ATLAS WAN Requirements at BNL

Slides Extracted From Presentation Given By

Bruce G. Gibbard

13 December 2004

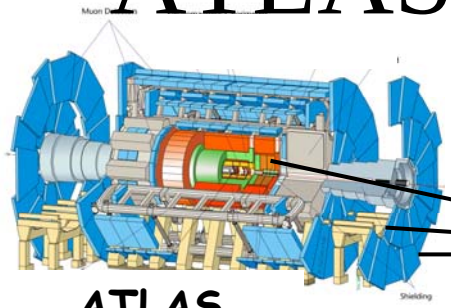
Office of Science
U.S. DEPARTMENT OF ENERGY

BROOKHAVEN NATIONAL LABORATORY

Primary Drivers of BNL WAN Requirement

- BNL has primary responsibility for two DOE programs which involve very large, internationally distributed collaboration and include distributed computing resources
  - The Relativistic Heavy Ion Collider (RHIC) for which it is the host institution
  - US Participation in the ATLAS experiment at CERN's Large Hadron Collider (LHC) for which it is the lead US institution for both the construction project and for computing facilities (US Tier 1 Center)
- For each project, BNL is responsible for:
  - Directly supplying at BNL a major computing facility for storage, production processing and analysis of data
  - Marshaling and integrating additional computing resources from a large number of institutions distributed around the world into a single coherent and effective ***virtual computing facility*** via the Grid and its underlying WAN infrastructure

ATLAS Distributed Computing Model



ATLAS Experiment

~PByte/sec

CERN:Outside Resource Ratio ~1:2
Tier0:(Σ Tier1):(Σ Tier2) ~1:1:1

Online System

< GBytes/sec

Tier 0 +1

CERN ~5M SI2K
>1 PB Disk
Tape Robot

Castor

~10 Gbits/sec

Tier 1

IN2P3 Center

RAL Center

INFN Center

BNL: ~2M SI2K; 2PB Tape Robot

HPSS

2.5 Gbps

Tier 2

~2.5 Gbps

Tier2 Center

Tier 3

Institute

Physics data cache

100 - 1000 Mbits/sec

Workstations

Tier 4

Tier 0: DAQ, reconstruction, archive

Tier 1: Reconstruction, simulation, archive, mining and (large scale) analysis

Tier 2+: Analysis, simulation

Tier 3+: Interactive analysis

US ATLAS Tier 1 Computing Facility

Functions

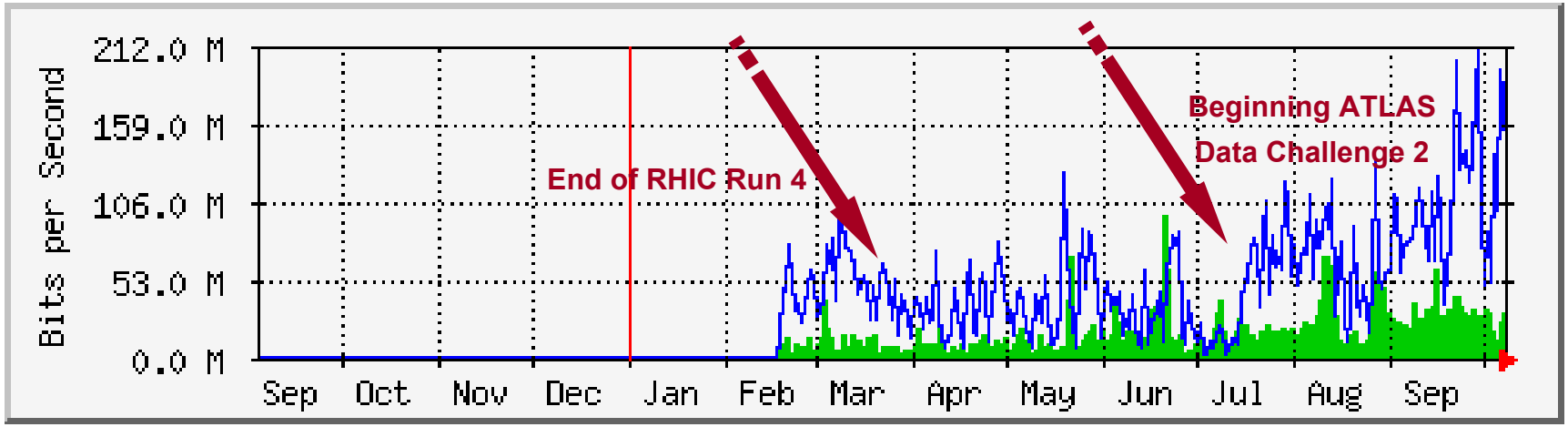
- Serve as primary U.S. ATLAS & ATLAS data repository
- Reconstruction 1/nth of ATLAS data archived at BNL
- Programmatically select and distill reconstructed data
- Support *"Chaotic"* high level analysis by individuals
- Generate Monte Carlo data
- Supply technical support for smaller US computing resource centers

RHIC and ATLAS Capacities at BNL

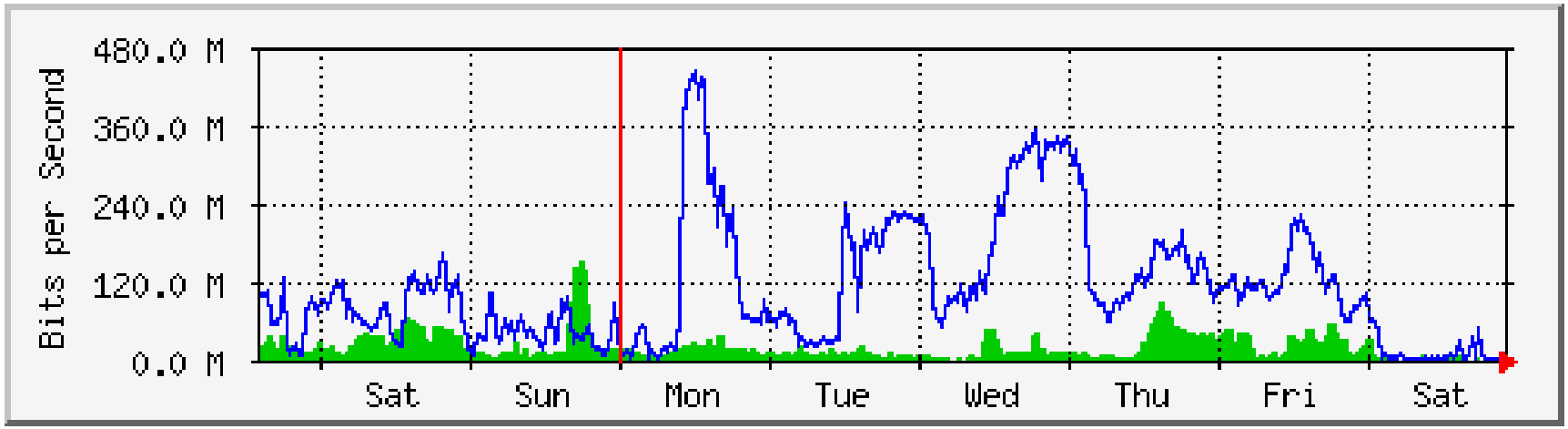
Year	2004	2005	2006	2007	2008	2009	2010
RHIC							
CPU (kSPECint2k)	1200	2999	3916	6122	8337	12025	15302
Disk Volume (TBytes)	400	818	1029	1478	1938	2705	3404
Tape Volume (PBytes)	4.5	5.6	11.2	11.2	22.5	22.5	44.9
ATLAS							
CPU (kSPECint2k)	205	302	842	1807	3930	6246	9720
Disk Volume (TBytes)	24	104	346	778	1730	2768	4324
Tape Volume (PBytes)	0.1	0.2	0.3	0.5	1.7	2.9	5.3
TOTAL							
CPU (kSPECint2k)	1405	3301	4758	7929	12267	18271	25022
Disk Volume (TBytes)	424	922	1375	2256	3668	5472	7729
Tape Volume (PBytes)	5	6	12	12	24	25	50

WAN Utilization

1 Day Average



30 Minute Average

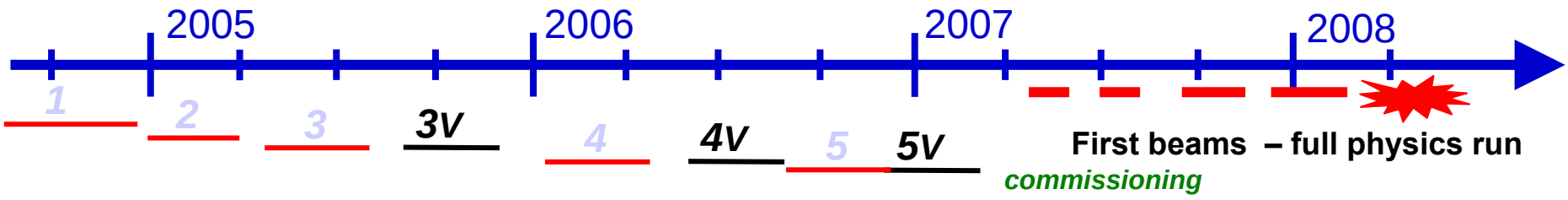


Drivers of ATLAS WAN Requirements

- ATLAS Data Challenges
  - Intended to exercise ATLAS application and Grid/production software, ATLAS computing model in context of current Grid hardware/middleware production versions (LCG, Grid3/OSG)
  - DC1 completed Oct 03
  - DC2 currently on going; Jun 04 => Jan 05
  - DC3 will begin Feb 06
- LCG Service Challenges (robust data transfers stress WAN)
  - Intended to stress test Grid services at maximum possible levels using advance R&D components where available
  - Some detail on next page
- ATLAS Startup and Operations
  - Initial LHC operations currently scheduled for 2nd half of 2007
  - First year of full LHC operations 2008

LCG Robust Data Transfer Service Challenges (Les Robertson)

- Service Challenge 1 – Dec 04 – 2 weeks sustained
  - Basic data transfer – 2 weeks sustained – 500 MB/sec disk => disk to Tier-1s
- Service Challenge 2 – Mar 05 – 1 month sustained
  - Reliable file transfer service – 1 month sustained – 500 MB/sec Mass Store (disk) => Mass Store (disk)
- Service Challenge 3 – Jul 05 – 1 month sustained
  - Acquisition/reconstruction/recording/distribution – Mass Store (disk + tape) – Mass Store (disk + tape)
  - Followed by Tier-0/1 model verification exercise
- Service Challenge 4 – Apr 06 – 1 month sustained
  - As above but including ESD skimming and rates to 1.2 GB/sec at Tier-0
  - Followed by Tier-0/1/2 scaled full model verification
- Service Challenge 5 – Nov 06 – 1 month sustained
  - As above but with rates to 2.5 GB/sec at Tier-0
  - Followed by Tier-0/1/2 100% full model verification



HEP/NP WAN Requirements at BNL

US ATLAS Tier 1 WAN Bandwidth Requirement Estimate

(Mbits/sec)

Year	2004	2005	2006	2007	2008	2009	2010
Remote Site(s)							
Tier 0 (CERN)	52	105	349	874	1,747	1,747	3,494
Tier 1's (~2 Peer sites)	37	75	250	624	1,248	1,248	2,496
Tier 2's (5 USA satellite sites)	64	128	428	1,069	2,139	2,139	4,278
Tier 3-4 (150 Individual users)	95	190	632	1,581	3,161	3,161	6,322
Total	**249**	**498**	**1,659**	**4,148**	**8,295**	**8,295**	**16,590**

BNL HEP/NP WAN Bandwidth Requirement Estimate

(Mbits/sec)

Year	2004	2005	2006	2007	2008	2009	2010
US ATLAS Tier 1 Req.	249	498	1,244	4,148	8,295	9,954	16,590
RHIC Computing Facility Req.	200	500	1,023	1,286	1,847	2,422	3,381
TOTAL	449	998	2,267	5,433	10,142	12,377	19,971
BNL HEP/NP Requirement	***OC12***	***OC48***	***OC48***	***OC192***	**2 X λ**	**2 x λ**	**3 x λ**

Qualitative Issues As Well

- Need to share effectively between a number of very different requirements (programs & services) – ***need differentiated services*** (be able to specify Quality of Service)
  - Long term programmatic bulk transfers (CERN => BNL, BNL => LBNL, BNL => Riken, etc.) – ***background activity?***
  - Short term programmatic bulk transfers (BNL => Tier 2's & Peer Tier 1's. etc.) – ***scheduled activity?***
  - High priority smaller chaotic transfers (Support for interactive analysis, calibration & metadata requests, etc.) – ***priority driven preemptive activity?***
- Need the ability to manage the network as a critical resource; much as resource scheduler/batch managers currently manage CPU resources in a multi-user environment
- MPLS/QoS project intended to address this need

WAN Dependence of Virtual Facilities

- Predictability and dynamic configurability are required to optimize use of network depended computing resources
- Greatly increased reliability and robustness of WAN is needed because of interdependency of distributed components of ***Virtual Facilities***
  - WAN is coming to serve as the backplane of a global computer (or at least as the LAN of a tightly coupled global computing facility)
  - WAN failures imply ***major*** disruption of a large increasingly monolithic widely distributed computing facility