

RCF Operations and Plans

Michael Ernst

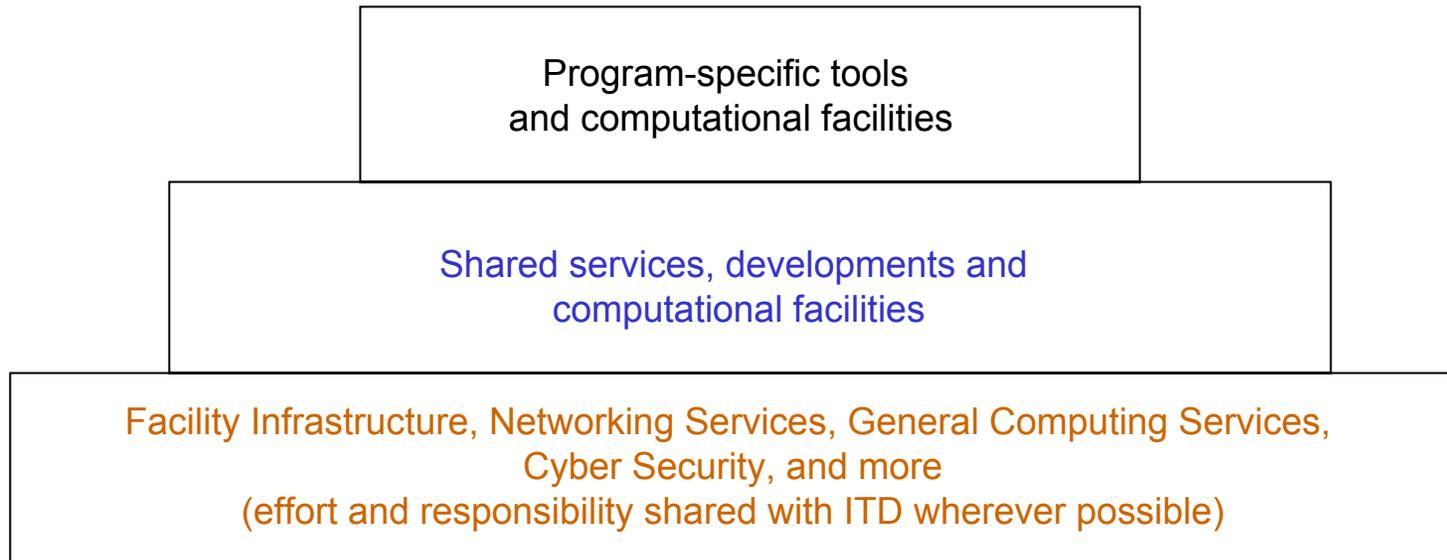
**DOE/Nuclear Physics Review of RHIC
Science and Technology**

7 - 9 July 2008

RHIC & ATLAS Computing

A layered model for providing support for Science Programs

- Share as much as possible - leverage, consolidate, focus on robust solutions – drive down risk and cost of operations



- No capacities for Research and only little for Development

RCF Mission and Scale

➤ Mission

- Online Recording of Raw Data
- Production reconstruction of Raw Data
- Primary Facility for Data Selection and Analysis
- Long time Archiving and Serving of all Data

➤ Scale

- Authorized staff of 20 FTE's
- Historically ~\$2M/year equipment replacement funding (25% annual replacement) – 2006 limited to \$1.3M, last year to \$1.7M, current year to \$1.7M again
 - Funds primarily used to improve storage and network infrastructure, and to address obsolescence
- Growth beyond originally planned scale will require an increase in funding

Computing Requirements Estimate

- A Comprehensive Long Range Estimate done by PHENIX, RCF and STAR in Fall/Winter 2005
 - Conclusions published as part of “Mid-Term Strategic Plan: 2006-2011 For the Relativistic Heavy Ion Collider”
- Input is Raw Data Volume for Each Species & Experiment by Year
- Model for Requirements Projection
 - Assume Facility resources need to scale with Raw Data volume
 - With adjustable parameters reflecting expected relative ...
 - Richness of data set (density of interesting events)
 - Maturity of processing software
 - Number of reconstruction passes
 - ... for each experiment, species, and year
- The assumption of planning for an annual equipment investment to satisfy the growing needs has turned into an administration of shortages to cover the bare minimum (allow the facility to continue to function)
- Given the revised funding profile and by incorporating recent information we (PHENIX, STAR and RCF) need to reconsider and eventually revise the plan

Requirements Estimate for a Particular Running Scenario

	FY '06	FY '07	FY '08	FY '09	FY '10	FY '11
Annual Requirement						
<i>Real Data Volume (TB)</i>	1700	2700	3500	4500	7200	8600
<i>Reco CPU (KSI2K)</i>	600	1000	2900	4700	8500	9700
<i>Analys CPU (KSI2K)</i>	310	570	1800	2700	4600	5400
<i>Dist. Disk (TB)</i>	220	480	1500	1700	3000	3600
<i>Cent. Disk (TB)</i>	30	60	190	260	450	560
<i>Annual Tape Volume (TB)</i>	2000	3200	4200	5400	8700	10300
<i>Tape bandwidth (MB/sec)</i>	690	920	920	1700	2100	2300
<i>WAN bandwidth (Mb/sec)</i>	1400	2000	2100	4300	5700	6700
<i>Simulation CPU (KSI2K)</i>	110	200	610	1000	1800	2100
<i>Simulation Data Volume (TB)</i>	330	530	710	900	1400	1700
Installed Requirement						
<i>CPU (KSI2K)</i>	2100	2800	6800	11800	20800	27700
<i>Dist. Disk (TB)</i>	480	720	1900	2600	4300	5700
<i>Cent. Disk (TB)</i>	200	200	290	400	650	880
<i>Tape Volume (TB)</i>	4300	7500	11800	17200	25900	36200
<i>Tape bandwidth (MB/sec)</i>	920	1400	1600	2500	3300	4000
<i>WAN bandwidth (Mb/sec)</i>	1500	2700	3500	6100	8700	11000

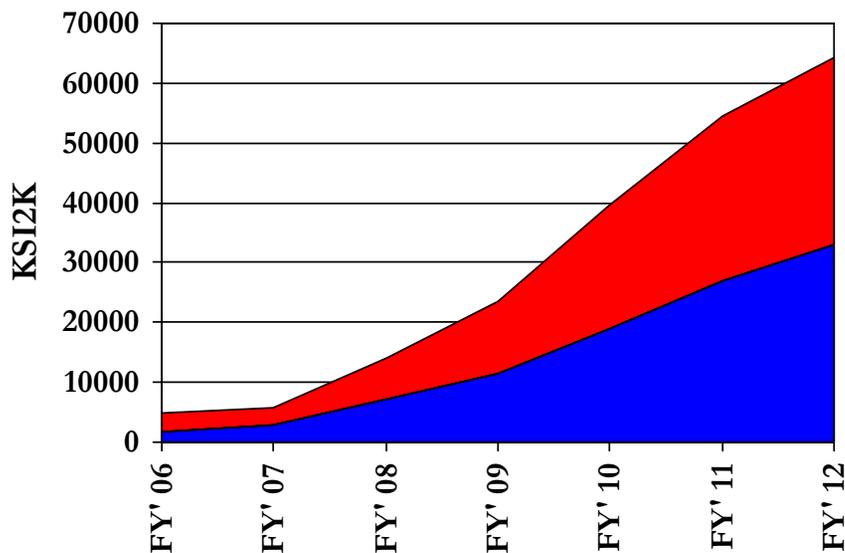
(anticipated as of 2/2006)

Funding Profile (\$K)

	FY '06	FY '07	FY '08	FY '09	FY '10	FY '11
<i>CPU + Distributed Disk</i>	270	770	1360	790	1270	1960
<i>Central Disk</i>	150	250	400	330	450	310
<i>Tape Storage System</i>	640	590	250	1060	570	250
<i>LAN</i>	120	190	250	270	320	180
<i>Overhead</i>	130	200	250	270	290	300
Total Annual Cost	1310	2000	2510	2720	2900	3000

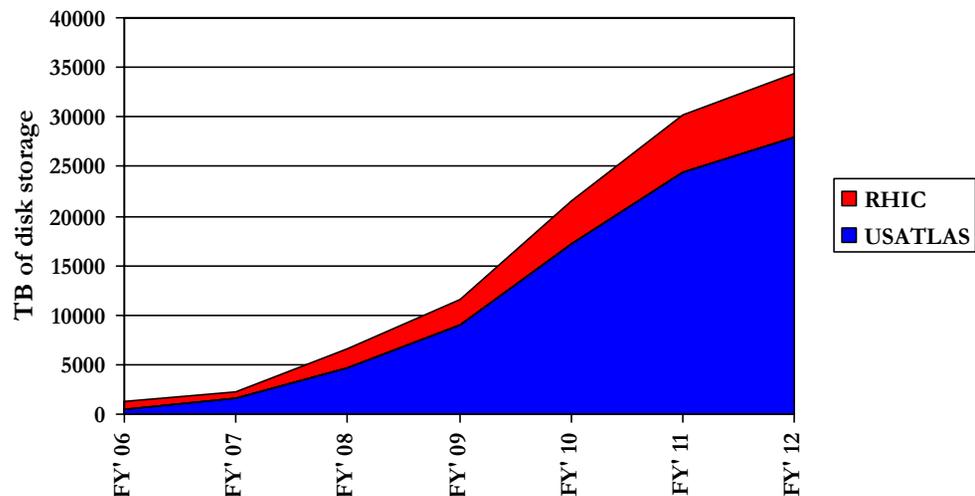
Revised Mid-Term Plan (\$K) 1700 1700 2000 2500 3000

Expected Computing & Disk Storage Capacity Evolution at the RACF



Processing Power
65 MSI2k in 2012

Disk Storage Capacity
35 TB in 2012



RCF Staff

- Current authorized staff level: 20 FTE's
- Excellent synergy in the context of a co-located ATLAS Tier-1 Center in terms of operations
 - Very high level of commonality
 - A dramatic divergence in technical directions could change this, but this seems very unlikely
- It does not allow for aggressive involvement in new technologies
 - Effort spent primarily on Integration and Operation

	Current FTE's	Target FTE's
Linux Farms	3.5	3.5
Mass Storage	4.2	4.2
Disk	2.5	2.5
User Support	2.9	2.9
Fabric Infrastructure	2.6	2.6
Wide Area Services	1.8	1.8
Administration	1.5	2.5
Total	19.0	20.0

Strategies and best practices for RHIC Computing

- Drivers that mandate a strategy of continuous refresh of computing facilities
 - Maintaining old or non-aligned (with the current/future program) tools and software infrastructure is costly in effort
 - Each unique solution costs multiple FTEs (at Experiments and RCF)
 - Robotic storage and tape technology must move forward – costs of robot slots must figure into the economic model
 - Density of tapes doubles every ~ 3 years
 - Strategy is to migrate data and keep “online” in robotic storage
 - **Requires additional tape drives to copy data**
 - Sharing of storage resources between Programs is essential

RHIC replenishment of Computing Facilities

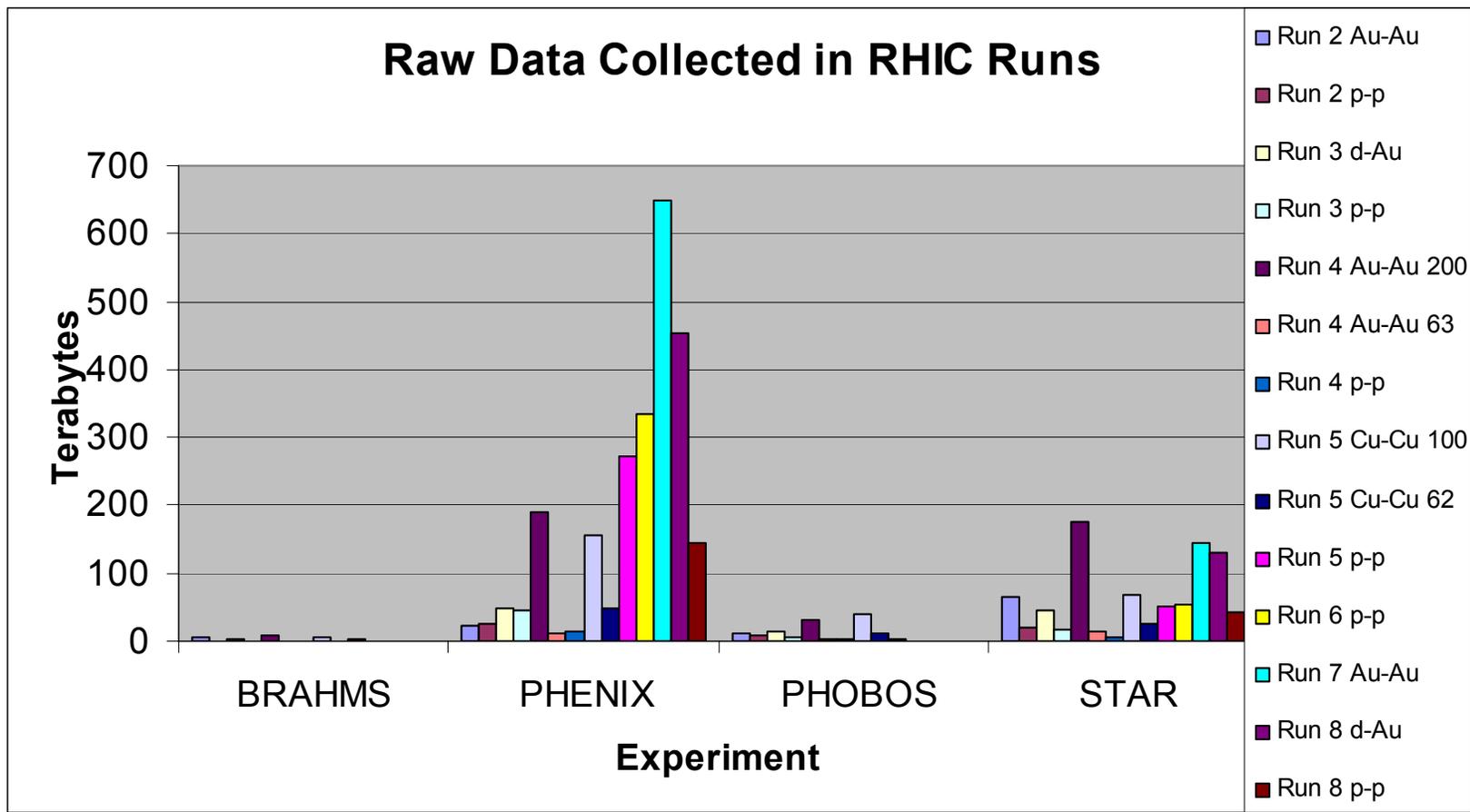
- About 25% of the CPUs will have to be replaced each year
 - Expect heavy demand on analysis computing due to PHENIX & STAR Detector and DAQ upgrades (higher trigger rates)
- About 33% of the disks will have to be replaced each year
- Servers and Network equipment will have to be replenished on either a 3 or 4 year cycle

RCF Capacities deployed as of June 2008

Prior to FY08 purchases	PHENIX	STAR
CPU (MSI2K)	~1.5	~1.8
Disk (TB)	~540	~520
Tape (PB)	4.0	2.2

- Will barely meet the required capacities as outlined in mid-range plan once FY'08 capacities are installed (~August, 2008)
 - Shortfall: 800 kSI2k and 200TB distributed disk

A lot more Data archived in 2007 and 2008 – and a lot more to come ...



PHENIX and STAR are working on Detector and DAQ upgrades that will increase the amount of data by a factor of ~2-3 for PHENIX and a factor of up to 10 for STAR

PHENIX – Two Upgrade Components

In ~2010 the working assumption is a 50KHz collision rate in a very narrow (+/- 10cm) vertex range (currently +/- 35cm).

- Two orthogonal upgrade projects are both going to significantly increase the data volume.
- PHENIX is adding Silicon detectors and new calorimeters. In particular the Si detectors increase the PHENIX channel count dramatically (factor of ~4), increases the data volume by a factor of 2 - 3
- Need to upgrade DAQ rate capability to cope with the increased Luminosity and collision rate (CDR underway)

Current and expected Event Size

System	Au+Au Central	Au+Au MB	p+p
Existing Systems	389 kB	191 kB	88 kB
VTX pixels	90 kB	90 kB	90 kB
VTX strips	39 kB	xx kB	15 kB
FVTX	165 kB	xx kB	44 kB
NCC	64 kB	xx kB	3 kB
MuTrig	0 kB	xx kB	0 kB
Total	747 kB	xx kB	240 kB

Annual Data Volumes – Recent and expected

Year	2007	2008	2009	2010	2011	2012
Raw Data [TB/year]	650	590	1200	1600	2500	2500

STAR Homogenous Software Framework

- Historical separation of online / offline computing framework was a built-in model (light weight online framework)
 - Seemed like a good idea initially
 - NOT desirable in workforce constraint situation
 - “thin” DAQ group, stretched S&C team
 - Delays and out-of-sync features between offline and online (both directions)
- Resources made available to integrate online DAQ code to offline
 - Full integration achieved by the end for Run 8
 - Numerous immediate benefits
 - DAQ1000 and new detector data available to offline
 - DAQ1000 based sector ready for reconstruction
 - Offline EventDisplay available to online
 - Online Data Quality Assurance plots can run offline and vice versa
 - Framework, workforce
 - Code development effort reduced (common code, common framework) in the long term
 - Offline code and build frameworks can be used in online (validation, test suite, regression tests)
- Achieved a common online / offline STAR build / run-time framework
- Preserving the valuable workforce for (more) creative activities
- Allowing to address advanced tasks seamlessly such as
 - quasi-online, offline high level trigger
 - Automated calibration, automated data quality assurance

STAR DAQ Upgrade (DAQ1000)

DAQ upgrade is raising questions and could require a change in the data model

- Possible high level trigger quasi-online event reconstruction
 - Discussion at management level; no requirements yet
 - The idea would be to ship data to (the) processing farm(s) (before it is migrated to Tape) and use the CPU there
 - **Leverage existing resources, existing personnel**
 - **Need to understand implications on network topology**

Current and expected Data Volume

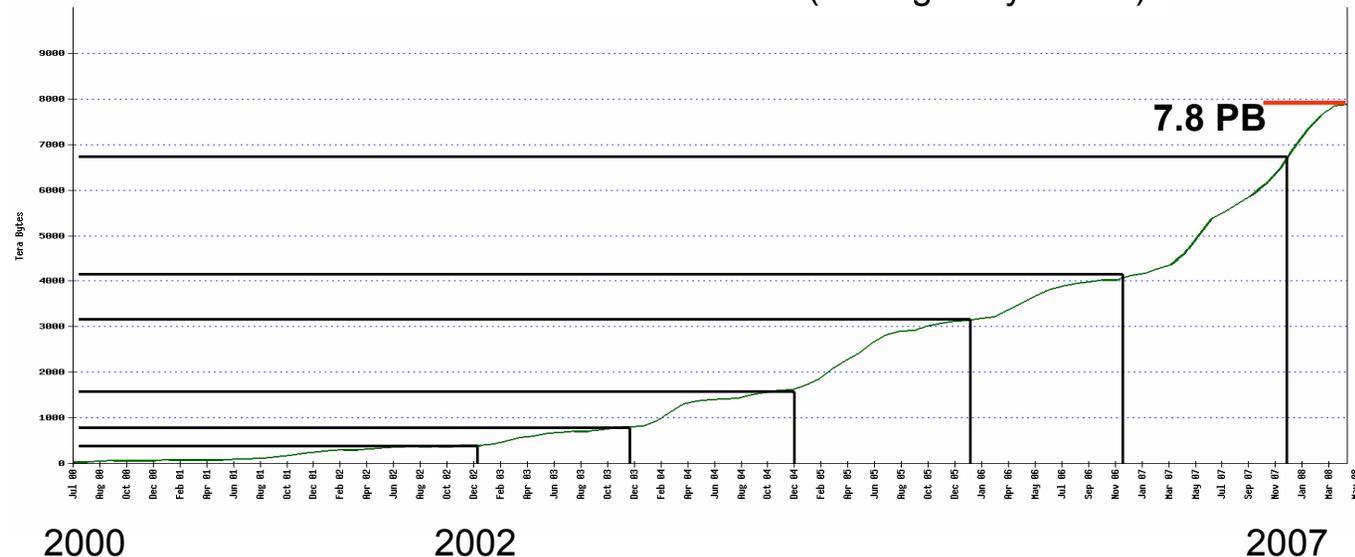
	Projected TB on tape (all)	Projected Raw TB	Acquired/re-scoped Raw	Expected Raw for embedding (15% level)	Expected derived MuDST TB (1 pass, MuDST only) Tier 2 scale
FY08	870	115+320	165 (37%)	16	33
FY09	1720	220+640	650 (<i>p+p only</i>)	65	130
FY10	3000	500+1000	1500	150	300
FY11	4160	680+1400	2080	208	416
FY12	4160	680+1400	2080	208	416

Managing Petabytes of Storage

- At RACF we currently have nearly 8 Petabytes of near-line (actively used) long lived tape storage
 - Projecting > 30 PB by the end of FY 2010
- Significant effort required to assure
 - Integrity and Protection of data
 - Access to data – locally and globally, with enormous peak loads
 - Standard Grid access protocols (SRM)
 - High-performance storage management solution (dCache & XrootD today)
 - Need to look at next generation solutions while scaling up and operating

Have migrated almost all data from end-of-life tape libraries (In addition: (9940B => LTO3 exchange at no cost)

Data Volume archived at the RACF (managed by HPSS)



Mass Storage System Milestones

Substantial gains in HPSS staff understanding in following areas

- Operations – better identification and remediation of system failure modes
- Capacity planning – more complete performance testing to guide future hardware purchases
- Disaster recovery – exercised recovery plans

Other Areas

- Moved data on 8000 9940B tapes to LTO-3 at no cost for media.
- Significant number of new recorded metrics and logs
- Closed many “feedback loops” to automatically adjust HPSS configuration based on system condition and load profiles.

Run 9 Preparation

- Sinking of raw data at 300 MB/sec/per experiment demonstrated during Run 8.
- Combined PHENIX & STAR 600MB/sec demonstrated with current hardware.
- Believe that system can handle up to 1-1.2 GB/sec when combined with ATLAS operation
- Upgrade of Network Link to 10 GE between STAR Counting House and RCF in Fall 2008
- Areas of Concern
 - DST disk cache NOT designed for LTO-3/4 bandwidth
 - Already experience stability issues due to load
 - ~50% loss in mounts/hour in 9940B/LTO-3 to LTO-3/4 transition
 - >5GB file size for peak LTO-4 performance
 - LTO-4 may require 10GE upgrade of inter-mover network

The Volume Constraint – The PHENIX Analysis Model

- Starting with Run4, volume of reconstructed data (~80TB) was too large to keep all files resident on central disk for random access
 - Even before that, there were problems with disk performance at peak times
 - Cumulative effect: variety of data sets over the years means that users today still need access to files dating back to Run3
- Several evolving strategies
 - Produce smaller files with reduced amount of information that can be kept on disk: electron files, high momentum track/cluster files
 - Organize analyses that need to go over the entire data set: the “Analysis Train”
- The initial idea of an “analysis train” evolved from mid ‘04 to early ‘05 into the following plan
 - Reserve a set of the RCF farm (fastest nodes, largest disks)
 - Stage as much of the data set onto the nodes’ local disks; run all (previously tested on ~10% data sample: “the stripe”) analysis modules
 - Delete used data, stage remaining files, run, repeat
- One cycle took ~ 3 weeks
 - Very difficult to organize, maintain data
 - Getting ~200k files from tape was very inefficient
 - Users unhappy with delays

Slides by C. Vale, PHENIX Computing Coordinator

From the Analysis Train to the Analysis Taxi

➤ Since ~ summer '06

- Add all existing distributed disk space into dCache pools
- Stage and pin files that are in use (once!)
- Close dCache to general use, only users phnxreco (mostly write) and anatrain (read/write) have access: performance when open to all users was disastrous - too many HPSS requests, frequent door failures, ...
- Users can “hop in” every Wednesday, requirements are: code tests (valgrind, insure), limits to memory and CPU time consumption, approval from WG for output disk space
- Typical time to run over one large data set: 3-6 days

➤ Currently used by ~300 different PHENIX Analysts

The data

Entire data set(s) staged from HPSS into dCache disk (once) and kept there

New rides start every week

Condor jobs are submitted for each “fileset” (~10GB chunk of input data), which is then copied from dCache into the local area of the executing node

All the modules that need a given fileset run over it

Database keeps track of failed jobs for each module, which are then resubmitted

Develop and test analysis code using small central disk-resident sample

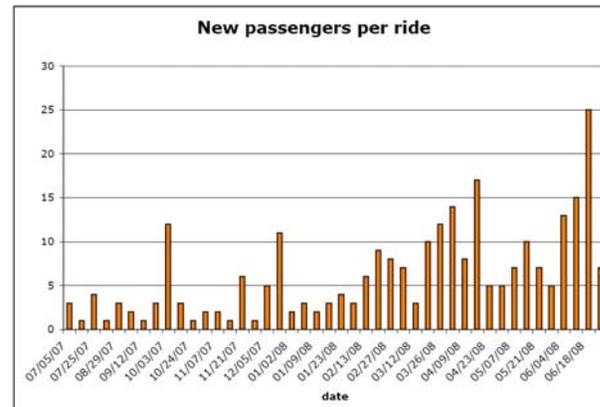
Get approval from WG for usage of space for analysis output

Check-in the code and fill a web-based form to get a ride

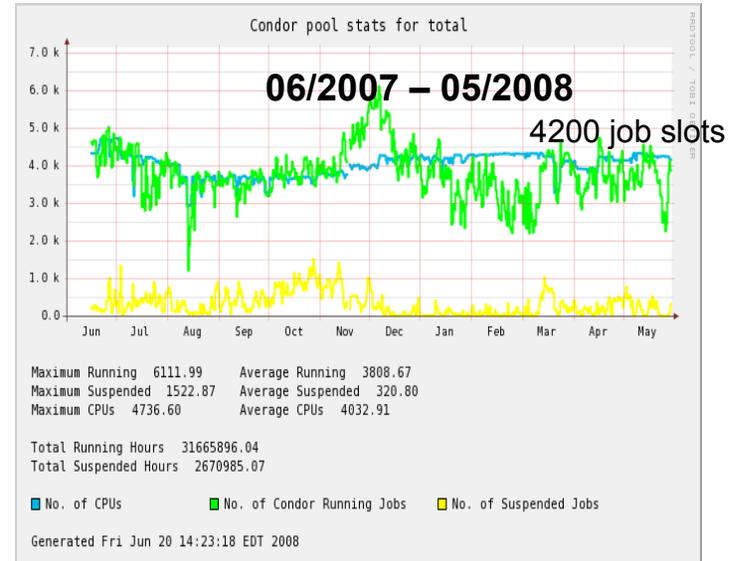
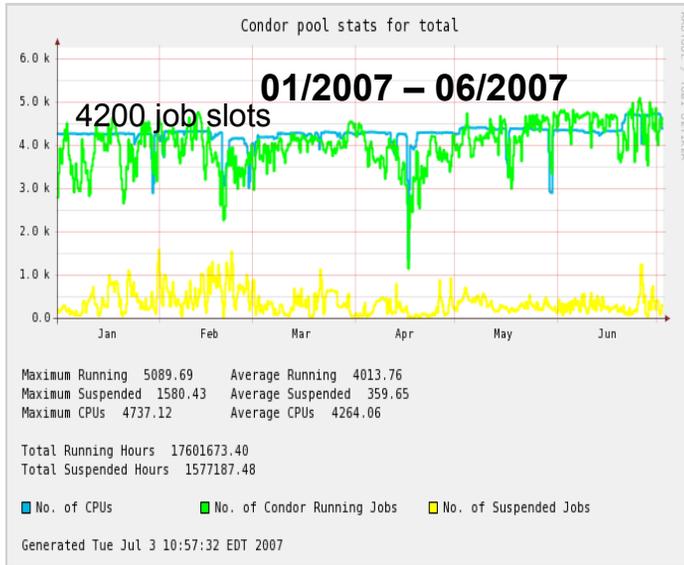
Can check on the status of each module's progress online, as well as deactivate/reactivate it

The users

The process

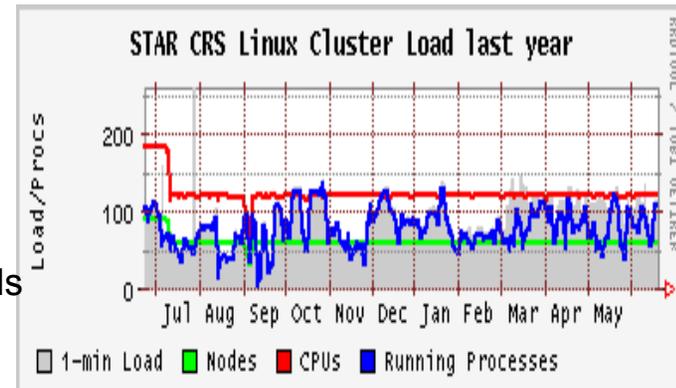
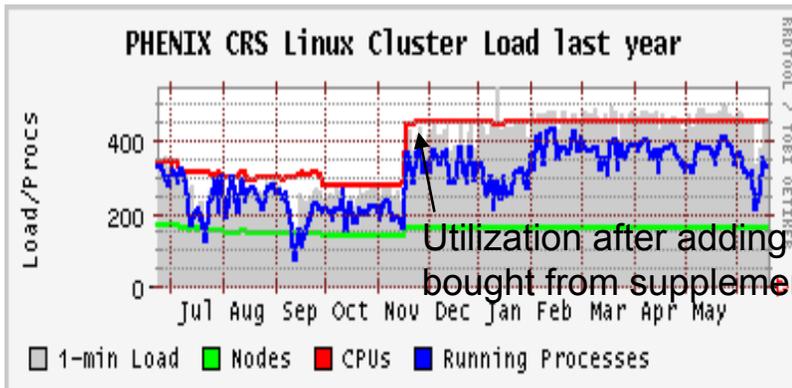


Condor Occupancy (RHIC & ATLAS)



➤ Occupancy remained at 94% between the two periods

Utilization of PHENIX and STAR Production Farm



“Full Function” Disk Service - From last year’s Review

➤ Read/Write (Posix compliant), reliable, high performance and high availability – NFS served RAID systems

- Historically

- ~150 TB of Sun served RAID 5 disk
- ~70 TB of Panasas (appliance) served RAID 5 disk

- Acquisition in 2006

- ~100 TB of Nexsan & Aberdeen Linux served RAID 5/6 disk

- Movement to lower Tier of RAID disk vendors last year

- Product from expensive vendor failed to fulfill expectations
- Inexpensive RAID systems unable to sustain the load
 - **Too many concurrent processes**

- Very bad situation in early 2007

- Many service disruptions due to old and unreliable equipment
- Services distributed on too many different products
- Negative impact on user efficiency (losing jobs, eventually losing data)
- Two FTE’s constantly occupied to keep the service operational

BlueArc Operations & Outlook

- Central Disk Consolidation: BlueArc Titan Cluster w/ 191TB usable space
- Stability of the system (in production since ~9 months) has been good:
3 unplanned service outages
 - 1 configuration problem (slow fail over)
 - 2 instances HW problems (failed to fail over)
- ~3 instances of degraded performance (single head failure)
- 7 disk failures out of 1056 disks in the system
- Outlook for FY'08 procurement
 - Upgrade from 3 heads to 2 next generation Titan 3200 heads
 - 2x10GbE Network upgrade
 - Addition of 65 TB of SATA storage

Grid and Network Services

- **Computing models of RHIC Experiments incorporate Grid Technology**
 - **Desire (necessity?) to utilize substantial distributed resources is driving evolution towards Grid Computing**
 - **Started with simulation, moving towards analysis**
 - **LBNL, Wayne State, NPI / Prague, KISTI (in preparation) etc. for STAR**
 - **Riken, Vanderbilt, IN2P3, etc. for PHENIX**
 - **Same staff engaged in U.S. ATLAS Grid effort also supports RHIC wide area distributed computing with**
 - **Support for Grid tools and services as well as network expertise**
 - **GridFTP, SRM, ...**
 - **High volume network transfer optimization**
 - **Support for involvement (of STAR) in Open Science Grid**
 - **OSG software deployment and integration of resources into OSG**
 - **OSG administration**

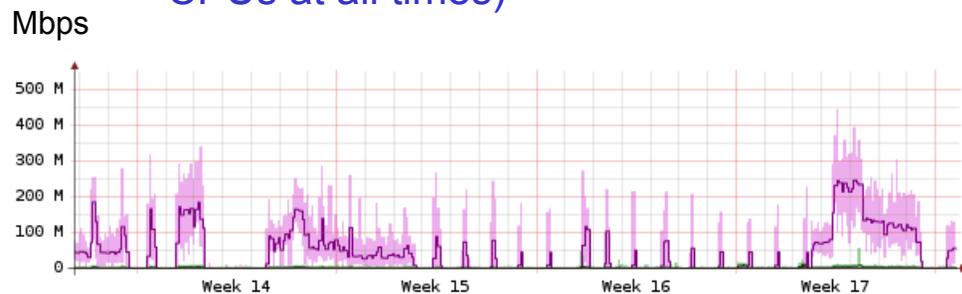
STAR Grid Computing Achievements

➤ Use of the Star Unified Meta Scheduler (SUMS) showing no sign of scalability limitations

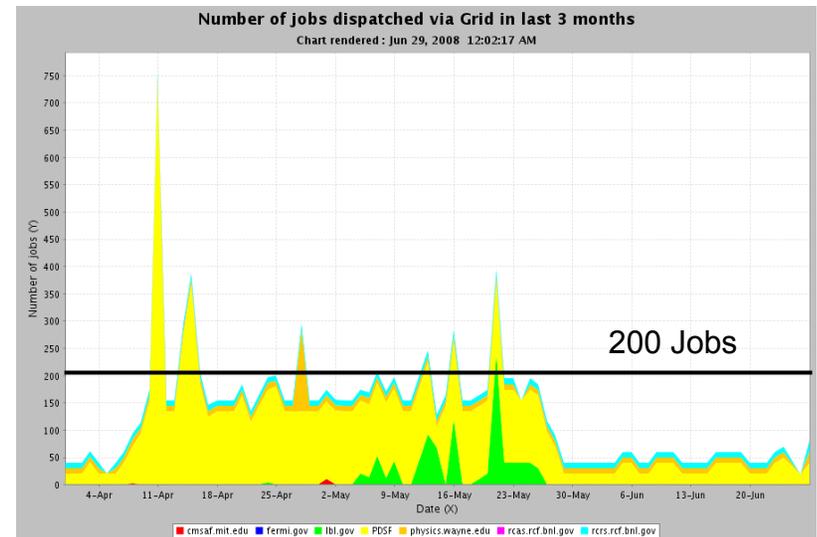
- Daily at RCF / PDSF (different batch system, same job description)
- Expanding to Tier-2 for seamless integration of SE
 - Prague and DPM (below)
- Used in Grid context daily (next slide)

Data Transfer to NPI / Prague pilot Tier-2

- Using Storage Resource Manager (SRM)
- SUMS used seamlessly with different SE
- Fully functional analysis site (100-150 CPUs at all times)



Grid Jobs in last 3 Months



Can submit from BNL to other sites

- PDSF main contributor on this time slice
- Fermilab (OSG) also a significant provider of on-demand resources

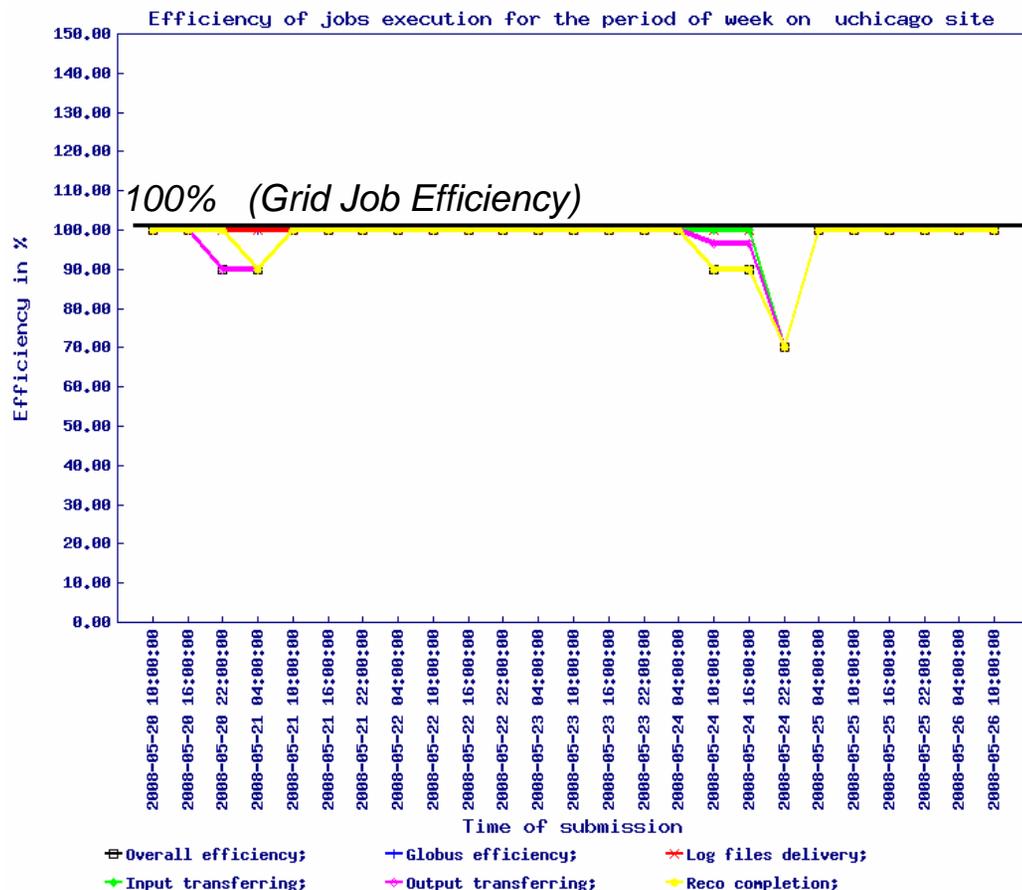
Slides by J. Lauret, STAR Computing Coordinator

STAR Computing Achievements

- Grid job stability outstanding
 - Efficiency > 97%
 - Operation support from OSG helps
 - **All Monte-Carlo production has moved to Grid-based operation**

- Investigation of Virtual Cluster (VC) / Cloud computing
 - Similar efficiencies
 - Full STAR reconstruction can be run
 - Full data flow validated

- VC would allow
 - Harvesting more resources (embedding production)
 - Make easier provisioning of the STAR software stack to ANY STAR site with minimal effort

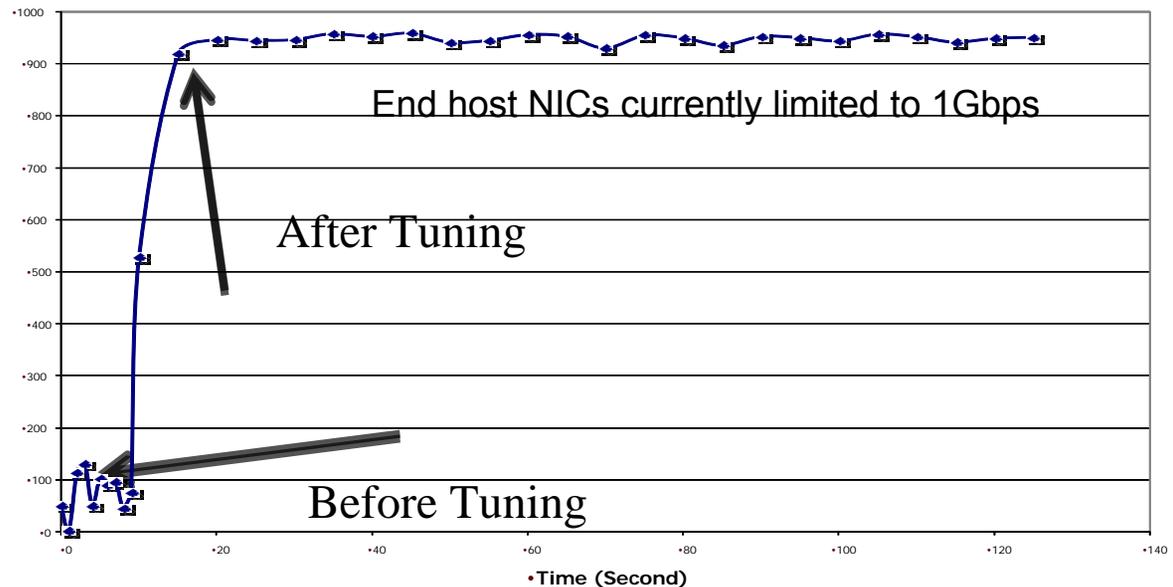


Possible STAR Tier-0 at KISTI / Korea

➤ Tier-0 at remote site is shaping

- Network tuning underway
- STAR intends to transfer data in quasi real-time
- Targeting real-data processing
 - Decrease latency for analysis
 - To increase production and to make more resources available to analyses
 - More production passes
 - Would make some resources available for high level trigger

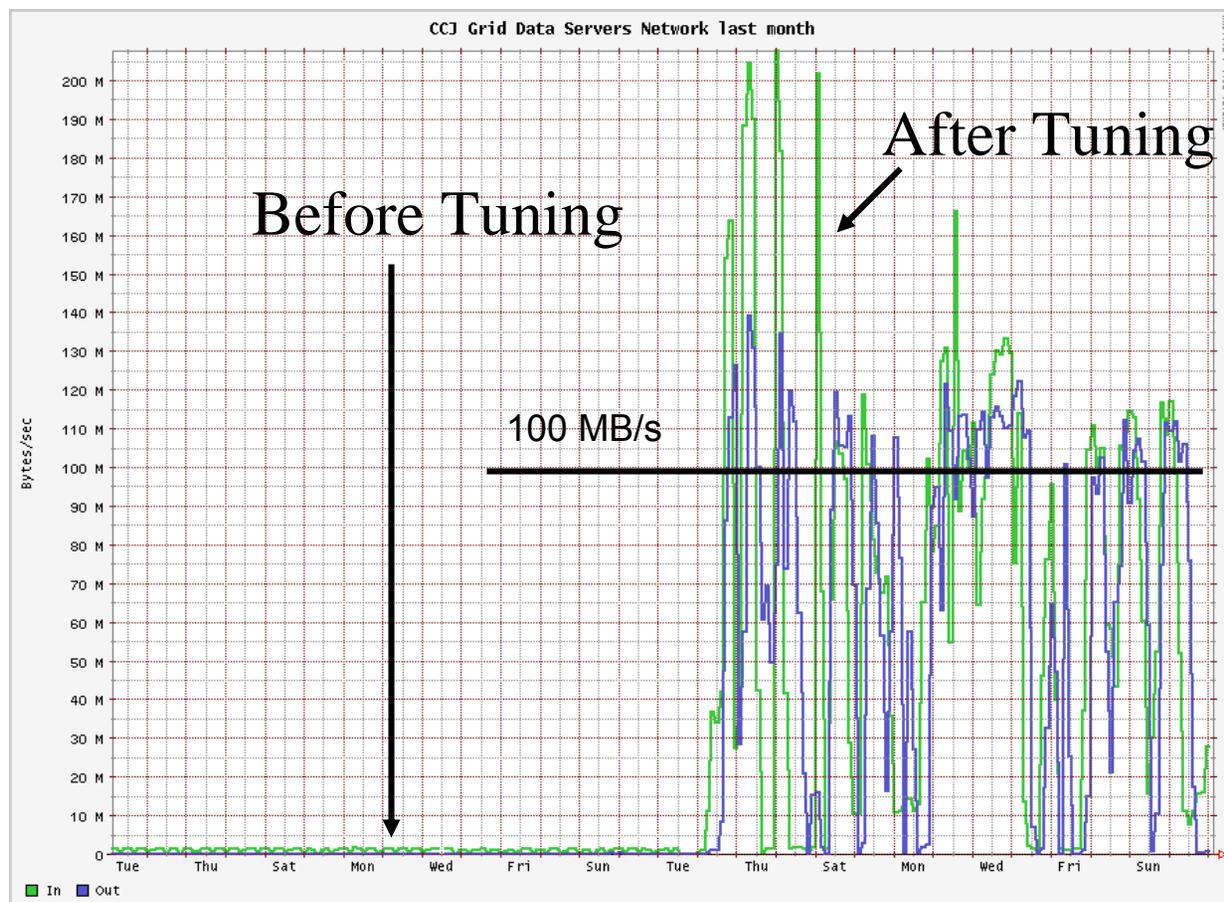
Preliminary results from
Evaluation of BNL / KISTI
connectivity



PHENIX World-wide Distributed Production

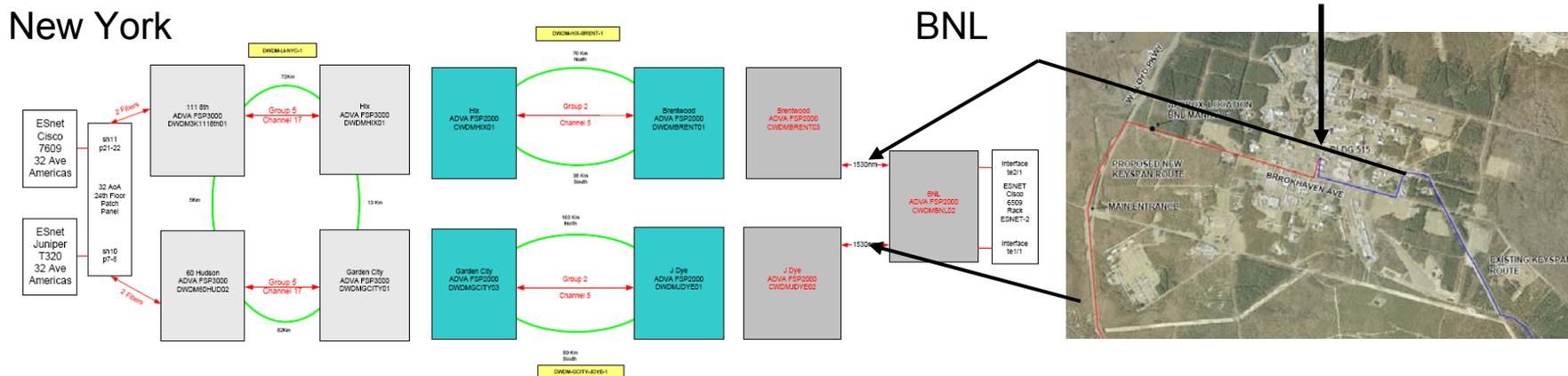
In view of limited resources at BNL export of RAW data to collaborating Institutions In Japan (CCJ) and France (IN2P3) as part of the PHENIX standard production methodology

Since 2005 production data transfers at unprecedented scale (up to 600TB per Run) utilizing Grid transfer technology (GridFTP) across trans-oceanic / long latency networks



Wide Area Network

- BNL's current WAN bandwidth provided by ESnet is 20 Gbps
 - 10 Gbps best effort IP (shared by entire lab) and 10 Gbps Lightpath to CERN (ATLAS)
 - 10 Gbps will be added for ATLAS Tier-0 / Tier-1 connectivity in Sep 2008
 - Excellent technical support from ITD Networking and ESnet
- Connection has now the desired redundancy and diversity between NY and BNL



Physical Infrastructure – From last year's review

- Have reached limits in all areas
 - Without additional space RACF will not be able to accommodate the next robot and the upgrade to processing power and disk storage
 - Reallocation of existing space to RACF allows 2008 expansion
 - ❑ Additional power & cooling is needed each year
 - Need expansion of space in 2009 and beyond
 - ❑ Working with ITD, BNL Plant Engineering and BNL Management on a plan
 - ❑ **Very tight schedule**
 - ❑ **Funding still not (entirely) secured**
 - ❑ **Progress is not as good as we had hoped for**
 - ❑ Technical and organizational problems
 - ❑ Improving since end of last year
- **This is our top concern at the moment**

Progress on Physical Infrastructure

1. Renovation of existing area of ~2000 sq. ft. adjacent to the computing facility, ready for occupancy in Oct 2008 \$1,200k
 - Funding lined up, construction has started
2. Data Center Expansion (6,400 sq. ft.), ready for occupancy in July 2009 \$4,750k
 - Funding lined up, design complete, construction to start ~Oct. 2008
3. Purchase and Installation of UPS (1 MW), needed in 2009 \$1,250k
 - RACF getting 300KW of UPS Power from NYBlue to cover 2008/9 needs
 - Received \$950k of supplemental funds from DOE/HEP
 - Rest will be paid out of ATLAS Program Funds
4. Power upgrades to allow full occupancy in 2010 – 2012 \$4,100k

While most of the infrastructure was/will be furnished by BNL we are further seeking for funds from DOE/HEP for 4.

Computing Issues – Funding

- Funding for computing remains a concern
 - Funding for computing, storage and network infrastructure less than planned / needed
 - Falling behind planned capacities – barely making single production pass per year
 - Request for supplemental funds for network core switch upgrade was not considered a priority this year
 - **The essential is “preserved”, internal traffic may not scale, no resilience / auto-failover**
- Funding issues pros & cons
 - Cons:
 - 1/3 replacement each year not entirely possible in recent years
 - Stretching lifetime sometimes leading to instability, taxing on personnel
 - Potential risk: Upgrade operation disruptive (bulk replacement / swap rather than 1/3 replacement, possible network reshapes in middle of data taking)
 - Squeezed user analysis (some moved off of BNL to remote sites)
 - Pros:
 - Squeeze on the storage side lead to inexpensive disk solution, distributed disk model
 - Approaches motivated by economics – implications on manpower non-trivial
 - Squeezed user analysis (some moved off of BNL to remote sites)
- RACF staffing concerns
 - Increased usage of inexpensive distributed disk is taxing on personnel
 - dCache, Xrootd needs personnel to maintain availability and required performance
 - Constant staffing level with growing data volume, capacities and complexity

Other Concerns

➤ Scalability

- Data volume has grown
- With growing data volume the Complexity has grown and will further grow with Detector and DAQ upgrades. Issues include
 - Increasing number of distributed services used by the Experiments
 - Grid Computing to transparently integrate usage of remote resources
- Detector R&D continues to draw manpower/expertise from S&C at STAR (simulation, tracking), team spread thin over increasing number of tasks
- Funding issues at remote institutions have an impact on BNL S&C (core) team
 - Projects are moving from institution-based support to a BNL-centric support
- Local effort remains about the same, but the expectations as well

Outlook

- **Plans to evolve and expand facility services to meet expected needs**
 - Are based on successful adjustments of technical directions
 - Remain within the mainstream of NP and HEP computing
 - Requires agreed and planned for increases (capital and operating) in 2009 and beyond
 - Drive down cost for operating and computing support
 - Improved monitoring and problem resolution
- **Funding shortfall until (at least) 2010 likely to require revision of the Mid-range plan (computing related)**
 - Resources external to BNL at collaborating Institutions and open for opportunistic usage at others via the Grid are vital to accomplish the scientific mission
 - Grid technology is likely to change future RHIC computing
 - Building on OSG Middleware and support
- **Physical infrastructure expansions and improvements**
 - Projects to expand by 8.500 sq.ft. in fall 2008 and summer 2009 are well underway
 - Funding for Power Infrastructure needed in 2010 - 2012 an issue

Backup Slides

RHIC Computing Facility (RCF)

- Organizationally established in 1997
- Staffed as a GROUP IN Physics Department
- Equipment physically located at Brookhaven Computing Facility
 - ⊖ BCF operated by ITD
- Currently co-located and co-operated with the ATLAS Computing Facility (ACF), the U.S. ATLAS Tier-1 Regional Center
 - ⊖ ACF ramping up quickly, currently
 - ACF capacities are ~ 65% for processing, 121% for disk capacity
 - ACF staff level ~ 75% of RCF

Principal RCF Services

- General Collaboration and User Support
- Processing Services (Linux Farm)
 - Programmatic Production processing
 - Individual and Group Analysis
- Online Storage (Disk)
 - Data storage for work area (Read / Write)
 - Data serving for Analysis (> 90% Read)
- Mass Storage (Robotic Tape System)
 - Raw Data recording and archiving
 - Derived Data Archiving
- Grid & Network Services

Scientific Computing Services (shared)

Shared services at a high level are

- Grid Services
 - STAR and U.S. ATLAS are using the OSG Middleware and services (compute and storage element, accounting, site availability/reliability monitoring, etc.)
- Storage and Data Caching services
 - ATLAS, PHENIX and STAR are using dCache and XrootD as data serving technology
- Data Processing services
 - Processor Farm and Local Resource Management (Condor)
- Networking (Site and WAN)
- Facility itself (power, cooling, monitoring, planning)
- Central core services – such as Web, Backup, Storage Area Network, Database Administration, and more

Director RACF

Michael Ernst

Deputy Director (RCF)**Admin. Assistant**

Maureen Anderson

**Processing and
General Services**

Tony Chan

Storage

Shigeki Misawa

**Grid MW and
Services**

Dantong Yu

Linux Farms + SysAdmin

- Tony Chan
- Chris Hollowell
- Richard Hogue
- Alexander Withers
- Tristan Ziska

Computer Fabric + SysAdmin

- Robert Petkus
- Mizuki Karasawa
- John McCarthy
- Jason Smith
- Morris Strongson
- James Pryor
- (Frank Burstein)

Operations & Infrastructure

- Richard Hogue
- Kevin Casella
- Enrique Garcia

Mass Storage

- Shigeki Misawa
- Ognian Novakov
- John Riordan
- Grace Tsai
- David Yu
- New Hire (MSS H/W Ops)

Storage Mgmt &
Data Movement

- New Hire
- Hironori Ito
- Jane Liu
- Ofer Rind
- Iris Wu
- New Hire (dCache Ops)

Central Storage

- Maurice Askinazi
- Dave Free

Production, Data Base
and User Support

- Dantong Yu
- John DeStefano
- Carlos Gamboa
- Yuri Smirnov
- Tomasz Wlodek

General Software Env. and
Software Development

- Dimitrios Katramatos
- John Hover
- Jay Packard

Grid Middleware
(OSG / WLCG)

- John Hover
- Jay Packard
- Xin Zhao

Experiment / RCF Interaction

- **Weekly Liaison Meeting**
 - ⊖ Addressing operations issues
 - ⊖ Review recent performance and problems
 - ⊖ Plan for scheduled interventions
- **Experiments / RCF Annual Series of Meetings to develop Capital Spending Plan**
 - ⊖ Estimate scale of need for current/coming run
 - ⊖ Details of distribution of equipment to be procured
 - ⊖ Most recent in early Spring for FY-07 funds
- **Periodic Topical Meetings, examples**
 - ⊖ ~Annual Linux Farm OS upgrade planning
 - ⊖ Replacement of Central Disk Storage
- **Other User Interactions**
 - ⊖ Web site
 - ⊖ Ticket System (Request Tracker (RT – Open Source)
 - Fully
 - ~3000 Tickets for RHIC & ATLAS Services (last 12 months)

Strategies and best practices for RHIC Computing

- Computing facilities and the tools and software that make them run cannot stagnate
- Drivers that mandate a strategy of continuous refresh of computing facilities
- Space, power and cooling infrastructure is very expensive
 - Computers and disks more than 3 years old use up these resources in a very inefficient & costly way
 - Maintenance costs for old equipment is high
 - Where possible trade maintenance costs against capital expenditures
 - Risks of running equipment that is past end-of-life (vendor doesn't support) are high and operations costs (people) are high

Managing Scale and Complexity

- This is a huge enterprise to run and monitor
 - We collect enormous amounts of data on usage, servers, services – e.g. automated monitoring and alerts
 - But still not enough automated problem solving and intelligence
- Has to be sub-divided for scaling issues
 - e.g. one “head node” can only manage so many CPUs
 - One file server can only efficiently serve so many CPUs and so much data
 - One robot can only handle so many tapes and drives
- Yet has to be “virtualized” and run efficiently across all these boundaries
 - There is still work to do in the context of both RHIC and ATLAS

Curation of Data

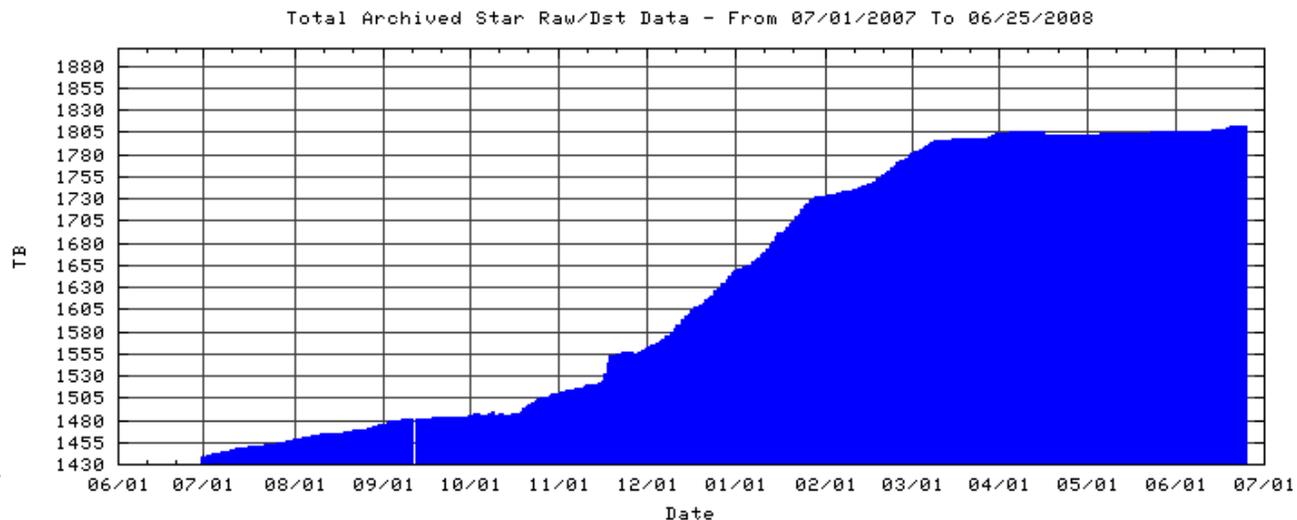
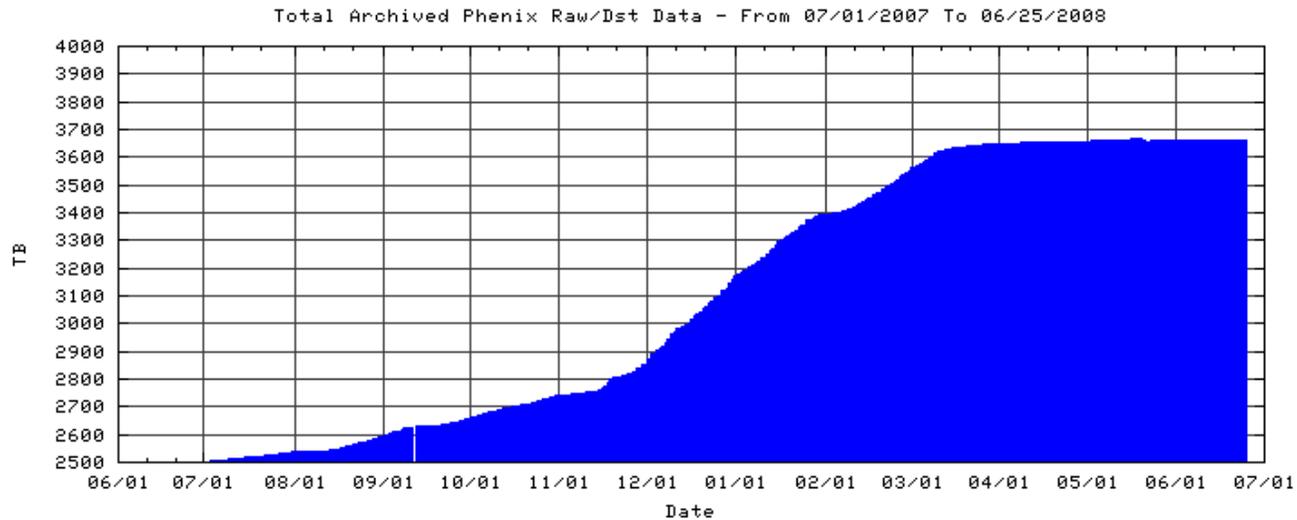
The activity of managing the use of data from its point of creation to ensure it is available for discovery and re-use in the future (also includes managing vast amounts of data sets for daily use)

- Ensure integrity – must keep data “live” in robot
- Protect from loss - maintain data across 2 buildings (from 8/2009 on)
 - Replication of some datasets
- Robots come in discrete units with tape drives attached to them
 - Have some “pass-through” capabilities with adjacent robots
 - Needs care managing data placement and access patterns
- Tape Drives for continuous repack and migration of data will be needed
- Tape drive plant need to grow continuously
 - Most of the use is for serving data for analysis
 - Complex tradeoff between tape and disk – continue to refine this model and measure usage – and optimize
 - Replenishment of tape technology on a ~3 year cycle

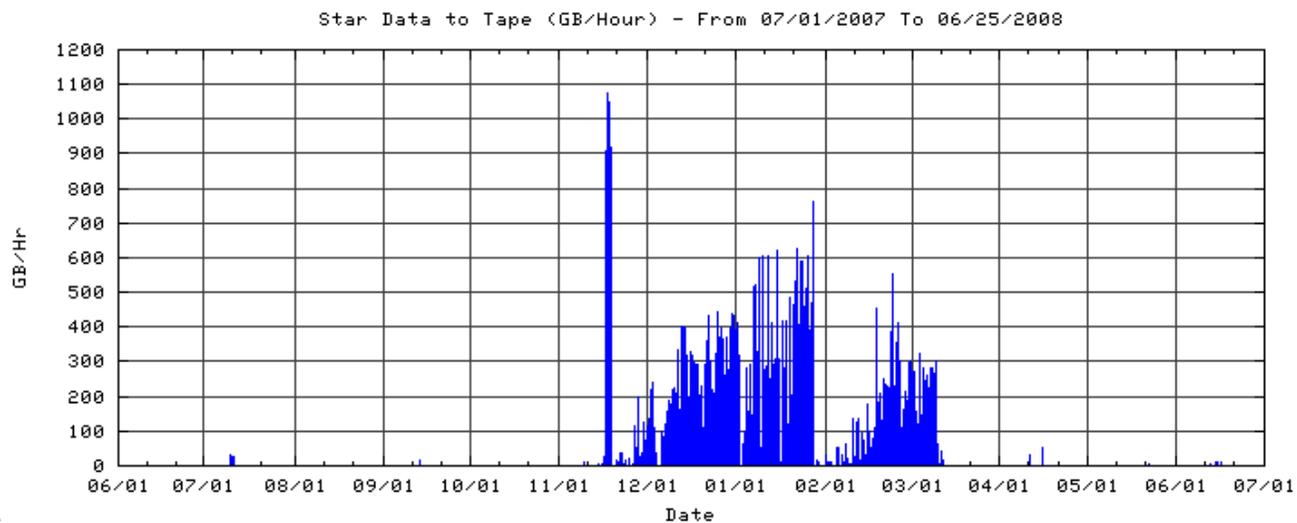
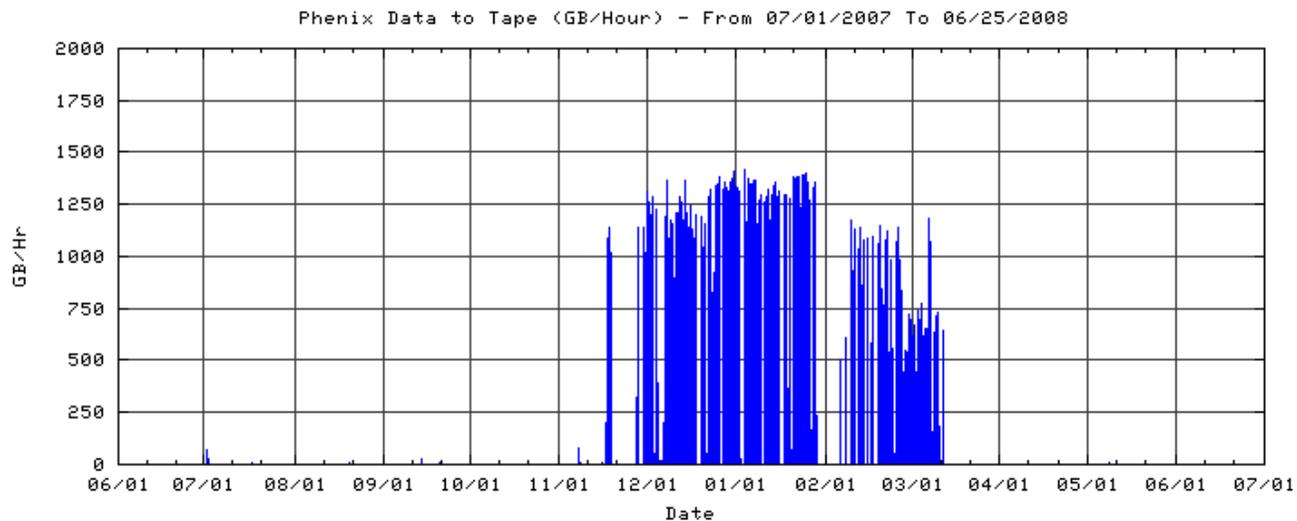
HPSS Statistics

- # RHIC Files – 32,961,900
- Total Space used by RHIC – 6.2 PB
- Cartridges Used
 - ⊖ 7125 9940B Cartridges
 - ⊖ 12519 LTO-3 Cartridges

Total Archived DST/Raw Data



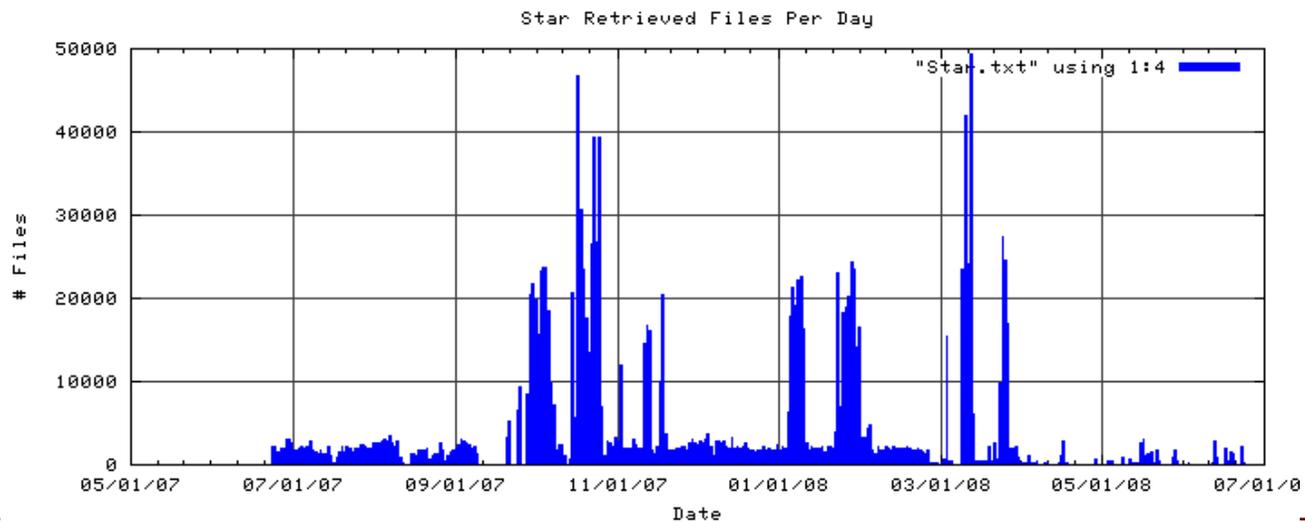
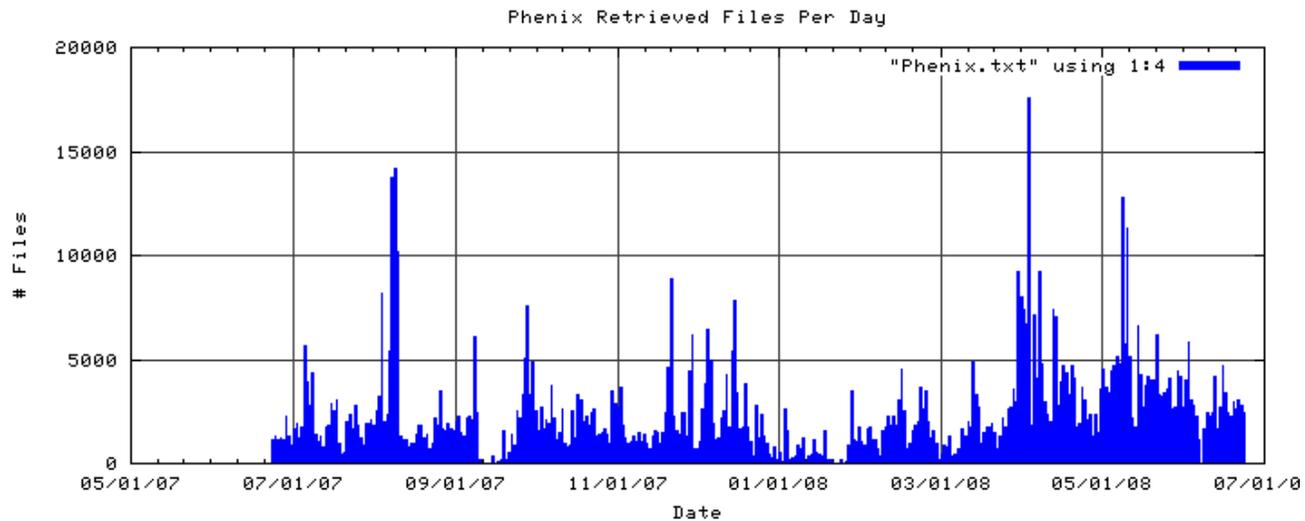
Raw Data Rate to Tape



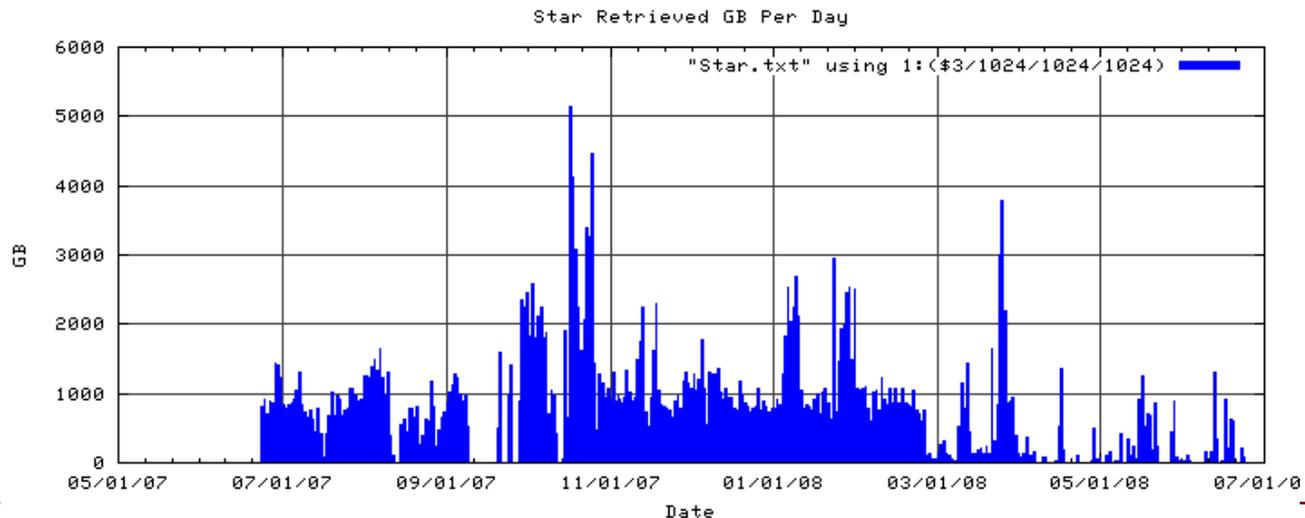
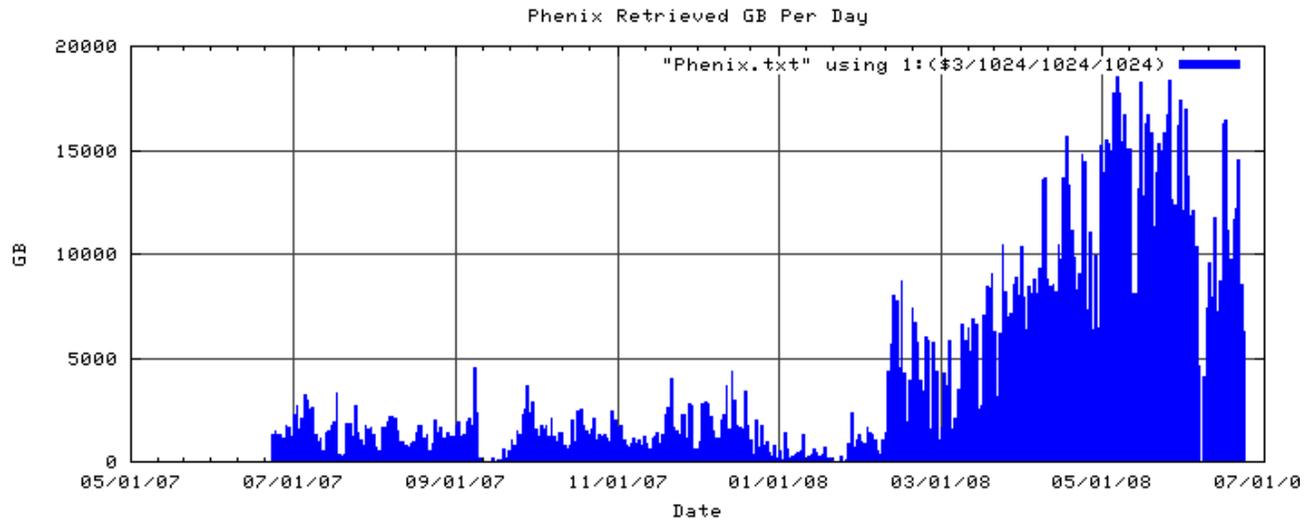
Data Retrieval Statistics

- Total Data Read in 1 year - 1.88 PB
- # Files read in 1 year – 2,500,975
- Avg. size of retrieved file – 788MB
- Avg. # files staged - 6853/day (5 files per minute)
- Avg. data staged – 5276 GB/day
- Avg. mount rate – 945 mounts/day

Files Retrieved



Retrieved Volume per Day



Milestones

- Major reconfiguration of hardware in HPSS.
 - ⊖ Redistribution of cartridges in silos and of silos/tape drives to experiments, triggered by addition/removal of silos and addition of LTO-4 tape drives.
 - ⊖ Physical move of all STAR movers to ease future expansion.
 - ⊖ Redistribution/addition of mover hardware
 - ⊖ Upgrade of all AIX based hardware

More Mass Storage System Milestones

- Significant number of new recorded metrics and logs
- Closed many “feedback loops” to automatically adjust HPSS configuration based on system condition and load profiles.
- Moved data on 8000 9940B tapes to LTO-3 at no cost for media.

HPSS Equipment overview

- New Core server
- Monitoring/gateway servers (6)
- 18 Movers (50/50 Star/Phenix split)
- 5 Fibre Channel Disk Arrays (~32 TB) 50/50 Star/Phenix split
- 1GE channel bonded network inter-mover network

Mass Storage Equipment (cont'd)

- 3 SUN/STK SL8500 silos 18,933/28,536 slots used (shared with Atlas)
- 3 9310 Silos 11,000/17,188 slots used (shared with Atlas)
- 26 LTO-3, 10 LTO-4, 35 9950B drives (Star LTO-3/Phenix LTO-3/4)

Mass Storage Equipment Changes

- 10 new LTO-4 drives (Sept 2007)
- 10 new LTO-3 drives (June 2007)
- 1 old 9310 traded in (had 4 now have 3)
- 1 new SL8500 (had 2 now have 3)
- 2 additional Linux movers
- Eliminated remaining AIX based movers
- Retired 9940A tape drives.

Mass Storage Equipment Lifecycle

- Core/Monitoring/Gateways 1 year old
- Movers mix of 1/2/3 years old
- Disk arrays (4 and 3 years old)
- All 9310 silos end of service life (EOL) 2010
- 9940B SAN infrastructure EOL 2009

Areas of Concern

- Retirement of residual 9940B data problematic
 - ⊖ Copy time – small files on remaining cartridges
 - ⊖ Remaining 9940B cartridges may require sanitization before exchange.

- Future retirement of LTO-X not likely to create commercial interest
 - ⊖ No residual value for LTO-X tapes unlike 9940B

- EOL of 9310 silos problematic

Areas of Concern

➤ Tape loss

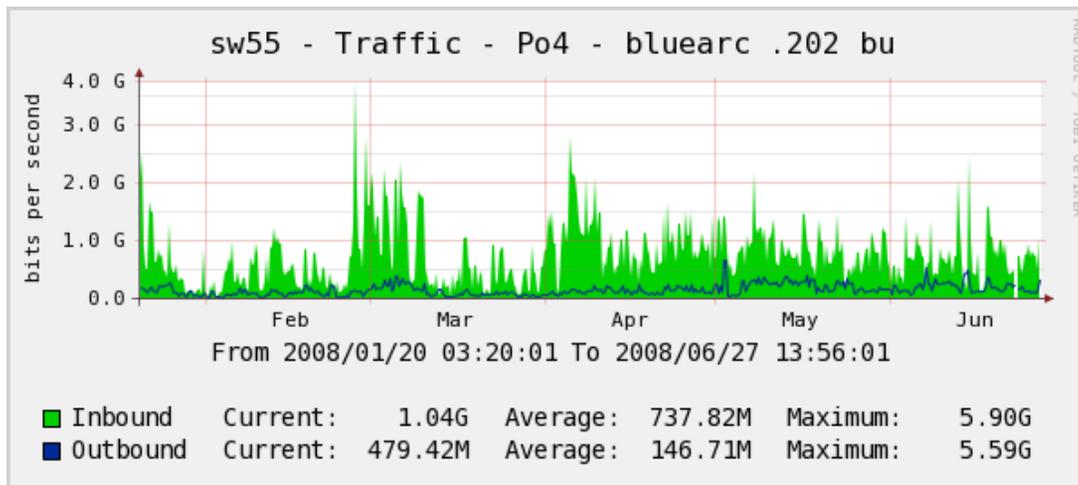
⊖ ~50 9940B since start of FY 2000.

⊖ ~10 LTO-3 since start of FY2006.

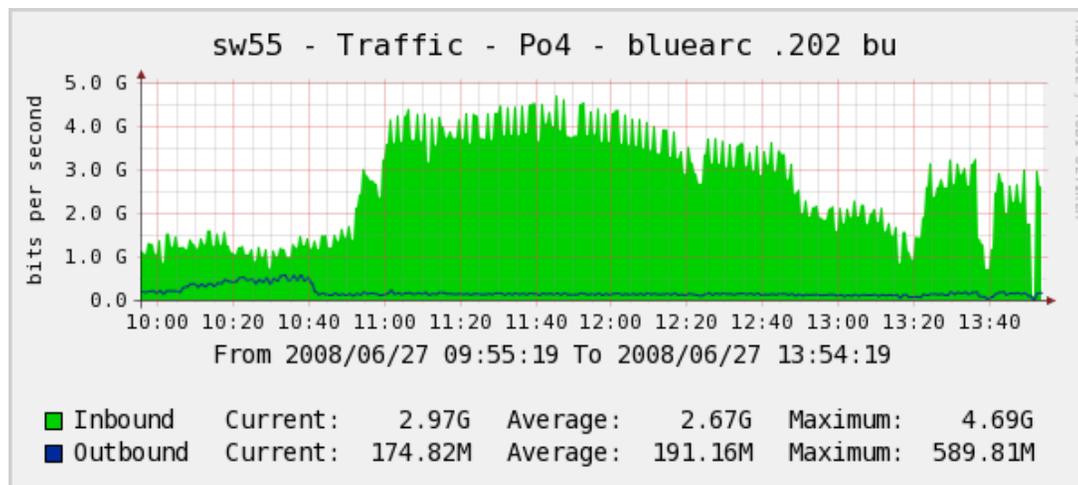
➤ Problems with library service caused by reorganization of vendor support organization. Hope to have this resolved soon.

BlueArc Network I/O

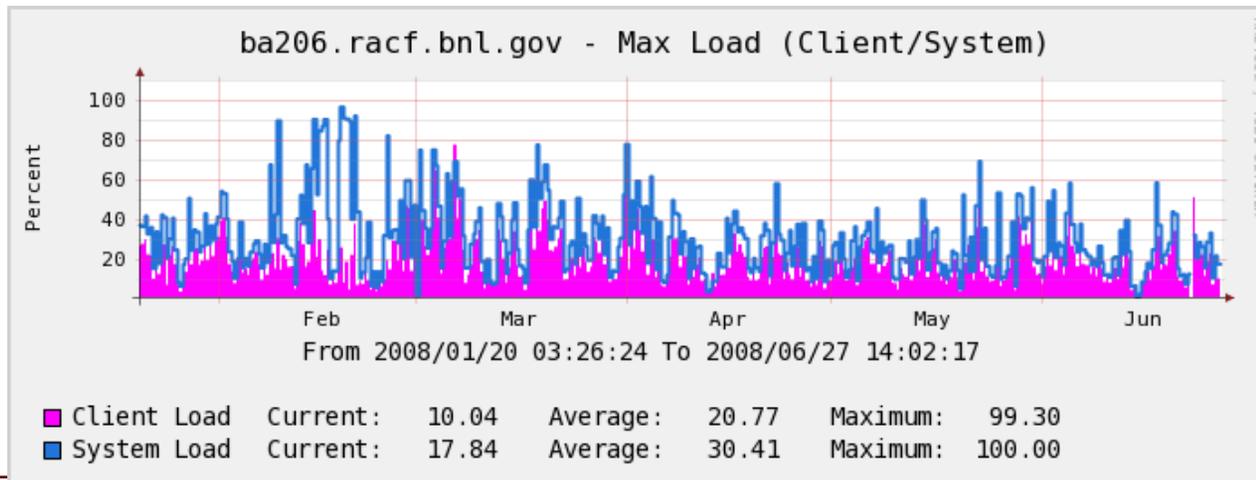
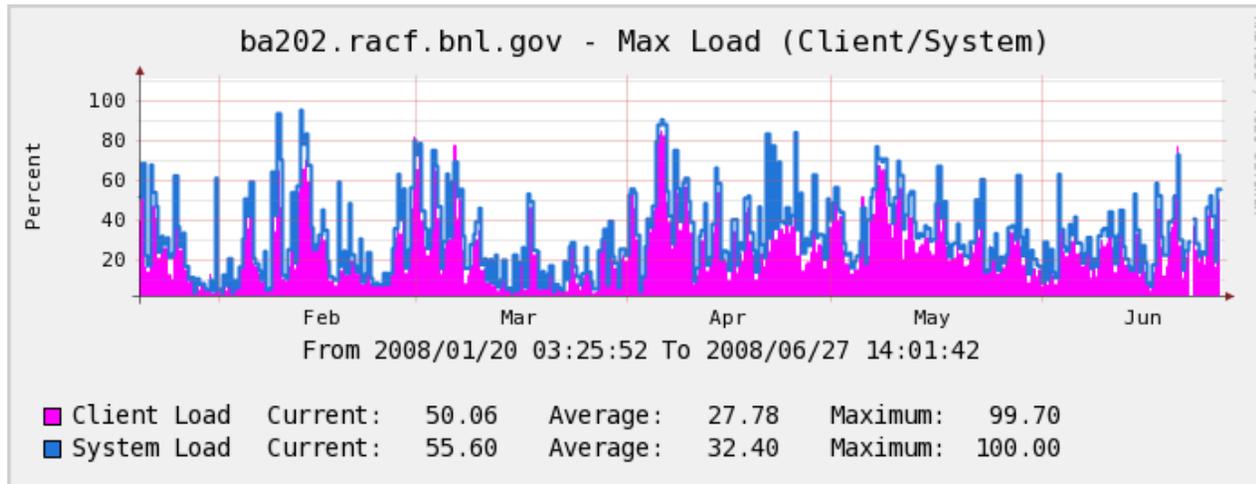
Last 6 months



A typical day



BlueArc Server Load



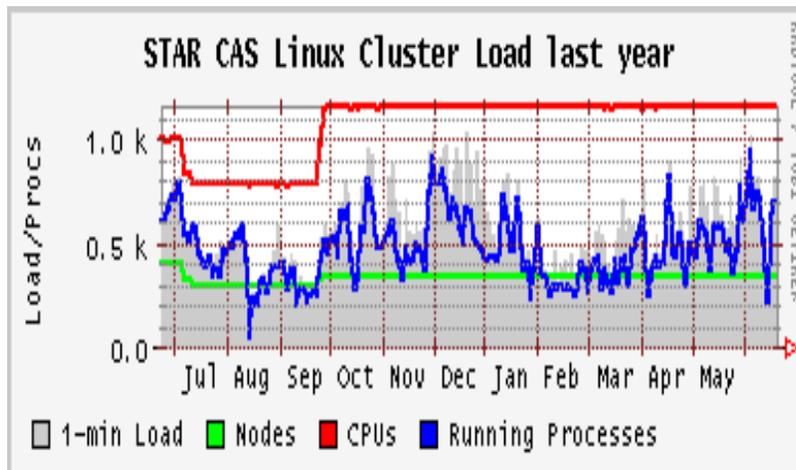
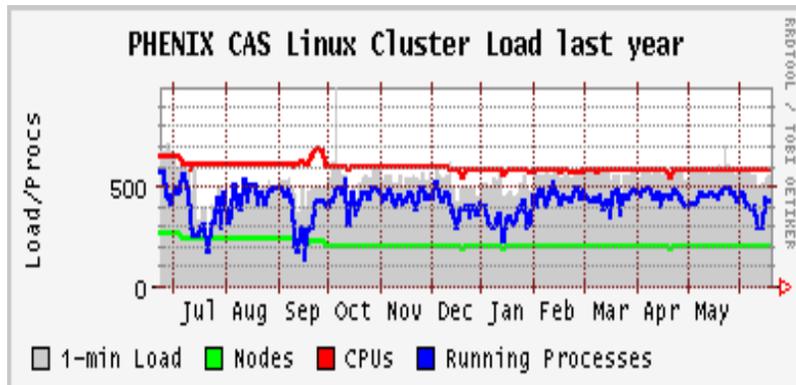
Compute Servers

➤ Three Generations of Linux CPU rack mount systems

- Dual CPU (single core) systems (2,800 SI2k per box)
- Dual CPU (dual core) systems (4,600 SI2k – 10,000 SI2k per box)
- Dual CPU (quad core) systems (16k – 22k SI2k per box)
- Currently 1,200 compute servers with 2,400 CPU's (3800 cores, 3.3 MSI2k)
- Delivery expected by end July of ~120 additional Dual CPU / Quad Core machines (8 cores / box) with 2.7 MSI2k
 - Multi-core CPU technology also addresses power/cooling barrier by finessing non-linearity of power consumption with clock speed
- Expect to address future requirements by continuing to follow Moore's Law price/performance in commodity market (multi-core, 64 bit advances)

PHENIX & STAR Processing Utilization

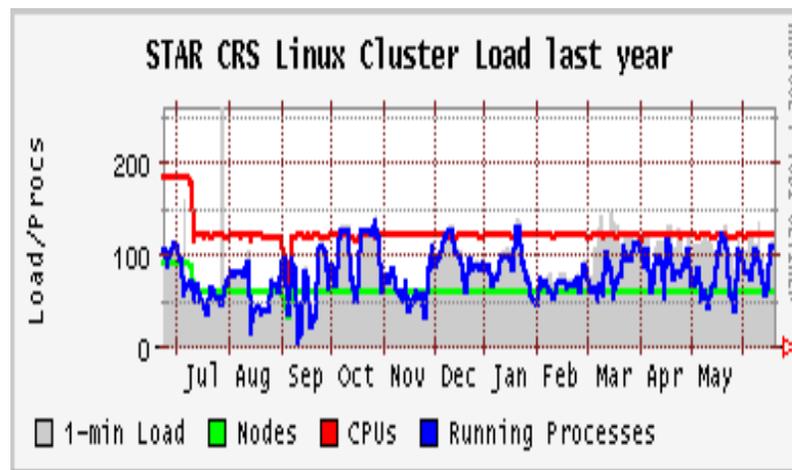
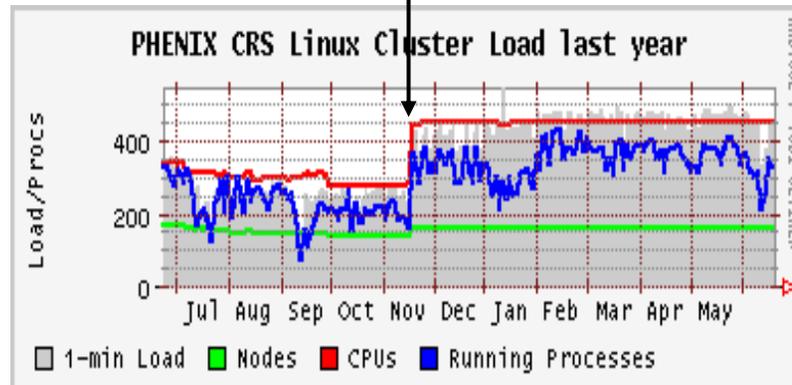
Analysis



Reconstruction

Additional CPU's (256 cores)

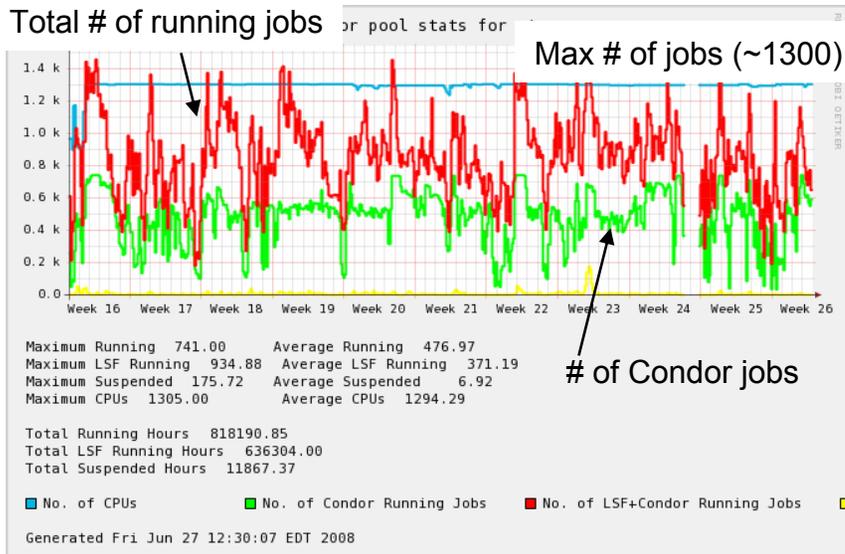
From supplemental funds



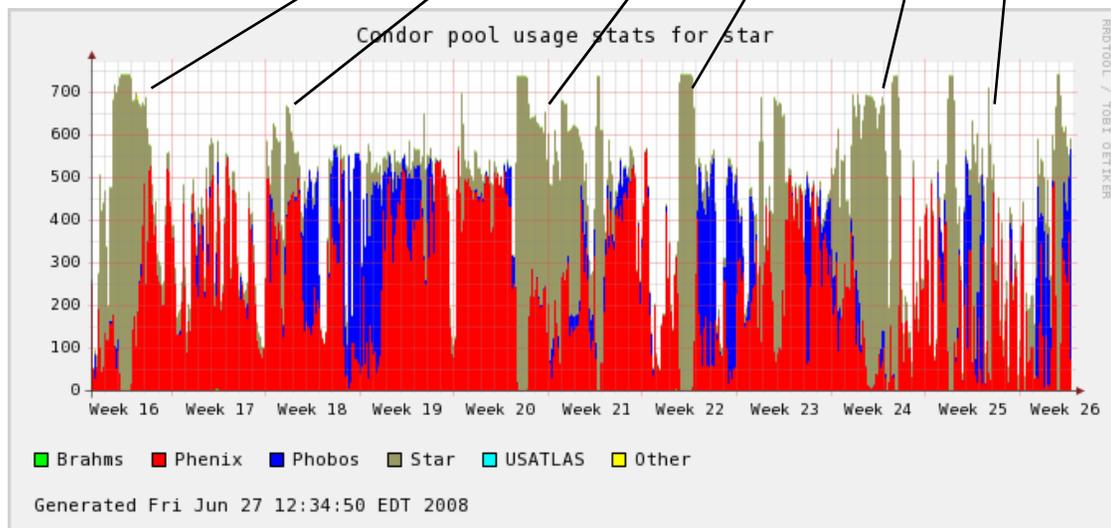
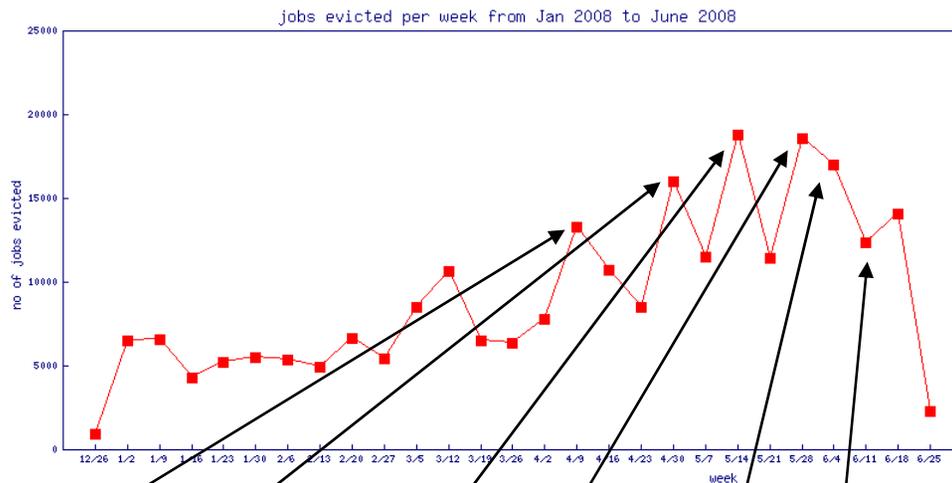
Resource Sharing among Experiments

- Goal was to make idle cycles available in processor farms to other user communities without impact to “owner”
- Mechanism is to evict “guest” jobs when “owner” needs cycles
 - Two hour grace period to let the job complete

STAR Batch Queue Statistics



Number of evicted jobs submitted to general queue



Condor Usage in the last 12 months

File Edit View Go Bookmarks Tools Help

https://web.racf.bnl.gov/Facility/LinuxFarm/gq_year3.html

Red Hat, Inc. Red Hat Network Support Shop Products Training

Statistics from 06-26-2007 to 06-25-2008

		no. jobs completed						
		destination						
		phenix	phobos	star	brahms	atlas	rcf	total
source	phenix	<u>3311363</u>	<u>839451</u>	<u>716132</u>	<u>1358314</u>	<u>283284</u>	<u>61427</u>	6569971
	phobos	<u>40079</u>	<u>1871981</u>	<u>97694</u>	<u>63647</u>	<u>55490</u>	<u>366</u>	2129257
	star	<u>110</u>	<u>461</u>	<u>852802</u>	<u>4061</u>	<u>34814</u>	<u>140</u>	892388
	brahms	<u>664</u>	<u>88</u>	<u>116</u>	<u>104934</u>	<u>3</u>	<u>2013</u>	107818
	atlas	<u>74083</u>	<u>75333</u>	<u>24477</u>	<u>63392</u>	<u>10418042</u>	<u>55107</u>	10710434

		no. jobs evicted before completion						
		destination						
		phenix	phobos	star	brahms	atlas	rcf	total
source	phenix	<u>313709</u>	<u>35630</u>	<u>56489</u>	<u>64400</u>	<u>24491</u>	<u>8193</u>	502912
	phobos	<u>6201</u>	<u>27647</u>	<u>6064</u>	<u>3475</u>	<u>2783</u>	<u>19</u>	46189
	star	<u>67</u>	<u>13</u>	<u>18153</u>	<u>2090</u>	<u>2806</u>	<u>9</u>	23138
	brahms	<u>65</u>	<u>4</u>	<u>63</u>	<u>1769</u>		<u>51</u>	1952
	atlas	<u>2423</u>	<u>5536</u>	<u>5223</u>	<u>3585</u>	<u>186377</u>	<u>191</u>	203335

		total effective runtime hours consumed by completed jobs						
		destination						
		phenix	phobos	star	brahms	atlas	rcf	total
source	phenix	8712683.01	494542.01	286828.54	545754.76	334890.74	76352.97	10451052.03
	phobos	22600.88	6471016.96	75023.67	75133.08	19074.01	189.68	6663038.28
	star	124.99	459.51	3568350.45	35979.34	183686.49	199.29	3788800.07
	brahms	173.32	43.29	41.2	57892.55	3.26	438.97	58592.59
	atlas	7318.09	228072.25	13424.89	36906.61	11376085.68	8857.55	11670665.07

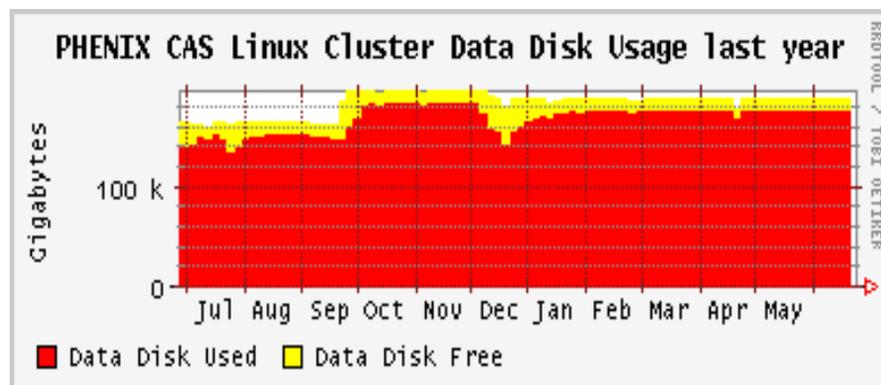
		total ineffective runtime hours consumed (including jobs removed)						
		destination						
		phenix	phobos	star	brahms	atlas	rcf	total
source	phenix	628384.16	137427.66	240391.52	296880.18	147514.09	81389.16	1531986.77
	phobos	70456.15	368297.31	14704.08	24588.15	20807.91		498853.6
	star	1997.01	375.76	234159.16	39218.49	103859.22	11.97	379621.61
	brahms	44.37	4.16	33.02	6540.12		12.04	6633.71
	atlas	18750.15	13867.23	13182.63	20743.52	329677.22	22.62	396243.37

Done

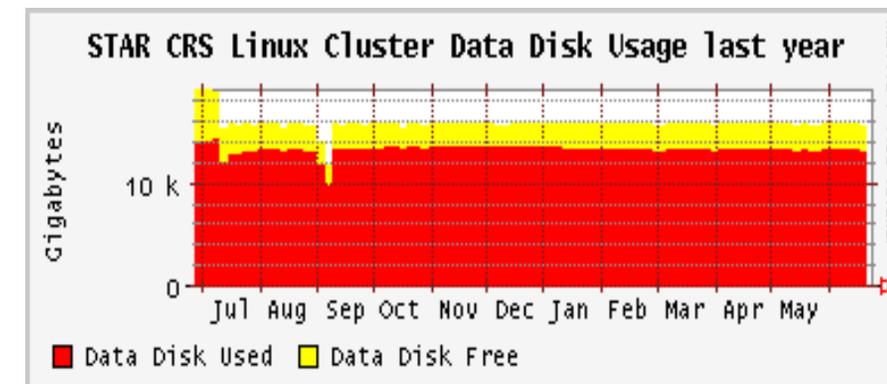
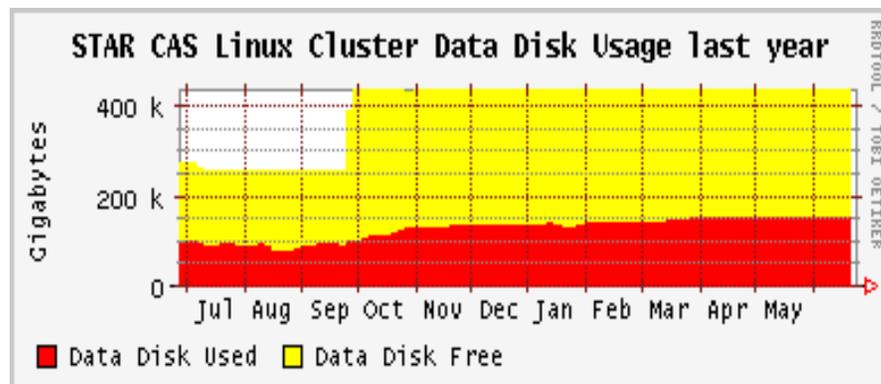
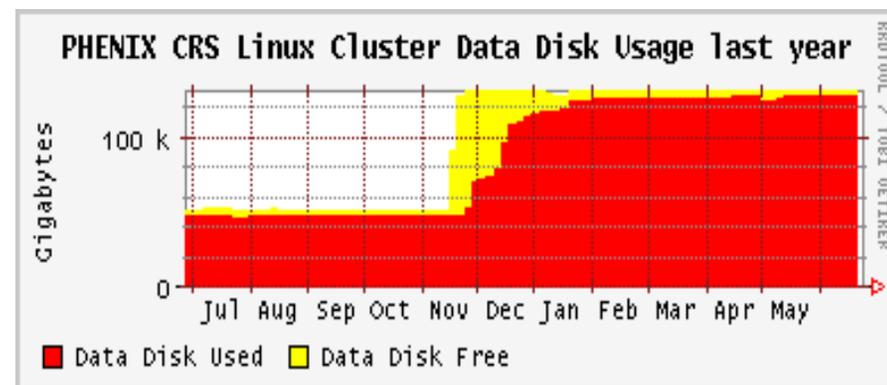
- Condor usage by RHIC experiments increased by 50% (in terms of number of jobs) and by 41% (in terms of CPU time) since 2007.
- PHENIX executed ~50% of their jobs in the general queue.
- General queue jobs amounted to 37% of all RHIC Condor jobs during this period.
- General queue efficiency increased from 87% to 94% since 2007.

PHENIX & STAR Distributed Disk

Analysis



Reconstruction



U.S. ATLAS Computing Capacities

➤ The U.S. ATLAS Tier-1 center at BNL will complete its four year procurement ramp-up for initial ATLAS/LHC operation in the fall of 2008

BNL Tier-1	CPU T1	5.4MSI2k	Tier-1 Processing Nodes
	Disk T1	3 PB	dCache
Fall 2008	Network	20Gb/s	CERN ↔ BNL
	People	20FTE	Integration and Operation

- In 2011 there will be
 - 18 MSI2k just for the pledged Tier 1 computing and
 - 16 PB of Tier-1 disk

A New Operational Model for the RACF

- RHIC facility operations is a system-based approach
- ATLAS needs support for (mostly) remote users
- Service-based operational approach better suited for a distributed computing environment
- New SLA for RACF incorporates service-based approach
- Mapping of services to related systems

A Dependency Matrix

File Edit View Insert Format Tools Data Window Help

Arial 10

M13

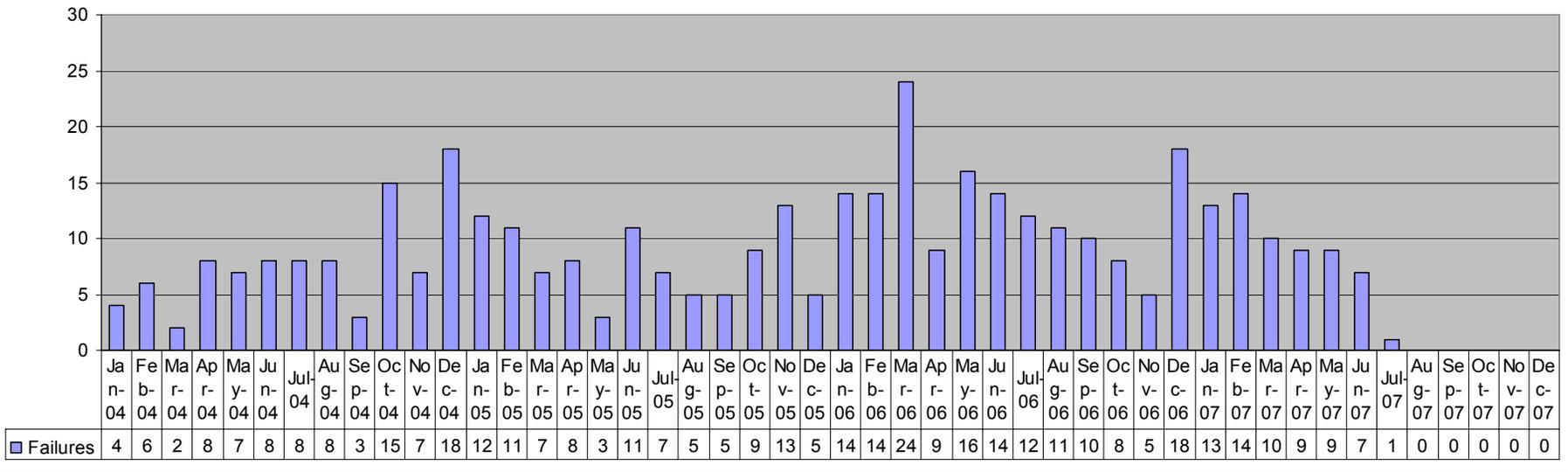
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Service Class	C	C	C	C	U	C	C	C	C	C	C	C	M	M	U	U	U	M	C
2		Access to mass data archive - RMC	Access to mass data archive - ATLAS	Access to local NFS storage	Access to NFS storage	SMB File Servicing (SMBFA)	Data Catalogs	User Databases	Grid job execution	Local job execution	PANDA Service	Central Reconstruction Service	Gateways	Web documents	Email	Printing	Data Protection	Accounting	Software Subversion Service	Support and Monitoring
3	HPSS	X	X									X								
4	dCache		X																	
5	DG2		X																	
6	FTS		X																	
7	PANDA										X									
8	Gatekeepers								X											
9	Gatekeepers								X	X										
10	Farm								X	X			X							
11	Condor/LSP								X	X			X							
12	User Databases						X	X												
13	NFS			X																
14	AFS				X															
15	Backup																X			
16	Kerberos	X	X	X	X				X	X										
17	MyProxy		X	X					X	X										
18	VOMS		X	X					X	X										
19	GUMS		X	X					X	X										
20	DNS/LDAP/NTP	X	X	X	X	X	X	X	X	X			X	X	X					X
21	SDI		X						X	X										
22	Gratia																	X		
23	RT																			
24	Nagios													X						X
25	Ganglia													X						X
26	Web svr													X	X					
27	Network and Firewalls	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
28	Power	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
29	A/C	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
30																				
31																				
32	Critical	C																		
33	Managed	M																		
34	Unmanaged	U																		
35																				

IME Status

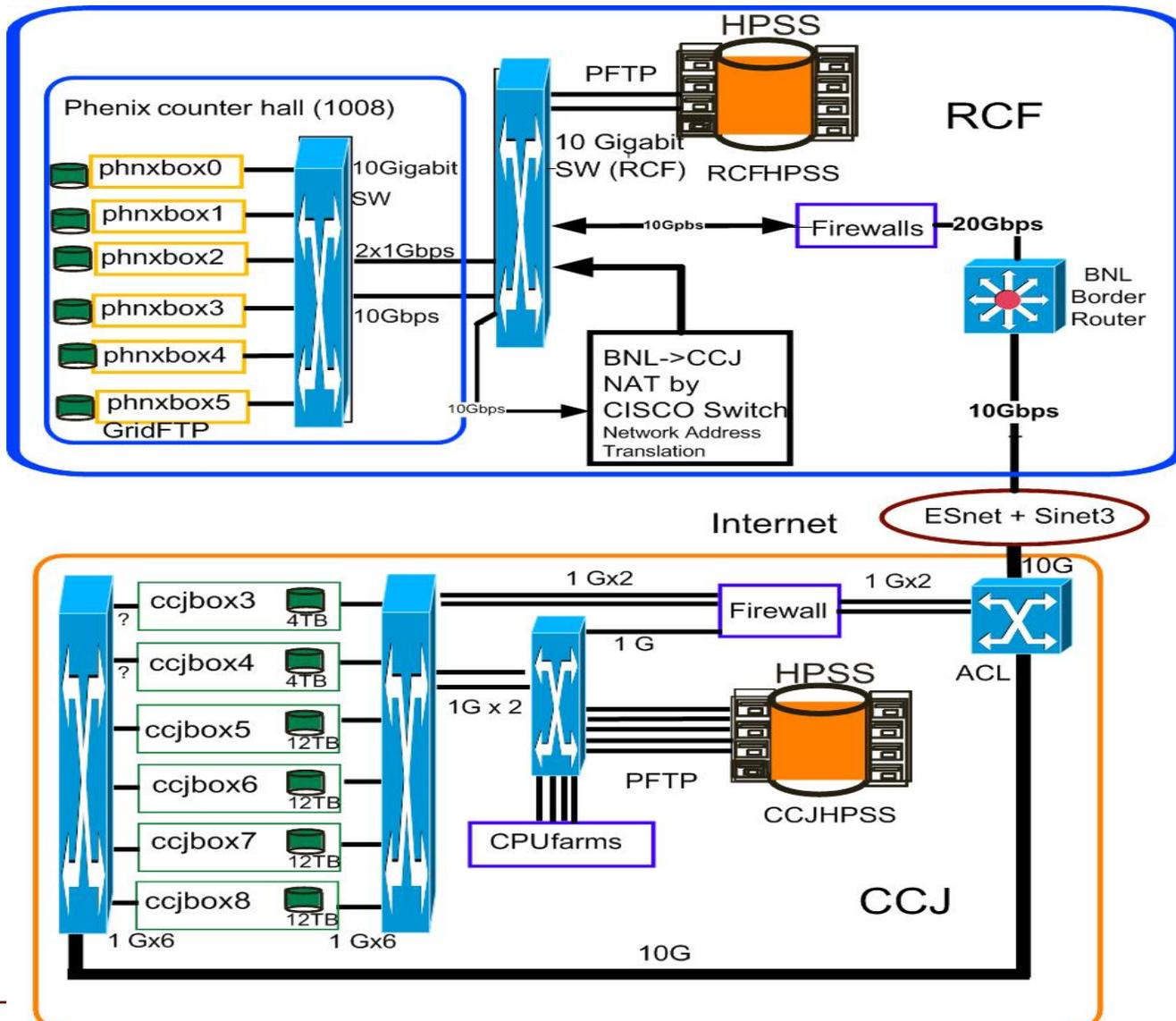
Sheet 1 / 3 tony@voyager:/home/tony

Central Disk Failures over Time

GCE Failures 1/04 - 12/07



PHENIX Data Transfer Infrastructure



Adding Space in 2008 and 2009

