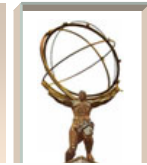
•Finished jobs:

- 155199 | LCG
- 64710 | LCG-CG
- 82202 | LCG-CG-DQ
- 1903 | LCG-DQ
- 79010 | NORDUGRID
- 140728 | OSG
- 52158 | SACLAY

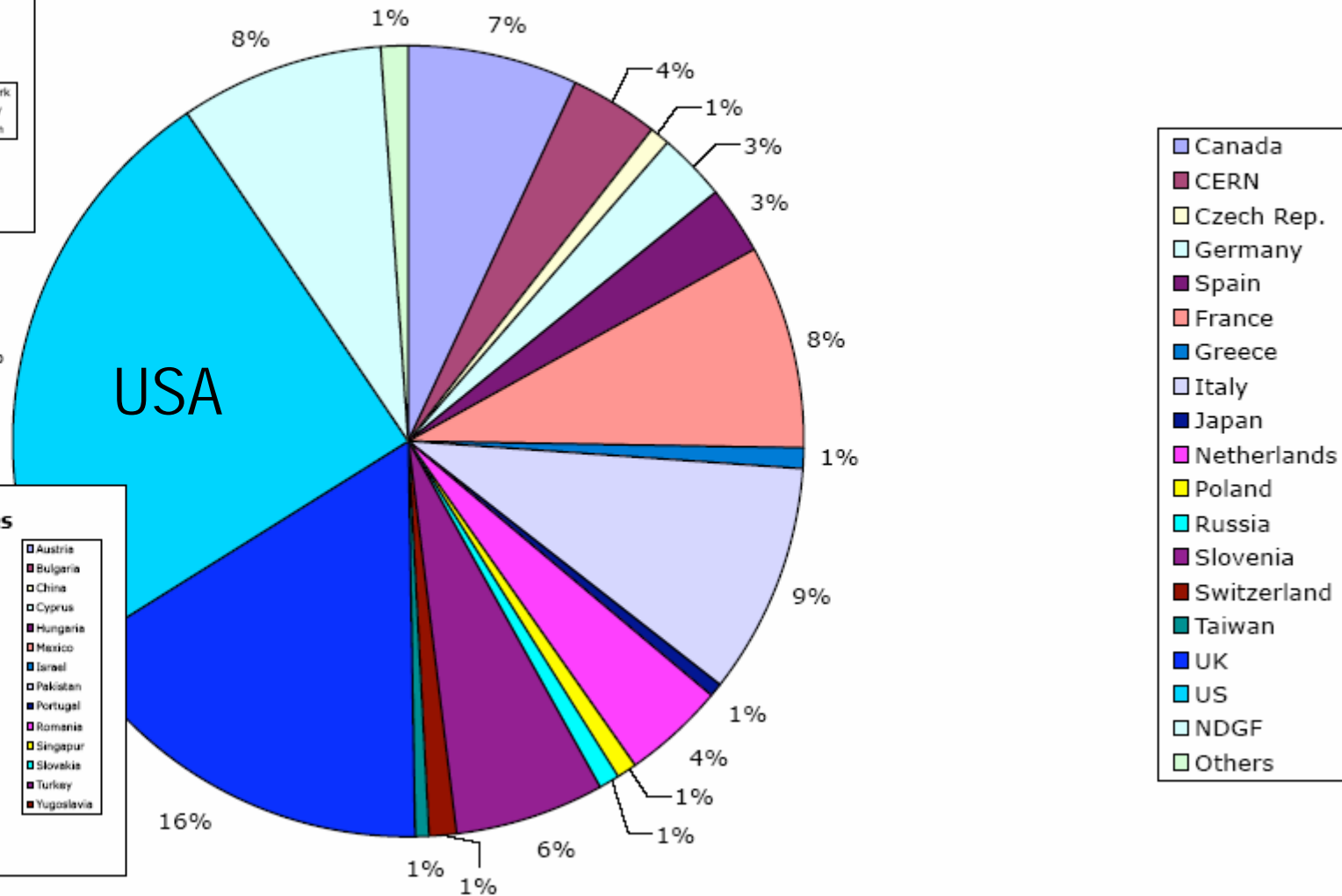
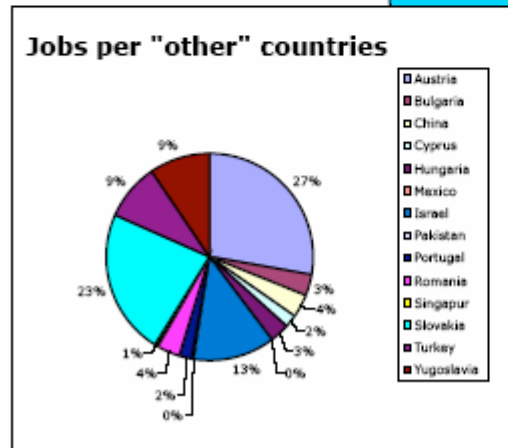
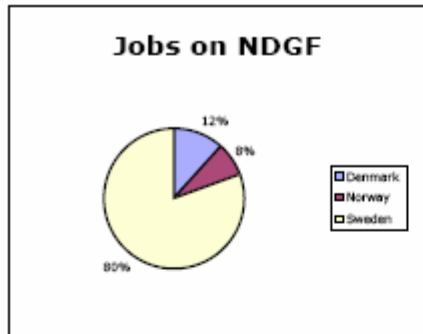


•Note each job has 50 events for physics samples, 100 events for single particles

Current CSC Production



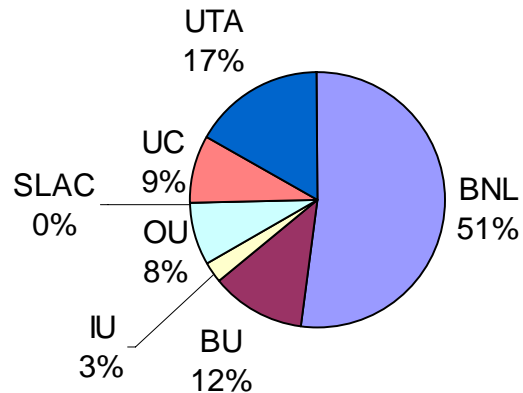
ATLAS production - Number of jobs per country



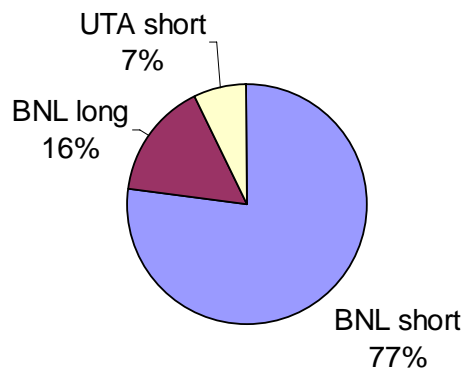
U.S. Production 1st Half 2006



Panda Production Jobs (6 months)



Panda Analysis Jobs (6 months)



- U.S. provided 24% of managed CSC production through PanDA

- Half of U.S. production was done at Tier 1, remainder at Tier 2 sites

- PanDA is also used for user analysis (through pathena)

- Currently, PanDA has 54 analysis users who have submitted jobs through the grid in past 6 months

- Analysis jobs primarily run at BNL. We are testing also at UTA. Soon all T2 sites will be enabled.

Funding Targets

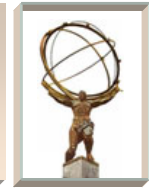


U.S. ATLAS Research Program Target Chart (AYk\$s) as of June 2004 Strawman Guidance

Category	WBS	Description	FY06	FY07	FY08	FY09
Computing	2.2	Software	4,533	5,403	5,648	5,648
		Direct to Experiment	4,141	4,917	4,919	4,774
		OSG/LCG/etc	310.0	241.8	251.5	174.4
		Cost of Analysis Support Center	82	244	478	700
	2.3	Facilities	5,365	9,394	9,952	9,952
		Tier 1 Labor/M&S	1,823	3,276	3,834	3,834
		Tier 1 Equipment	1,412	2,200	2,200	2,200
		OSG/LCG	357	419	419	419
		Tier 2 Labor/M&S	637	1,500	1,500	1,500
		Tier 2 Equipment	637	1,500	1,500	1,500
		Production	500	500	500	500
	2.0 Total Computing	9,898	14,797	15,600	15,600	

Bottom line not changed since last review

Reconciling Requests with Target



- As shown in Feb, we still have about a \$3M difference in requests over our targets in 2008 and beyond.
 - As shown then, we put about \$1M for software and \$2M for Facilities as requests to the Management Reserve (MR).
- We are still working on what our actual 2006 spending will be
 - Late hires at T1, late purchases at T2
 - Should allow us to get where we want on 2007 with minimal call on MR (< \$500k ?)
- 2008 and beyond still a problem
 - We are working closely with the ATLAS Computing Model Group to understand our T1/T2 needs out to 2012
 - New LHC running assumptions COULD lead to some savings in a later ramp of hardware
 - Software: Emphasis on user support, if not enough MR, we will have to cut some Core effort

Need to Ramp up the T2 hardware



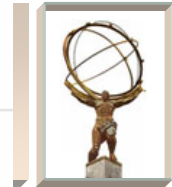
- The Research Program has a large FY06 Rollover
- We are begging for more money for FY07, but the agencies are saying use our rollover.
- Use it or loose it

The T2 Meeting Last May



- Tier 2 Planning Wiki:
 - <http://www.usatlas.bnl.gov/twiki/bin/view/Admins/TierTwoPlanning>

Tier2 Planning



- ↓ [Tasks and Milestones](#)
- ↓ [Questions / Issues for facilities](#)
 - ↓ [NET2](#)
 - ↓ [SWT2](#)
 - ↓ [MWT2](#)
- ↓ [OSG Software and Panda validation](#)
- ↓ [References](#)

Tasks and Milestones

- [TierTwoFacilities](#) and [TierTwoNetworking](#)
- [TierTwoStorageDataServices](#)
- [TierTwoPolicyAccounting](#)
- [TierTwoOperationsUserSupport](#)

Questions / Issues for facilities

NET2

- **Current set of action items:** [NortheastTier2ActionItems](#)
- Notes taken during the May 2006 Tier2 workshop:
 - Main set of questions concern storage system to use.

SWT2

- **Current set of action items:** [SouthwestTier2ActionItems](#)
- Notes during May 2006 Tier2 workshop:
 - Deployed 160 nodes
 - Number of headnodes - 8 TB
 - 16 TB SAN storage: IBRIX filesystem. To study
 - Additional xx TB available in distributed storage (single SATA drive)
 - Q: use dcache, or resilient dcache
 - Platform Rocks 3.3.0 (RHEL X86_64) - additional Dell rolls
 - Platform freeware scheduler Lava (looking into)
 - Networking: oc12 from ntg to houston peering site. future: LEARN and NLR? Need to look into peering.
 - OU: same as UTA, 1/4 scale. Plan on putting SRM/dCache on top of IBRIX filesystem.
 - OU: network connectivity option is through NLR.

Southwest Tier2 Action Items

Applies to ALL T2!

- ↓ [Introduction](#)
- ↓ [Facilities](#)
- ↓ [Networking](#)
- ↓ [Storage and Data Services](#)
- ↓ [Policy and Accounting](#)
- ↓ [Operations and User Support](#)

Introduction

These are some notes to follow-up on [TierTwoPlanning](#) tasks and deliverables as discussed at the May [Tier2 workshop](#) in Chicago. This is being updated for the [Tier2 workshop](#) at Harvard, August 17-18, 2006.

Facilities

Follow-up from the [TierTwoFacilities](#) working group:

- ...

Networking

Follow-up from the [TierTwoNetworking](#) working group:

- ...

Storage and Data Services

Follow-up from the [TierTwoStorageDataServices](#) working group:

- ...

Policy and Accounting

Follow-up from the [TierTwoPolicyAccounting](#) working group:

- ...

Operations and User Support

Follow-up from the [TierTwoOperationsUserSupport](#) working group:

- ...

Goal:
Fill-in/Update these Wikis



Progress on T2 Services?



Tasks and Deliverables

To the following deliverables has to be added what is specified within ATLAS Computing TDR or other documents. Total space requirement (in TB) have to be added and crossreferenced with the content of the [TierTwoFacilities](#) document.

Each T2 has to:

- be able to connect to Tier1 through FTS channel (no work necessary, the channel is open by the Tier1)
- provide a SRM managed storage: backend storage can be dCache (SRM/dCache) or other efficient storage (high cost solutions) A suggested SRM/dCache server would require at least 2, better 3 or more machines.
 - a CORE machine providing a file server serving at least 1 TB using PNFS. Possibly this machine (the DB and PNFS manager) should be separate on a powerful machine (leave all the other daemons of the core and dors on a the other machine): BNL has 2 3.0 Xeon, 4 GB
 - one or more machines with DOORS (gsiftp, dcap, SRM). You may like a separate machine for the SRM door.
 - provide some reader/writer nodes (at least one reader/writer) An alternative would be a SRM/DRM solution on top of a high performance shared FS, like Ibrx, Panases, Luster, GPFS
- provide a DQ2 server

Here is a list of tasks/deliverables. Most will have also related milestones in the next section:

1. Provide a dCache server suitable for SC4 (v1.6.6.5) and production:
 - At least 2 machines (see above) with 1 TB (min to have a realistic use)
 - dCache installation (Easy to bring up! - BNL will offer support)
 - Tune-up and improvement can be done later
 - update of dCache when needed
2. Provide a DQ2 server
 - have a DQ2 server functional (v.?)
 - update DQ2 server when needed
3. Prototype and recipe of T2 DB infrastructure
 - Test DB service for Condition and Tag
 - start with current model
 - investigate different models
 - provide requirements
 - provide a deployment recipe (Pacman?) and documentation
4. Readiness for the Calibration Alignment challenge (CAC will be this fall, after release 13, at the end of September)
 - Have a DB service (one machine with DB server hosting Tag and Condition DB - will be finalized by prototyping):
 - Tag: probably dedicated mysql server
 - Condition: probably file based DB like SQLite, but may be mysql
 - access will be only from within T2 site (all clusters part of it)

Tier 2 Documentation



- Uniform Web pages
- Up to date snapshots of hardware configuration
- Kill/Erased/destroy old web pages!