Facilities Operations and FDR
Introduction, Schedule and required Capabilities

Michael Ernst

Brookhaven National Laboratory

U.S. ATLAS Transparent Distributed Facility Workshop

Chapel Hill, NC

3 – 6 March 2008

HEP office of high energy physics

BROOKHAVEN NATIONAL LABORATORY

The Computing Model

• Resources Spread Around the GRID

•Interactive Analysis
•Plots, Fits, Toy MC, Studies, ...

Tier 3

•Reprocessing of full data with improved calibrations 2 months after data taking.
•Managed Tape Access: RAW, ESD
•Disk Access: AOD, fraction of ESD

DPD

30 Sites Worldwide
Tier 2

•Derive 1st pass calibrations within 24 hours.
•Reconstruct rest of the data keeping up with data taking.

AOD

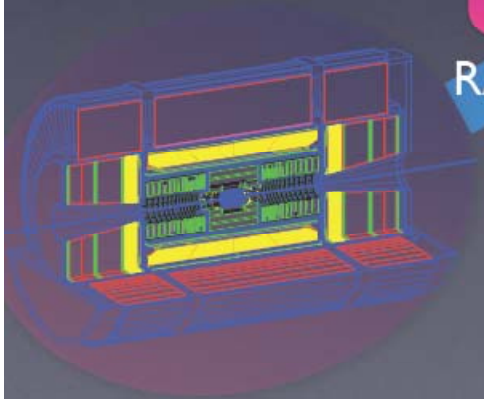
Tier 1

10 Sites Worldwide

RAW/ AOD/ ESD

Tier 0

RAW



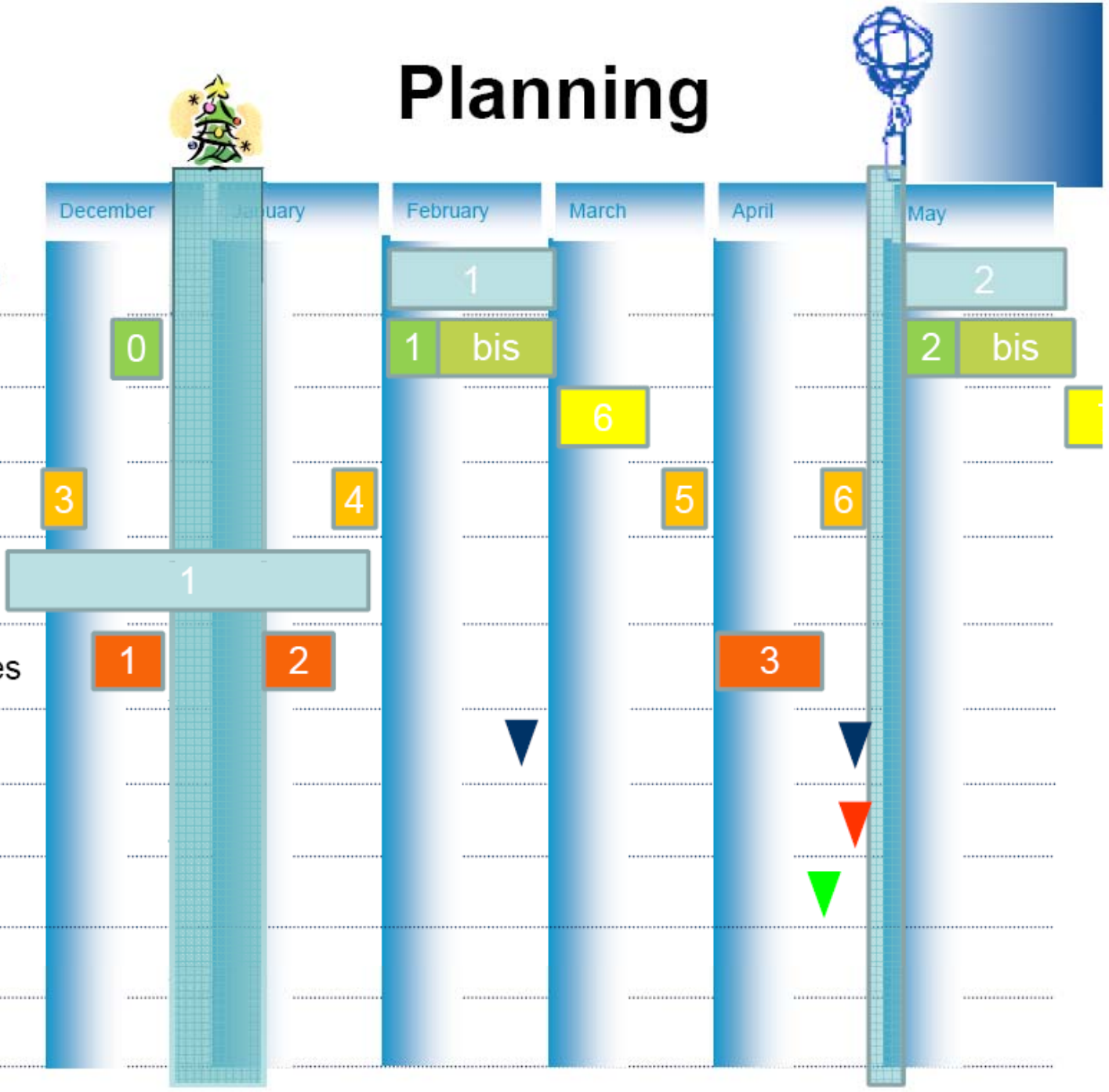
CERN Analysis Facility

•Production of simulated events.
•User Analysis: 12 CPU/ Analyzer
•Disk Store: AOD

•Primary purpose: calibrations
•Small subset of collaboration will have access to full ESD.
•Limited Access to RAW Data.

A. Farbin

Planning


December
February
March
April
May
CC Readiness Challenge
Full DressRehersal
Cosmic Ray tests
Functional Tests
SRMv2 testing
Time for fixes and updates
S&C Week
Jamboree 2 (?)
WLCG Workshop
1
2
0
1
bis
2
bis
6
3
4
5
6
1
1
2
3

FDR-1

- Brief introduction to what's happened in FDR-1
  - FDR-1 run took place ~ as scheduled in week of 4 February
  - 8 1h runs for two days (Tu/We) physics events only
  - 10 1h runs for third day (Th) physics events plus e/γ fakes in 2 runs
  - all data at a luminosity of 10E31
  - Data quality / alignment processing happened
  - Fired off bulk reconstruction at the Tier-0
  - Different reconstruction version each day
  - Need reprocessing soon at the Tier-1s
  - No (useful) TAGs from this processing
  - Data distribution (RAW (~1TB), ESD, AOD) took place on Mo-Th of following week, to Tier-1s

FDR / AOD Replication to ATLAS Tier's

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

FDR-1 - Some Results

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

FDR-2 Timescale

- We're now in week 9, and May is weeks 19-22 - a guess for FDR-2 is week 21
- Has to be negotiated with run coordination (SFO & Tier-0 use)
- Then we have a maximum of 8 weeks (56 days) from now for sample preparation
- In terms of weeks:
  - FDR - 12 13.0.40.2 validated, 13.0.40.3 ready?
  - FDR - 12-5 simulation & digitisation (some needs 40.3)
  - FDR - 5-2 event mixing at BNL
  - FDR - 2 deadline for any code to be run in Tier-0 deadline for release
  - FDR - 1 frozen software ship events from BNL to CERN
  - FDR week: upload to SFOs, run
  - Note that the weeks overlap
    - Schedule is already very tight: less than 10h data? Take the priority list seriously

Resources in ATLAS - Any Changes?

- We need to increase most dataset volumes by 10% to account for inclusive streaming
- The DPD looks like being about the same size as the AOD
  - ATLAS wants the full DPD at each T1 and to follow the AOD share
- All Numbers on the following slides are preliminary

Processing Times (per Event)

- We learned last week that FDR simulation with new showering takes 2300kSI2k-sec
  - The assumption (inflated from 100) was 400
- Times now are
  - Simulation: 1950 kSI2k-sec
  - Processing/reprocessing: 30kSI2k-sec (should be 15kSI2k-sec!)
- A lot of difficulties expected
  - Need to carefully determine the event mix to save on CPU

Event Sizes

- Event sizes now taken from FDR-1
  - Some good news but mainly very bad
  - RAW 1.6MB, ESD 750kB, AOD 170kB, TAG1kB
  - Simulated RAW 4MB, ESD 790kB, AOD 210kB
  - Assume DPD global size = AOD global size
- This is also bad (although reducing the simulation volume helps a little)
  - This will imply either less DPD or fewer AOD instances
  - The choice may depend on the year

Attempting to fit all this in…..

- The resource numbers are not likely to change upwards anytime soon
- We do not have a new schedule of LHC running
  - LHC Running Scenarios
    - Assume 4M sec in 08, 6M sec in 09, 10M sec thereafter
    - Assume no HI in 2008, assume lumi does not go to 1E34 until 2011 (affects event size, time per event)

Other proposed cuts

- Reduce the fractions of data on disk:
  - In T2 cloud
    - 30% of RAW and 75% of ESD in 2008
    - 9% of RAW and 15% of ESD in 2009 & 2010
  - In T1 cloud:
    - 10% or RAW on disk, still have 2 ESD copies

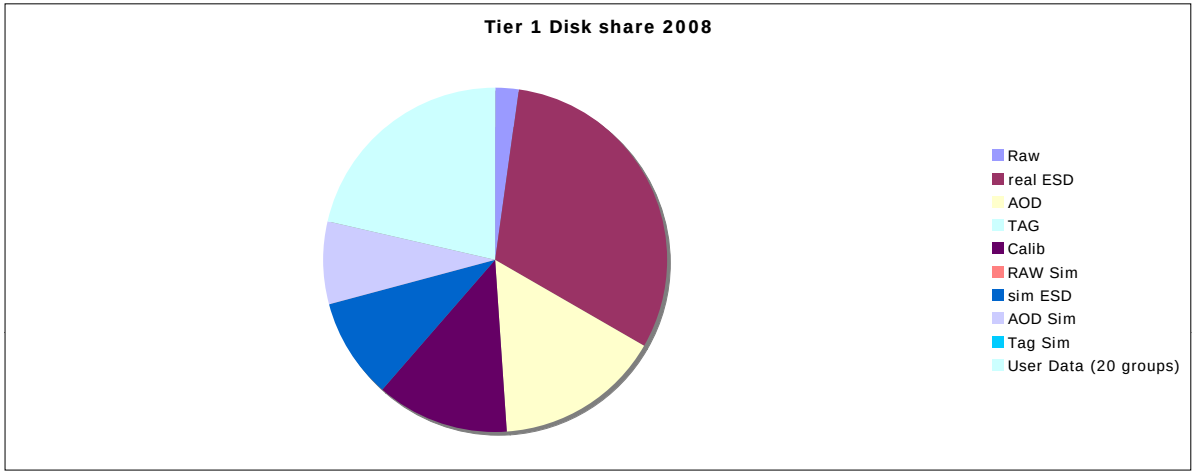
Effect of increased simulation time

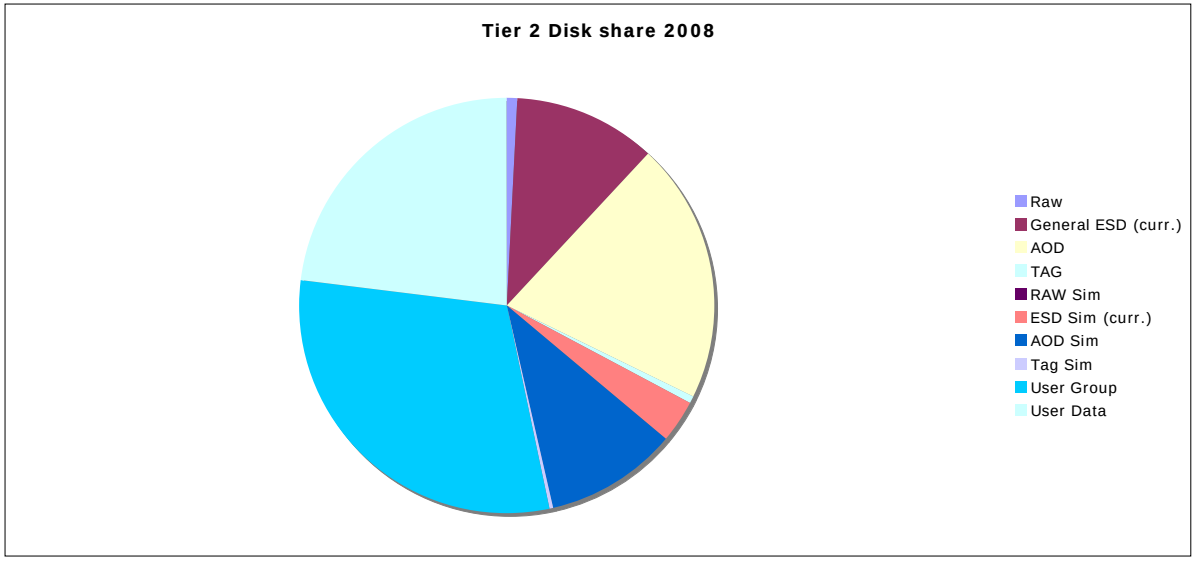
- In T2 cloud
  - Significantly cut back (~half) on per user disk and CPU at Tier 2s
- In T1 cloud:
  - 10% of RAW on disk, still have 2 ESD copies
- Reduce simulated data rate
  - 08: reduced by 40% to 11% of total
  - 09: reduced by 40% to 12% of total
  - 10% of total thereafter
  - This is really radical, but still leaves problems and is a direct consequence of increased simulation time

Implied use of Tier 2 Analysis Facilities

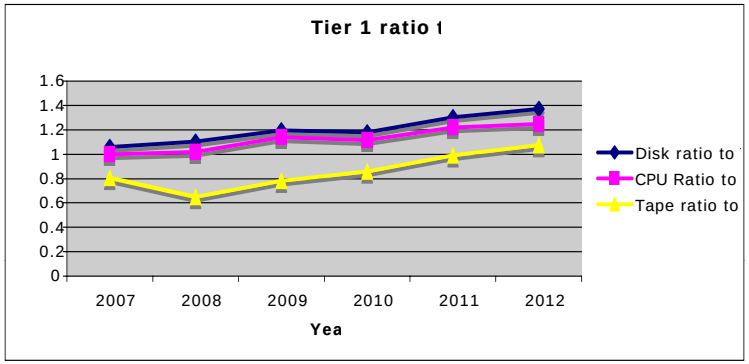
	2008		2009		2010		2011	
	CPU	Disk	CPU	Disk	CPU	Disk	CPU	Disk
User	28%	14%	28%	13%	12%	9%	11%	8%
Group	17%	36%	8%	42%	5%	44%	2%	37%
Simu	56%		64%		83%		87%	

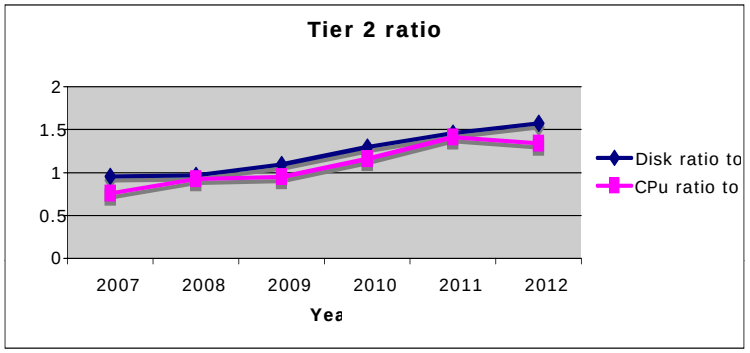
Disk Share 2008 (draft)





Away from CERN





- We may have to reduce reprocessing at Tier-1s if CPU limited
  - Only one reconstruction pass per Year?
- Disk will remain an issue
- The Tier-2 storage may mean losing 2-3 copies of the AOD/DPD
- Not clear even 10% simulation is sustainable by 2011

Simulation Budget?

- From the model, Roger Jones proposed the following simulation budget
  - 13.6MSI2k in 2008 for simulation (5.4MSI2k in T1s, 8.2 MSI2k in T2s)
  - This is 41% of the T1 & T2 combined CPU
  - Total budget for 1 instance of all April08-March09 simulation 850TB; all instances in Tier 1/Tier 2 system within 930TB
- The U.S. share is ~20%
- If Sites cannot use all installed disk or tape, resources should be released for other things!

Computing in U.S. ATLAS

- Computing resources used for production and analysis
  - BNL Tier-1
  - Five Tier-2's
  - A lot Tier-3's
  - Tier-1 and Tier-2's are funded through the Research Program, Tier-3's need other funding sources and/or leverage existing resources at Universities
  - All sites are organized hierarchically
- Personnel involved in production
  - Tier-1 site support (clusters, storage, networking)
  - Tier-2 site support (also helping Tier-3's)
  - Service support (servers, alarms, installation, integration)
  - Shift team
  - Facility Integration and Operation Coordinator
  - Production & Analysis Operation Coordinator
- Software systems for production
  - Panda (including pAthena) and ATLAS Distributed Data Management (DDM/DQ2, managed data "channels" by FTS)

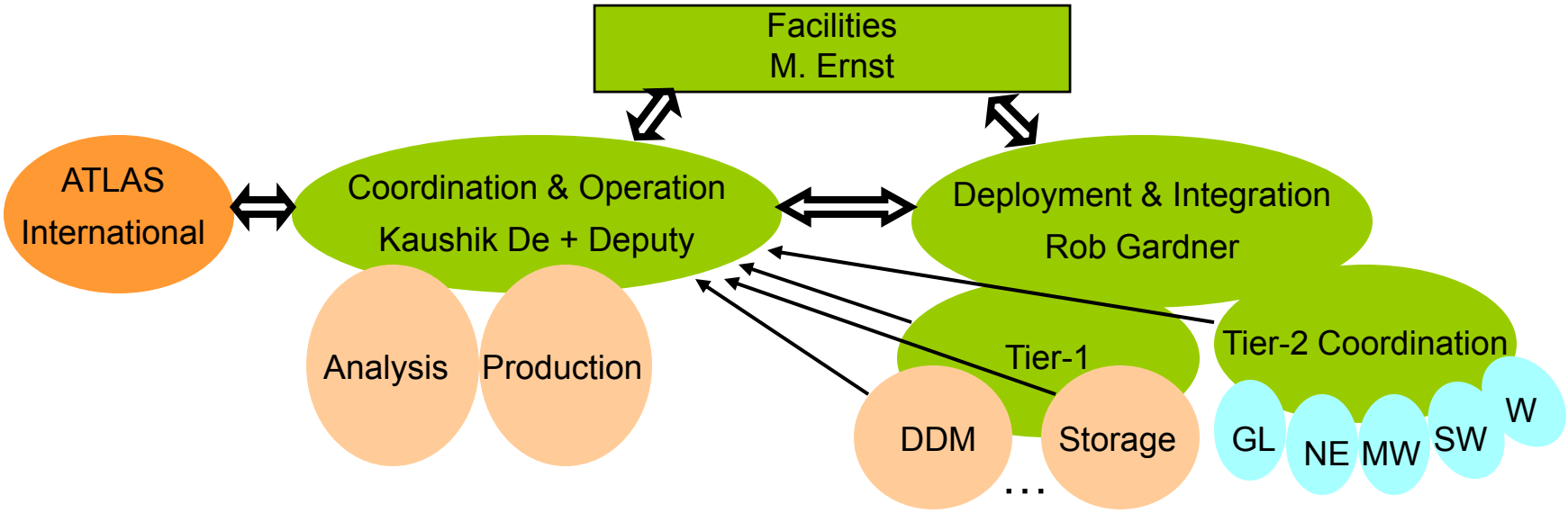
Data Location

- Tier 1 – main repository of data (MC & Primary)
  - Store complete set of ESD, AOD, Ntuple & TAG data on disk
  - Fraction of RAW and all U.S. generated RDO data
- Tier 2 – repository of analysis data
  - Store complete set of AOD, Ntuple & TAG's on disk
- Data distribution to Tier 1 & Tier 2's is managed
- Tier 3 – unmanaged data matching local interest
  - Data through locally initiated subscriptions
  - Mostly Ntuple's, some AOD's
  - Tier-3's most likely be associated with Tier-2 sites
  - Tier 3 model is still not fully developed – evolving

Facility Organization

Facility divided into two principal lines

- Production and Analysis Coordination & Operation
- Computing Deployment, Integration and Operation

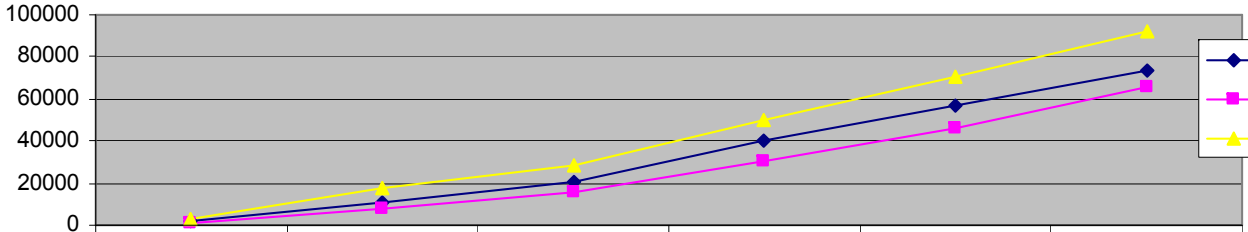


Production Operation Coordinator

- Responsible for Production, Analysis and DDM operation in the U.S
  - Overall responsibility for tracking and managing resources available to U.S. ATLAS
  - Define shift requirements and oversee shift activities
  - Establish resource allocation mechanisms and controls to implement RAC policies effectively
  - Define and maintain deliverables and milestones
  - U.S. ATLAS representative to ATLAS production and DDM operations
  - Representative to U.S. Physics Management
  - ATLAS representative to OSG for ATLAS Operations
  - Define and manage schedule for production software deployment and middleware (in cooperation w/ Integration Program)
    - U.S. ATLAS representative to OSG/VDT for release planning and validation

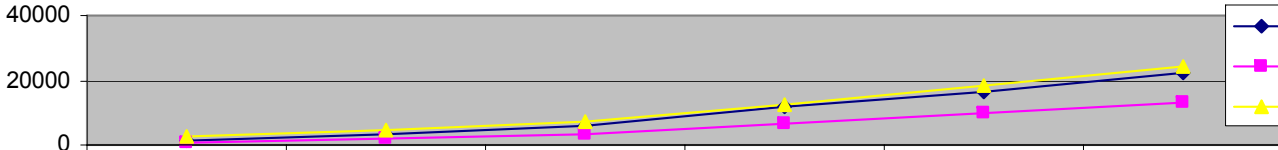
Resource Projection

Total ATLAS Tier-1 Resource Projection (revised October 2007)



	2007	2008	2009	2010	2011	2012
Total Disk (TB)	2090	10725	20922	40350	57053	73756
Total Tape (TB)	1246	8067	15787	29903	46503	65586
Total CPU (kSI2k)	3173	18124	28426	49576	70726	91877

Revised U.S. ATLAS Tier-1 Resource Projection (pledges, US is 23% of ATLAS, taking into account efficiency factors)



	2007	2008	2009	2010	2011	2012
Total Disk (TB)	1100	3136	5822	11637	16509	22328
Total Tape (TB)	603	1715	3277	6286	9820	13255
Total CPU (kSI2k)	2560	4844	7337	12765	18193	24008

Tier 1 Facility Capacity

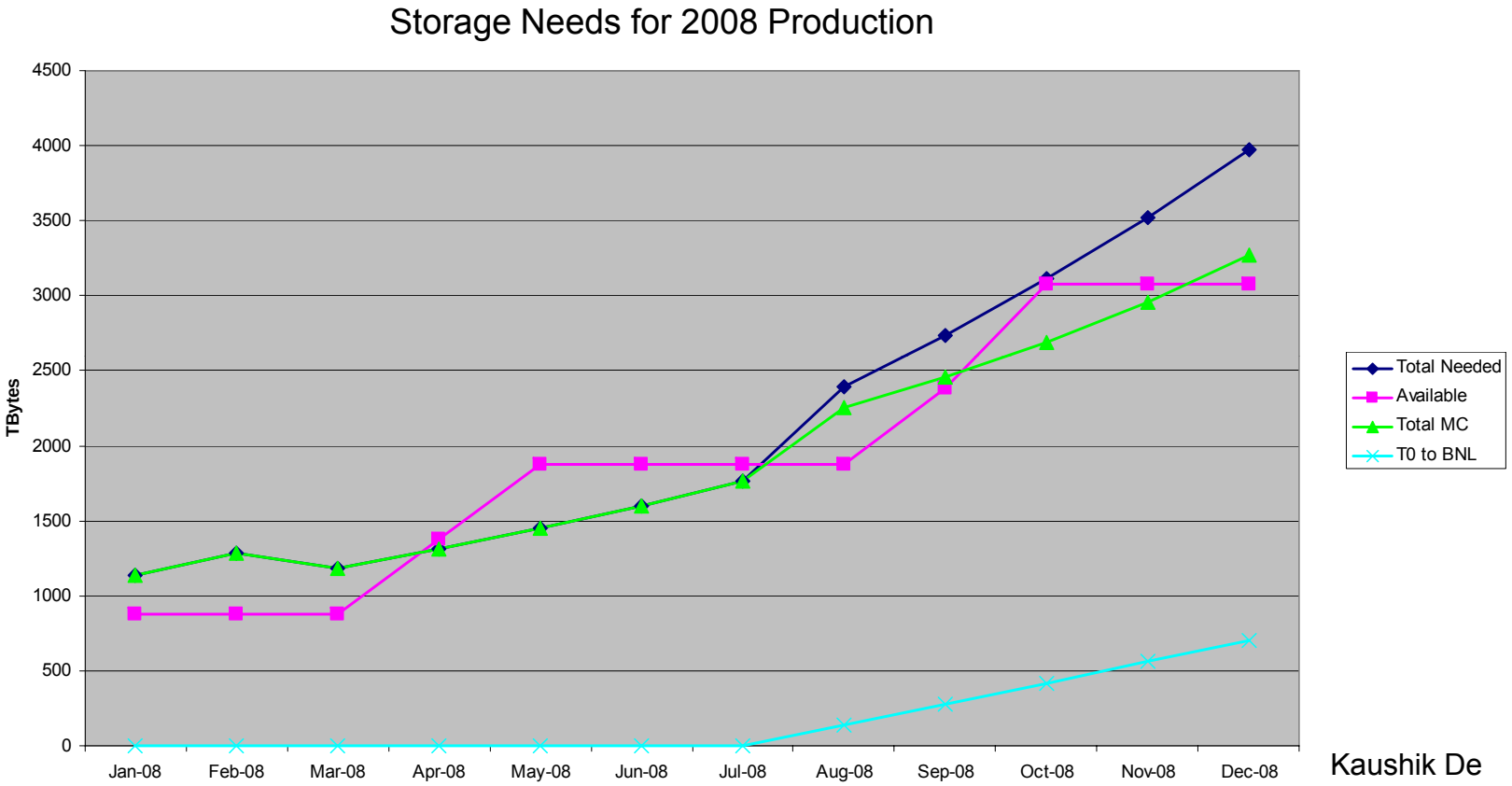
- Tier-1 Facility Capacity Profile (as of last years)
  - Includes Pledged + 50% locally controlled resources
  - The US must have such additional resources
    - Scale of additional US resources set to maintain full ESD copy and allow acceleration of analysis of one 20% data stream by a factor of two
    - In order to play a leadership role and to maintain a reasonable autonomy in Analysis

YEAR	2007	2008	2009	2010	2011
CPU (kSI2k)	2,834	7,140	11,598	18,838	26,875
Disk (TB)	1,556	4,610	8,921	17,262	24,427
Tape (TB)	993	3,284	6,276	11,996	18,781
WAN	2 x λ	2 x λ	3 x λ	4 x λ	4 x λ

wLCG Plan to pledge US model					
CPU (kSI2k)	2,560	4,844	7,337	12,765	18,193
Disk (TB)	1,100	3,136	5,822	11,637	16,509
Tape (TB)	603	1,715	3,277	6,286	9,820

Note: Capacity shortfall in FY'07 (0.4 MSI2k, 300 TB Disk)
Reason: Server Infrastructure replacement/improvements and Mass Storage System improvements (HPSS Cache and Tape Drives)

Storage Needs at the U.S. ATLAS Tier-1



Kaushik De

- Two Disk Procurements in FY'08: February and August
- Disk Resources proportional to CPU capacity in US Cloud
- Usage of Disk resources for MC Production contingent to Real Data Taking needs
  - After LHC-Startup CPU capacity at Tier-2s is needed for Analysis

Projections for U.S. Tier 2's

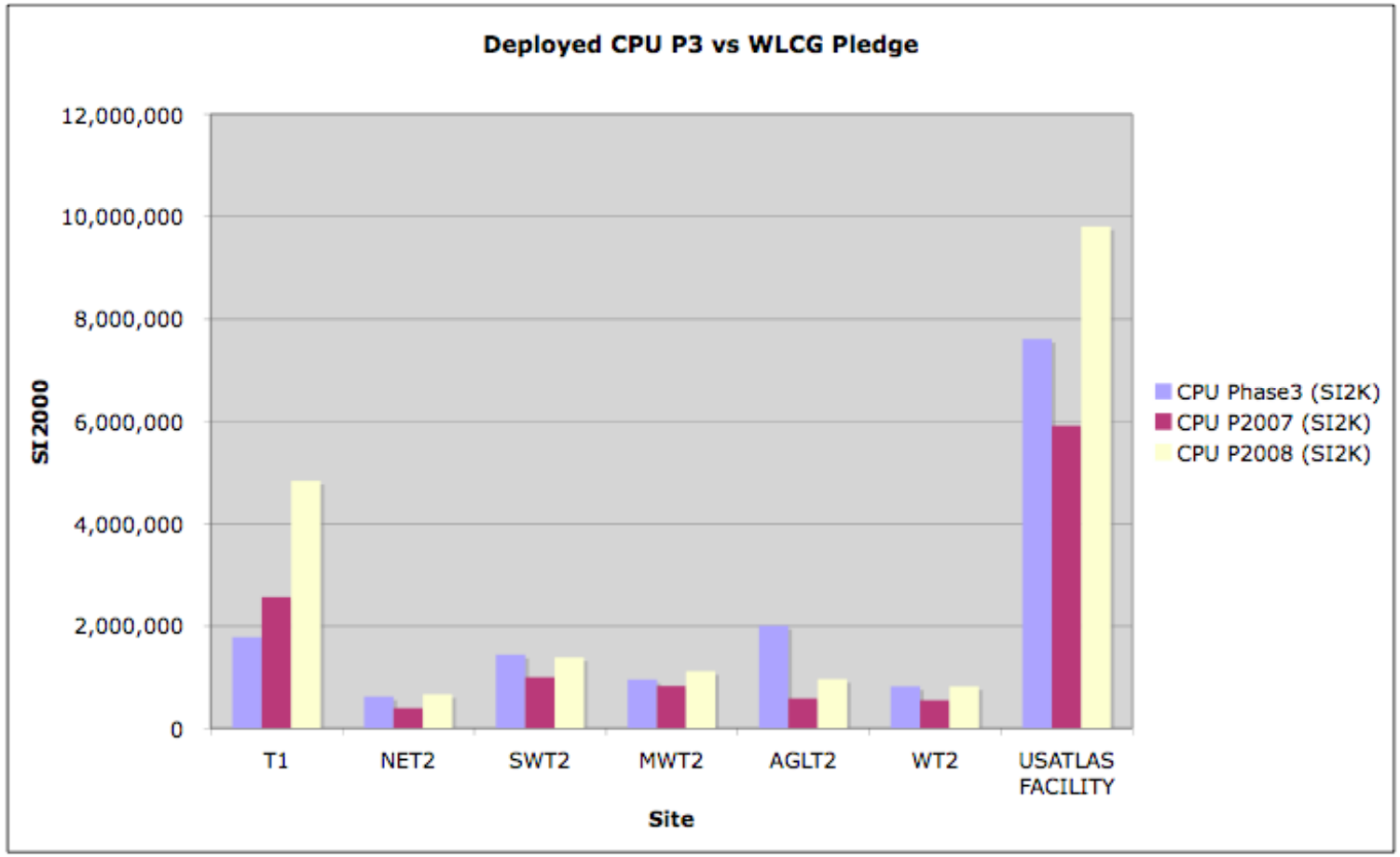
- Totals outline capacity committed to international ATLAS
- ~ 20% Capacity on top of totals retained under US control for US physicists

		2007	2008	2009	2010	2011
Northeast T2	*CPU (kSI2k)*	394	685	1,049	1,592	1,986
	Disk (TB)	103	244	445	727	1,024
Great Lakes T	*CPU (kSI2k)*	581	985	1,406	1,670	2,032
	Disk (TB)	155	322	542	709	914
Midwest T2	*CPU (kSI2k)*	826	1,112	978	1,282	1,785
	Disk (TB)	213	282	358	382	512
SLAC T2	*CPU (kSI2k)*	550	820	1,202	1,191	1,685
	Disk (TB)	228	462	794	1,034	1,462
Southwest T2	*CPU (kSI2k)*	998	1,386	1,734	1,986	2,514
	Disk (TB)	143	256	328	650	1,103
TOTAL US Tier 2's						
	CPU (kSI2k)	3,348	4,947	6,367	7,681	9,982
	Disk (TB)	842	1,567	2,467	3,482	5,015

Tier-2 Funding Profile (AY k$)					
Tier 2 Total	**3,000**	**3,000**	**3,000**	**3,000**

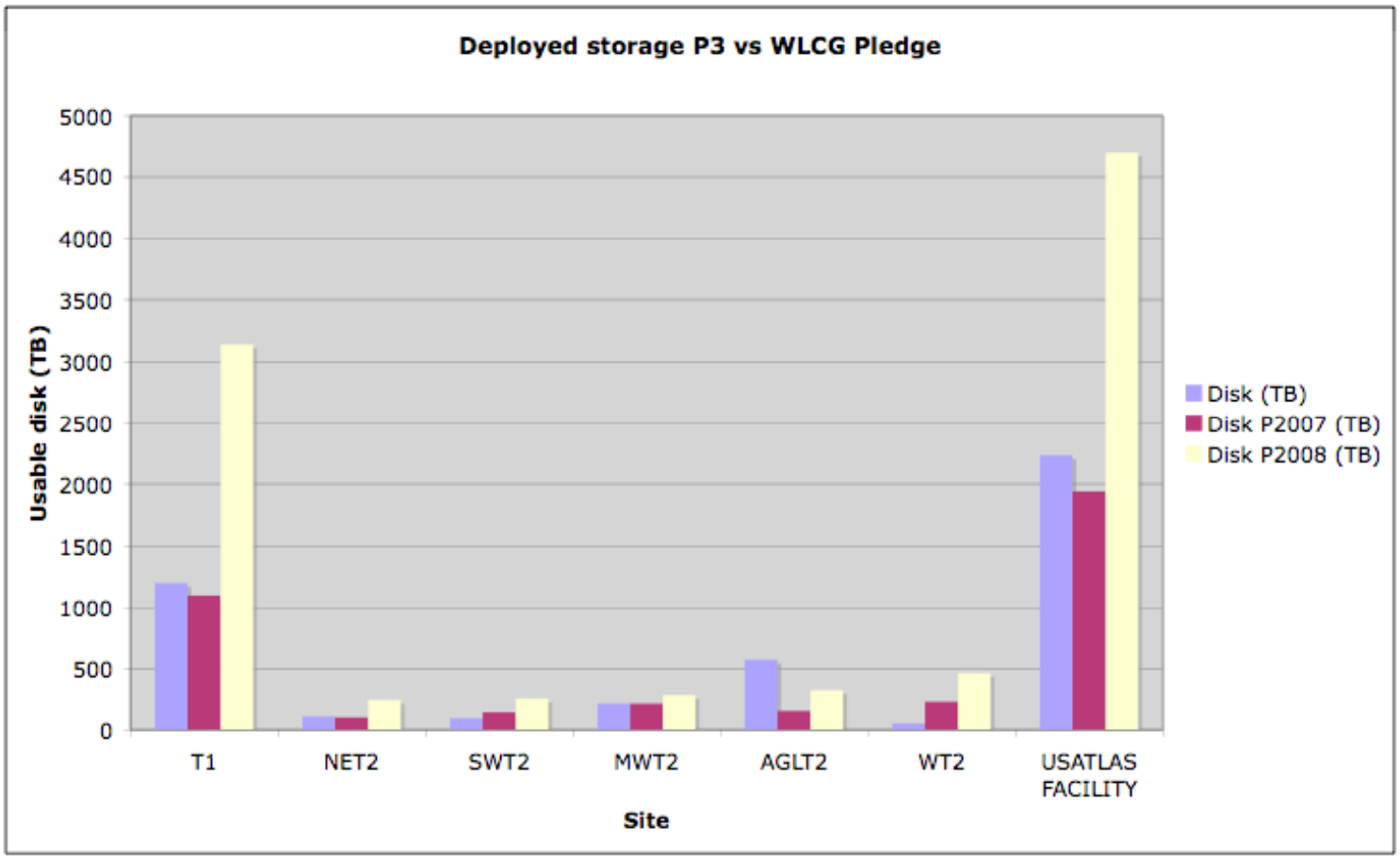
- CPU: On Target, Disk: Short by 20%
- Budget shortfall expected from 2010 on (replacement of obsolete equipment)

Deployed CPU versus WLCG Pledge



R. Gardner

Deployed Disk versus WLCG Pledge



R. Gardner

Facility Planning

The progressive growth of the Tier-1 and Tier-2 centers has been very helpful

- Provided ATLAS with critical resources for preparation
- Gained operations experience
- Identified and (partially) solved scaling limitations with grid and facility services

Milestone and Drivers for 2008

The 2006/2007 drivers were primarily designed to

- demonstrate the functionality of the computing model elements
- To start the transition of the experiment into stable operations
  - To be continued in 2008
- Participation in global ATLAS exercises
  - FDR, Cosmic Ray Runs, Functional Tests
- Goal is to have end-to-end data collection, distribution, and access
  - Includes DAQ and Tier-0 elements, Tier-1 and Tier-2 data replication, data management and operations
  - Application services
    - Reprocessing at the Tier-1 Center, and -if feasible- at Tier-2's
    - Analysis at the Tier-2 Centers
- Continuous operation with increasing functionality and activity level

Planning ahead …

- As of 30 November 2007 - Scope: ~6 months
- Milestones and Items on the following slides are guided by Production and Analysis needs
- U.S. ATLAS Computing Integration Program will translate them into the technical steps sites have to perform

Analysis

- Analysis at Tier-2s
  - Implement analysis queues at all 5 Tier 2's
    - Complete by 15 December
  - Replicate all R13 AOD's (and some selected R12 AOD's) to all Tier 2's
    - Goal: Complete by 31 December
    - Status: Complete as of 25 February
      - Compatibility issues delayed timely completion
      - Resolved by Nurcan plus excellent team effort
- Demonstrate automatic redirection of analysis jobs to all Tier 2's by pathena
  - Goal: Completed by 15 January
  - Status: Complete as of 25 February

Analysis

- Interactive Analysis
  - BNL PROOF farm available to all US ATLAS users for testing
    - Goal: Complete 31 January
    - Status: Delayed
  - BNL PROOF farm in production mode
    - Complete 31 March
    - Status: Likely to be Delayed (will hear more from Sergey)
  - Tier 2 PROOF farms available
    - Complete 30 June
    - Does this fit our model (User account management etc.)?
  - Support Tier-3 activities as part of the Computing Integration Meeting
    - Immediately, ongoing
    - Status: Still no regular participation from Tier-3 institutions

Storage Services

- At Tier-1
  - Evaluate Pinning with SRM v2.2
    - Complete by 21 December
    - Status: Delayed
  - Propose data placement plan for data at Tier 1, including pinning, disk only partitions etc
    - Complete by 31 December
    - Status: In progress
  - Develop and deploy software necessary to mange pinned files
    - Complete 15 January
    - Status: Delayed
  - Disk space reconfiguration according to computing model
    - Complete 31 January
    - Status: In progress
  - Develop and deploy disk-only dCache space management tools
    - Complete 21 December
    - Status: Delayed
  - User space management at Tier 1, including user management, cache cleanup
    - Proposal: Complete 31 December
    - Deployment: 31 January
    - Status: Delayed
  - LFC
    - Test system deployed: Complete 31 December
    - Test system production ready: 31 January
    - Migration to LFC completed for US, assuming successful tests: 28 February

U.S ATLAS Data

- Data Management
  - Deploy storage quota system US ATLAS wide
    - Complete by 28 February
    - Status: Delayed
  - DQ2 data deletion fully operational
    - Complete by 15 December
    - Status: In Progress
  - Complete DQ2 lost file tagging for US
    - Complete 15 January
    - Status: In progress
- What is data flow model in Pathena?
- What if researcher produces data at Tier3?
- How is the decision to archive made?
- Are Tier2's expected to maintain precious data indefinitely?
- User Data Lifetime?
- Consistency Checks?

Operations & Performance

- Incident tracking / Communication
  - Elog deployed & operational
    - Complete 15 December
    - Status: Up and Running at UTA
- Performance
  - Demonstrate 2007 WLCG pledge with 90% average efficiency
    - Complete 31 December
    - Status: More or less complete
  - Demonstrate 90% 2008 WLCG pledge
    - Complete 30 June
    - Contingent to funding situation

Summary

The U.S. ATLAS Facilities have made substantial progress toward an operational integrated computing facility for the start of the experiment

- The facilities, the Tier-1 and the Tier-2's, have performed well in ATLAS computer system commissioning and specific exercises
  - An Integration Program is in place to ensure readiness in view of the steep ramp-up
  - The Tier-2's are ready to provide resources for Analysis (still need the AODs)
  - Excellent contribution of U.S ATLAS Tier-2 Sites to high volume production in 2007
- The BNL Tier-1 serves as the hub and principal center of the US community, with scale-up for data taking underway
- U.S. ATLAS Computing Facility (in particular the Tier-1 Center) needs sufficient funding to be on track to meet the performance and capacity requirements of the ATLAS computing model
  - Need supplemental funds to provide local resources
- Overall, progressing well towards full readiness for LHC data analysis