U.S. ATLAS Computing Facilities

Bruce G. Gibbard

Brookhaven National Laboratory

Review of U.S. LHC Software and Computing Projects

LBNL, Berkeley, California

13-17 January 2003

BROOKHAVEN NATIONAL LABORATORY

Mission of US ATLAS Computing Facilities

- Supply capacities to the ***ATLAS Distributed Virtual Offline Computing Center***
  - At levels agreed to in a computing resource MoU (Yet to be written)
- Guarantee the Computing Required for Effective Participation by U.S. Physicists in the ATLAS Physics Program
  - Direct access to and analysis of physics data sets
  - Simulation, re-reconstruction, and reorganization of data as required to support such analyses

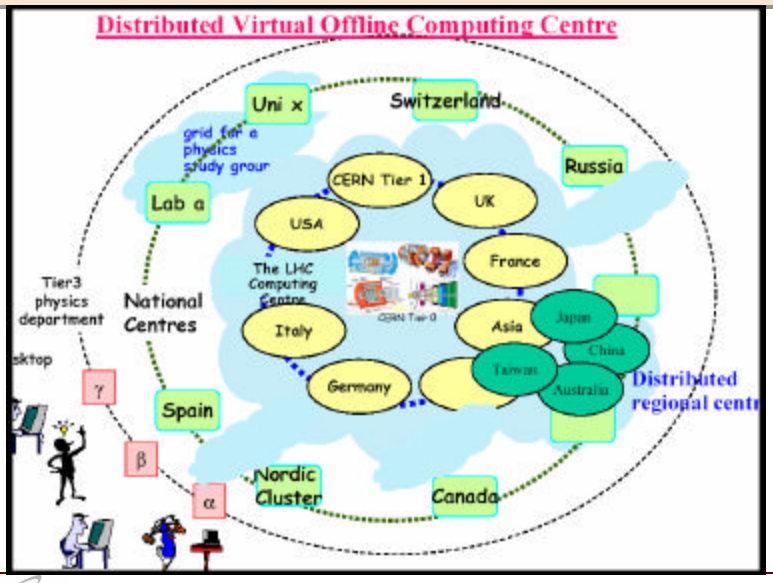
US ATLAS Facilities

- A Coordinated Grid of Distributed Resources Including …
- Tier 1 Facility at Brookhaven – Rich Baker / Bruce Gibbard
  - Currently operational at ~1% of required 2008 capacity
- 5 Permanent Tier 2 Facilities (Scheduled for selection beginning in 2004)
  - 2 Prototype Tier 2's now
    - Indiana U / (*effective FY '03*) University of Chicago – Rob Gardner
    - Boston U – Jim Shank
- Tier 3 / Institutional Facilities
  - 7 Tier 3 sites currently active in Testbed
- Tier 4 / Individual Desktop Users
- Program of Other Associated R&D Activities
  - Grid Projects (**PPDG, GriPhyN, iVDGL, EU Data Grid**)
  - Networking – Shawn McKee
  - US ATLAS Persistent Grid Testbed – Kaushik De

ATLAS Facilities Model

- ATLAS Computing Will Employ the ATLAS Virtual Offline Computing Facility to process and analyze its data
  - "***Cloud***" mediated set of resources including:
    - CERN Tier 0
    - All Regional Facilities (Tier 1's)
      - **Typically ~200 users each**
    - Some National Facilities (Tier 2's)
  - Rules governing access to and use of the Virtual Facility
    - Will be defined by ATLAS management
    - Will apply for all members of the ATLAS ***Virtual Organization*** (VO)
  - All member of VO must contribute to the Virtual Facility
    - Contributions in kind (personnel, equipment) or in funds
    - Contributions to be codified in MoU's agreed with ATLAS management

LHC Computing Facilities Model



ATLAS Facilities Model (2)

- Contribution Accounting
  - Accounting is based on CERN equivalence cost of contribution
  - As with detector M&O, level of contribution is based on number of physicists on the ATLAS author list
  - US author count is larger so contribution will need to be greater
    - MoU yet to be written
- Typically only a subset of resources at a regional or national center are Integrated into the Virtual Facility
  - Only integrated part counts as a contribution
  - Regional or national control over non-integrated portion retained
  - Retained portion is expected to be used to augment resources supporting analyses in which that region or nation is involved

Analysis Model: Having All ESD on Disk

- Enables ~24 hour selection/regeneration passes (versus ~month if tape stored) – faster, better tuned, more consistent selection
- Allows navigation for individual events (to all processed, *though not Raw*, data) without recourse to tape and associated delay – faster more detailed analysis of larger consistently selected data sets
- Avoids contention between analyses over ESD disk space and the need to develop complex algorithms to optimize management of that space – better result with less effort
- **Complete set on Disk at a single Tier 1 vs. WAN distributed across 3**
  - **Reduced sensitivity to performance of multiple Tier 1's, intervening network (transatlantic) & middleware – improved system reliability, availability, robustness and performance – At a $ cost, of course**

Required Tier 1 2008 Capacities by Model

	Tape Based Model	3 Center Disk Model	Standalone Disk Model
CPU (SPECint95)	209	329	500
Disk (TBytes)	365	483	1000
Tape (PBytes)	2	2	2
Disk (GBytes/sec)	10	20	20
Tape (MBytes/sec)	1000	200	200
WAN (Mbit/sec)	4610	9115	9115
		1/3+1/6 of ESD on disk	Add other 2/3 of ESD
	ESD pass each month	ESD pass each day	

- Cost impacts of new models are largely offset (relative to earlier cost estimates) by the combination of the LHC start-up delay and Moore's Law

Evolution of Plan for Tier 1 Facility

- Has Been In Response to Changes in
  - Budgetary Guidance as Discussed by Others
    - A late peaking funding profile
    - History of year to year decreases in funding expectation, especially in the near to mid-term
    - By design or happenstance, decreases have coincided with LHC schedule slippages so as to remain tolerable
  - ATLAS Computing Model & Requirements as Discussed Above
  - LHC Start-up & Data Challenge Schedules

Schedule Changes

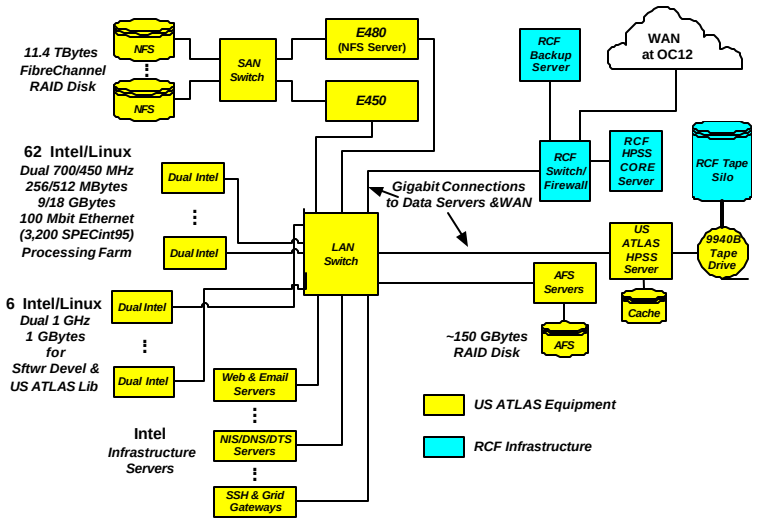
- LHC start-up now delayed 2 years since Nov 2000 review
  - 2005/2006 → 2006/2007 → 2007/2008
- ATLAS Data Challenges (DC's) sliding less dramatically
  - DC0 – 10^5 events: Nov/Dec 2001 → Dec/April 2002
    - Software continuity test
  - DC1 – nx$10^{\sim 7}$ events: Feb/Jul 2002 → July/Early 2003
    - ~1%+ scale test
  - DC2 – 10^8 events: Jan/Sep 2003 → Oct 2003/March 2004
    - ~10% scale test
  - DC3 – 5x10^8 events: Late 2004/Early 2005 ***– Newly Defined***
  - DC4 – 10^9 events: Late 2005/Early 2006 ***– Newly Defined***

Current Regional Center (Tier 1) Status

- Co-located/operated with RHIC Computing Facility (RCF)
  - A great deal of shared expertise and shared operational activities
    - **Intel/Linux, Sun/Solaris, HPSS, SAN/WAN/LAN, Cyber Security, LSF**
  - Some shared infrastructure components
    - **Robotics, backup system, firewall**
    - **WAN connection upgrade in July OC3 → OC12**
  - While of comparable size in 2008, the Tier 1 is currently small relative to RCF capacities being deployed for the RHIC FY 2003 run
    - **3% of 2050 Intel/Linux CPU's totaling 100 kSPECint95**
    - **10% of 115 TBytes of RAID disk @ 3 GBytes/sec**
    - **1% of 4.5 PBytes of robotic tape capacity @ 1 GByte/sec**
  - Near complete Tier 1 functionality with 4.5 FTE's on project (~2.5 FTE's doing direct fabric support) as a result of synergistic relationship with RCF

US ATLAS Regional Center (Tier 1) at BNL

November 2002



US ATLAS Regional Center At BNL

Currently

3.2 kSPECint95

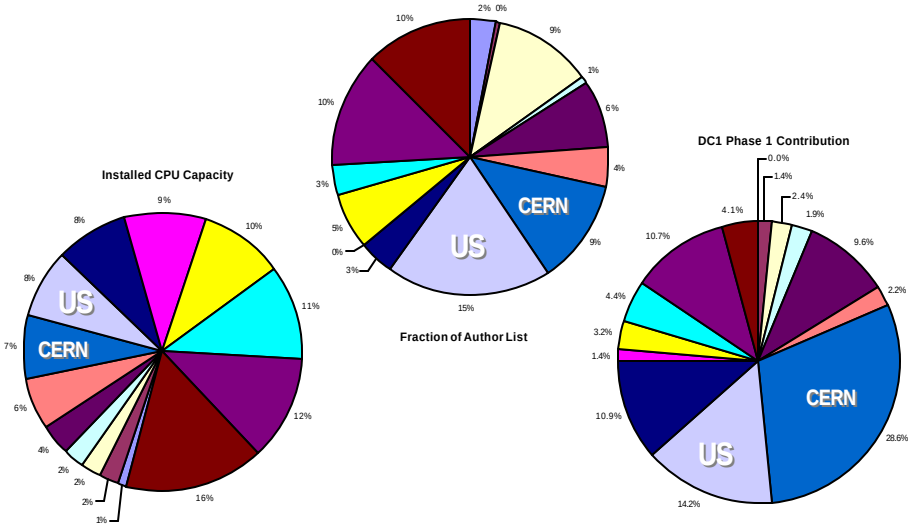
11.4 TB of Disk

30 MB/sec Tape

Tier 1 Utilization

- Current Ongoing Utilization
  - Simulations
    - Most Notably, Neutron Backgrounds
    - Additional Detector Studies – EM Calorimeter Response
- DC1 Phase I: US outperformed all except CERN
  - Installed CPU fraction was 8% of ATLAS total
    - ... compared to US author list fraction of 15%
  - Delivered DC1 Phase I data was 14% of ATLAS total
- DC1 Phase 2: Now Underway
  - Expect US Tier 1 to serve as one of handful of primary data repositories
    - CERN storage costs deemed excessive by ATLAS (~60 TBytes of data)
  - Consequently also expect to server as a primary analysis site
    - Between HLT TDR and physics topics a total of ~100 new users
      - **Perhaps 20 – 25 of them very active**
    - Concerns regarding new facility usage load with limited fiscal flexibility
      - **... but clearly an excellent exercise for the facility**

Facility DC1 Phase 1 Performance



Revised Equipment Spending & Capacity Plan

Tier 1 Capacity Profile

	2001	2002	2003	2004	2005	2006	2007	2008
CPU (SPECint95)	3	3	8	13	25	85	175	500
Disk (TBytes)	0.5	12	12	26	50	163	350	1,000
Disk (MBytes/sec)	40	90	90	520	1,000	3,400	7,000	20,000
Tape (PBytes)	0.01	0.05	0.05	0.10	0.21	0.32	0.86	2.05
Tape (MBytes/sec)	10	30	30	30	90	135	255	375
WAN (Mbits/sec)	155	155	622	622	2488	2488	9952	9952

Tier 1 Captial Equipment Cost Profile (At Year $k)

	2001	2002	2003	2004	2005	2006	2007	2008
CPU	$ 30	$ -	$ 100	$ 60	$ 105	$ 328	$ 344	$ 785
Disk	$ 100	$ 137	$ -	$ 177	$ 169	$ 558	$ 567	$ 1,315
Tertiary Storage	$ 46	$ 25	$ -	$ 120	$ 80	$ 23	$ 80	$ 30
LAN	$ 79	$ -	$ 20	$ 20	$ 90	$ 100	$ 250	$ 250
Overhead	*$ 20*	*$ 13*	*$ 10*	*$ 30*	*$ 36*	*$ 81*	*$ 99*	*$ 190*
Total	**$ 275**	**$ 175**	**$ 130**	**$ 407**	**$ 480**	**$ 1,089**	**$ 1,340**	**$ 2,570**

Revised Staffing Plan

- Major Re-estimation of Staff Levels Conducted
  - Based on support for 2 cycles of production operations for RHIC
  - … and on 2 years of operating a combined RHIC/US ATLAS facility
  - Reflects expectation that significant RHIC & US ATLAS synergy will continue in future
    - Very broad common computing platform and infrastructure base …
    - … and both are now on a path toward Grid based computing model via involvement in the same Grid projects and common local expertise
  - Significant reduction in out year staff level estimate

    25 FTE's → 20 FTE's
  - Ramp of staff up to this level is funding constrained
    - Optimal would be linear ramp to full staff level in '06
    - Budget consideration dictate slow start ramp to full staff level in '07

      … as shown in table

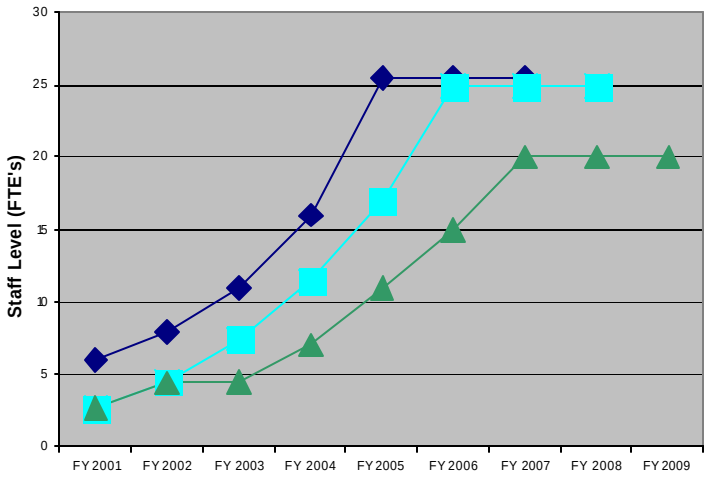
Revised Summary Tier 1 Cost Profile
(At Year $k)

($ Items below include overheads)	2001	2002	2003	2004	2005	2006	2007	2008	TOTAL
Staff Level (FTE's)	*2.7*	*4.4*	*4.5*	*7*	*11*	*15*	*20*	*20*	
Labor (fully loaded salaries)	$ 386	$ 661	$ 709	$ 1,159	$ 1,912	$ 2,738	$ 3,833	$ 4,024	**$15,422**
MST (travel, maint, licen, etc)	$ 167	$ 206	$ 231	$ 337	$ 452	$ 624	$ 875	$ 1,065	**$ 3,957**
Capital Equipment	$ 275	$ 175	$ 130	$ 407	$ 480	$ 1,089	$ 1,340	$ 2,570	**$ 6,465**
Total	**$ 828**	**$ 1,042**	**$ 1,070**	**$ 1,902**	**$ 2,844**	**$ 4,450**	**$ 6,047**	**$ 7,660**	**$25,844**

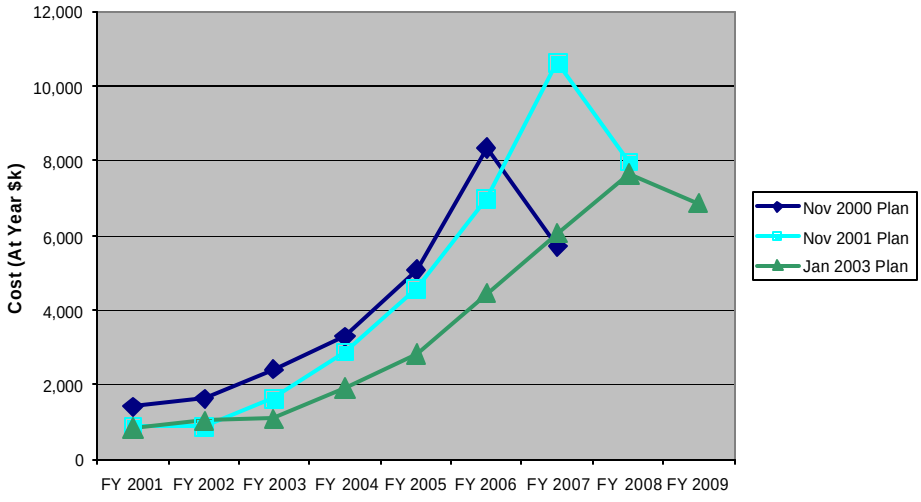
Comparison to Previous Tier 1 Cost Profiles

	2001	2002	2003	2004	2005	2006	2007	2008	Total
Nov 2000 Plan	$ 1,410	$ 1,609	$ 2,397	$ 3,270	$ 5,074	$ 8,346	$ 7,000	$ 7,000	$ 36,106
Nov 2001 Plan	$ 858	$ 857	$ 1,609	$ 2,869	$ 4,584	$ 6,993	$ 10,638	$ 7,993	$ 36,401
Jan 2003 Plan	$ 828	$ 1,042	$ 1,070	$ 1,902	$ 2,844	$ 4,450	$ 6,047	$ 7,660	$ 25,844

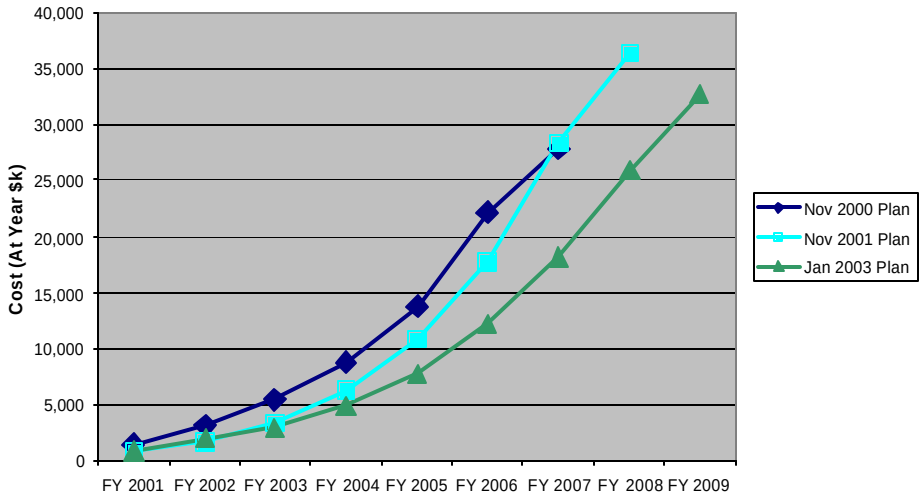
Evolution of Staffing Profile



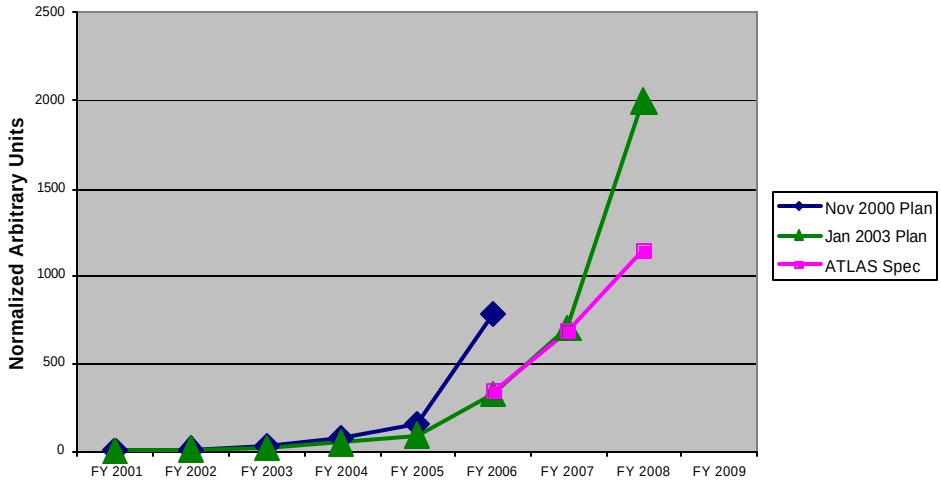
Evolution of Annual Cost Profile



Evolution of Integrated Cost Profile



Evolution of Capacity Profile



Summary of Tier 1 Grid Activities

- ATLAS (& STAR) Grid Activities
  - Partially PPDG Funded – 0.5 FTE (+ one time 0.5 FTE site AA*A*)
  - Grid/Network Monitoring
  - Jason Smith on iVDGL VDT Support Team
  - PPDG Site AA*A* (**BNL, FNAL, SLAC, LBL and JLab Participating**)
    - Interaction Between Grid and Site Security Models
      - **Many Administrative and Trust Issues Must be Addressed**
    - BNL Focus is on User Account Management
      - **Regional Centers *must* allow use by all Virtual Organization (VO) registered members → Need to grant some kind of local account**
      - **Fast Prototyping Tools to Import VO Data and Manage Local Accounts**
- ATLAS Specific Grid Activities
  - Pacman Cache Maintenance of Many Packages for US Testbed
  - Near term need/plan to integrate facility with LCG-1 (for next summer)
    - Orchestrated by Grid Deployment Board

Current Tier 1 Tactical Situation

- Limited FY 2002 funding forced choice between staff and equipment
  - Chose to grow staff by 2 FTE's to current total of 4.5 FTE's
  - Only FY 2002 equipment was from end-of-year supplemental funding, $200K
- Flat funding for 2003 leaves no choices
  - Anticipate no staff growth in 2003
  - Any Tier 1 equipment growth (needed for effective participation in DC2) will depend on repeat of supplemental end-of-year funding; likelihood unknown
- Profiles show:
  - Funding & Staffing are 1.5 - 2 years delayed relative to Nov 2000 plan
  - Capacities & Capabilities are ~1 year delayed (not necessarily inappropriate)
- Once the LHC schedule and agency budgets become predictable, a new detailed look at the Tier 1 plan, cost & schedule is needed