

---

# Run II Computing

Amber Boehnlein

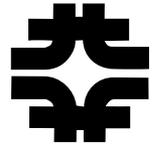
Fermilab

March 29, 2005

---

# Challenges in Run II Computing

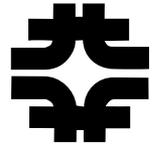
---



- Compared to Run I
  - data rates for Run II experiments have increased 20-30 times
  - Collaborations have doubled
  - the physics applications are slower
  - Reliance on COTS based systems
  - permanent storage is robotic
  - user expectations are higher.
- Staffing levels are comparable to Run I, and the computing is better meeting the experiments' needs.
- Operational support supplied by CD and experiment collaborators

# 2004-2005 Achievements

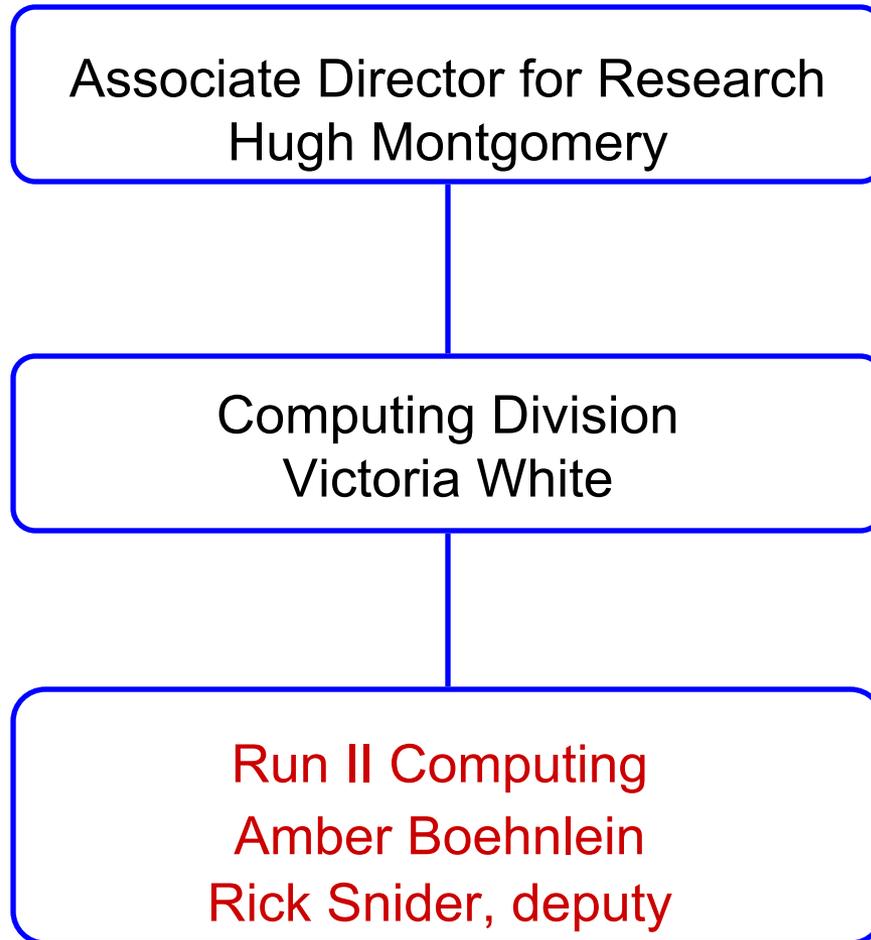
---



- Operations are smooth for both experiments
  - Key components in production for 4-5 years
    - Sequential Access via Meta data (SAM), dCache, Central Analysis Facility (CAF),
  - Joint operations department formed from CDF and DO CD departments
  - Combining pager rotations, expanding use of automated tools.
- Second-generation deployments
  - Deployment of database servers for CDF
  - New farm scripts for CDF
  - SAM deployment for CDF central systems
  - Completion of calibration DB access in RECO for DO
  - Monte Carlo production for DO using automated submission tools
  - Global Reprocessing for DO
    - 2003-100M events reprocessed offsite
    - 2005-Goal of 800M events reprocessed offsite
  - Hardware—replacing aging infrastructure components such as legacy SGI

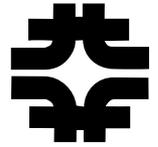
# Line Management

---



# Run II Department Roles

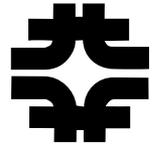
---



- Experiment specific support
- Production
- Data handling
- System administration
- CDF Online

# Production and Offline Support

---



14 FTEs in the Run II Department plus 1.5 FTE for database development and 0.5 FTE for DO Reconstruction Task force (16)

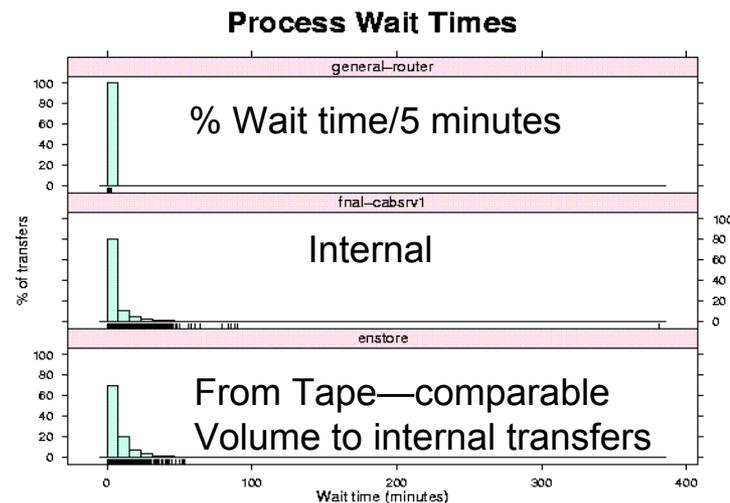
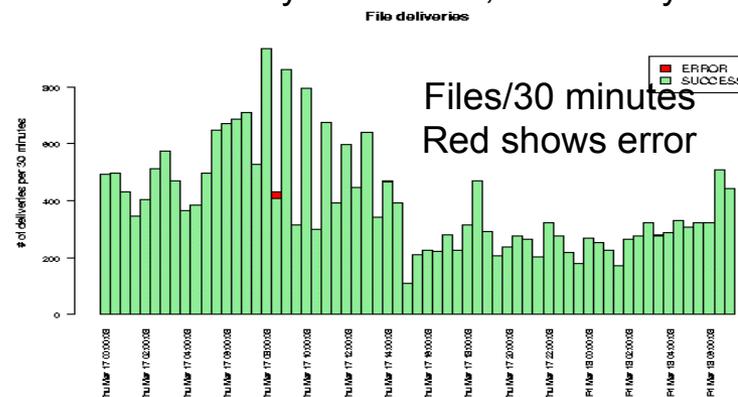
- Experiment specific tasks
  - Experiment Management (operations, physics, computing, software)
  - Offline Code development and releases
  - Experiment specific database
  - Preparing and Running Production executables
  - Includes Guest Scientists and Visitors needed to leverage experiment expertise
  - Physics Analysis

# Data Handling/Production



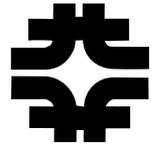
- 15M-25M Events logged per week/experiment
- Production capacity sized to keep up with data logging.
- MC production at remote sites
- Tape writes/reads
  - CDF 14 TB/ 20TB
  - DO 7 TB/30 TB
- Analysis requests
  - 750 M events/experiment analyzed
  - CDF: 150 TB/week
  - DO : 50 TB/week in 1000 requests

One of the DO analysis stations, recent day



# Data Handling Operations Effort

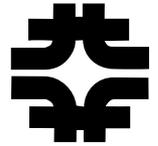
---



7 FTEs in the Run II Department plus 2 FTE direct support from other depts + 2 hires (11). This effort has been reduced by 2 FTEs in the past year

- Ongoing development to improve the services to improve maintainability and robustness and longevity
  - Increased reliance on Grid efforts
    - Improved monitoring for users and experts
- Daily operations for both experiments for SAM and dCache

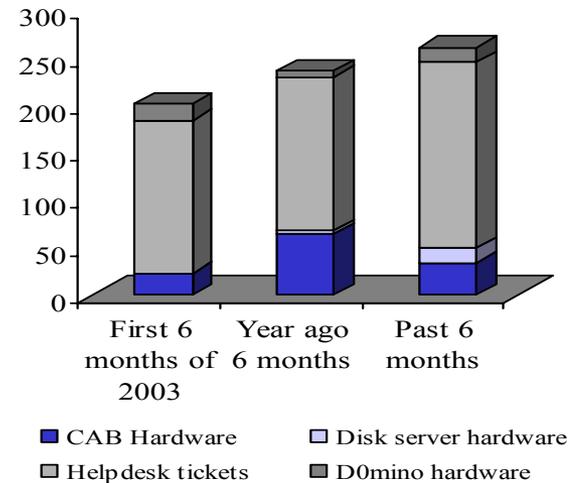
# System Administration/Online



9 FTEs + 3 hires (12 FTEs)

- 24/7 operations for critical systems
- Sizable operational plant
  - 1400 worker nodes
  - 200 file servers
- Introducing and perfecting automation
- CDF desktop support
- CDF online became a CD responsibility in FY2005, work combine operations with DO online—2 positions transferred from PPD
- Have been running short-staffed

Remedy tickets



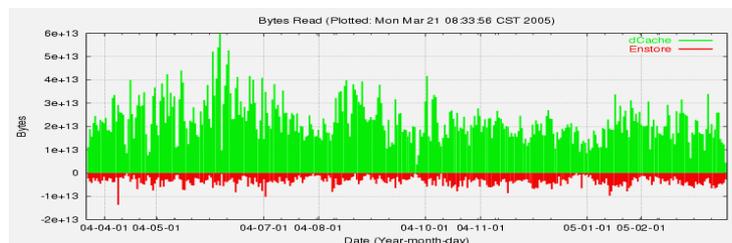
Using Remedy system  
Tickets/hardware/year  
Tracking in this way helps  
Us to understand which  
And how to mitigate  
operational issues

# CD Central Support



- Discussed in the breakout session
- Provides operational support

- Database systems
- Farms
- Hardware evaluations
- Networking
- Robotic storage
- Facilities
- General services: Equipment pool, e-mail, linux support, contract support, customer support



DCache and Enstore Reads/day for CDF for the past year

- Refining systems and evaluating hardware and scaling issues for all consumers and streamlining operations.
- CD evaluates and provides common tools to allow for uniform maintenance and operation of large systems.
- CD provides services that allow experiments to use common solutions as they move towards global and grid computing

# Experiment Contributions

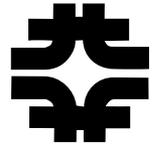
---



- Personnel ~ 20 FTEs/experiment
  - Experiment side operations, development and management
    - Database and database servers
    - Experiment specific analysis and processing infrastructure
    - Experiment resource estimates and the allocation of resources to meet experiment needs
    - Remote site administration operation
    - Running farm and MC production
    - Desktop computing
- Equipment
  - Resources provided by collaborations
    - Remote facilities for production and analysis
    - Equipment sent to FNAL for central facilities
- These are crucial contributions that can not be supplied by the FNAL CD.

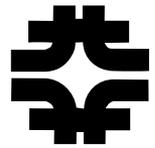
# Run II Computing Reviews

---



- Yearly Director's Review  
<<http://cdinternal.fnal.gov/CDEvents.asp#RUNII>>
  - Project progress, goals, needs are presented and reviewed for current and projected for out years
  - Bottoms up estimate
- Guidance in FY2002 review was \$2M/year/experiment, \$1.5 in FY2004
- CD developing economic model to better estimate costs for facilities

# CDF and DO Need Projections for FNAL Equipment



## CDF

FY	CAF CPU (SM)	Inter. CPU (SM)	Fam CPU (SM)	DB (SM)	Tape Drives (SM)	Disk (SM)	Network (SM)	Misc (SM)	Total (SM)
03A	0.31	0.08	0.19	0.15	0.20	0.34	0.23	0.02	1.5
04A	0.49	0.06	0.24	0.07	0.13	0.14	0.19	0.07	1.4
05E	0.42	0.10	0.18	0.05	0.43	0.29	0.31	0.05	1.8
06E	0.85	0.10	-	0.03	0.51	0.27	0.12	0.05	1.9
07E	0.73	0.10	0.18	0.03	0.48	0.17	0.08	0.05	1.8

## DO

	Purchased 2003	Purchased 2004	Purchase 2005	Purchase 2006	Purchase 2007	Purchase 2008
FNAL Analysis CPU	\$470,000	\$277,000	\$417,132	\$534,926	\$406,376	\$350,311
FNAL Reconstruction	\$200,000	\$370,000	\$454,269	\$717,742	\$443,490	\$362,546
File Servers/disk	\$111,000	\$350,000	\$357,000	\$356,000	\$293,000	\$276,000
Mass Storage	\$280,000	\$254,700	\$40,000	\$600,000	\$300,000	\$100,000
Infrastructure	\$244,000	\$140,000	\$547,000	\$200,000	\$200,000	\$200,000
FNAL Total	\$1,305,000	\$1,391,700	\$1,815,402	\$2,408,667	\$1,642,867	\$1,288,856

# Budget



SWF:

Corresponds to 37 FTEs for computing

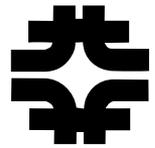
		<u>FY04 ACTUAL</u>	<u>FY05</u>	<u>FY06 PBR</u>	<u>FY07 FLAT</u>	<u>FY08 FLAT</u>	<u>FY09 FLAT</u>
		<u>BASE</u>	<u>BUDGET</u>		<u>TO PBR</u>		
1.1	<u>Accelerators</u>	0.0	0.0	0.0	0.0	0.0	0.0
1.2	<u>Collider Experimental Program</u>	18,383.4	18,673.6	17,526.5	17,541.2	17,532.8	17,693.0
1.2.1	CDF	6,769.7	6,789.2	6,089.3	6,109.8	6,087.2	6,131.5
1.2.1.1	CDF Operations	5,582.2	5,934.0	5,871.2	6,082.9	6,087.2	6,131.5
1.2.1.4	CDF Run II	1,187.5	855.2	218.1	26.9	0.0	0.0
1.2.2	DZero	7,885.5	8,372.5	7,652.8	7,501.9	7,476.8	7,527.1
1.2.2.1	Dzero Operations	6,652.9	7,206.8	7,257.8	7,475.3	7,476.8	7,527.1
1.2.2.4	Dzero Run II	1,232.5	1,165.7	395.0	26.6	0.0	0.0
1.2.3	Run II Computing	3,544.3	3,330.9	3,736.1	3,929.5	3,968.8	4,034.4

M&S

		<u>FY04 ACTUAL</u>	<u>FY05</u>	<u>FY06 PBR</u>	<u>FY07 FLAT</u>	<u>FY08 FLAT</u>	<u>FY09 FLAT</u>
		<u>BASE</u>	<u>BUDGET</u>		<u>TO PBR</u>		
1.1	<u>Accelerators</u>	0.0	0.0	0.0	0.0	0.0	0.0
1.2	<u>Collider Experimental Program</u>	9,189.3	7,928.5	6,812.2	6,782.2	5,316.4	5,351.6
1.2.1	CDF	2,347.8	1,838.2	1,746.1	1,731.1	1,486.6	1,492.3
1.2.1.1	CDF Operations	1,370.1	1,778.2	1,746.1	1,731.1	1,486.6	1,492.3
1.2.1.4	CDF Run II	977.7	60.0	0.0	0.0	0.0	0.0
1.2.2	DZero	3,245.2	2,625.0	1,600.8	1,585.8	1,335.8	1,335.8
1.2.2.1	Dzero Operations	1,719.1	1,576.0	1,600.8	1,585.8	1,335.8	1,335.8
1.2.2.4	Dzero Run II	1,526.1	1,049.0	0.0	0.0	0.0	0.0
1.2.3	Run II Computing	3,596.2	3,465.3	3,465.3	3,465.3	2,494.0	2,523.5

# Budget

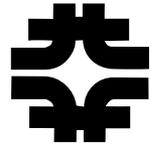
---



- 39 FTE of direct support (-2 as ongoing projects end)
- Approximately 36 FTEs direct support through 2009
  - Responsibilities likely to increase with constant staff
- **Equipment**
  - 2004—Contributions to Grid Computing Center as well as standard equipment budget
  - Making \$1.5M budget cover \$1.8M in needs requires experiment choices
  - Use CDF/DO/CMS/General resources to form FermiGrid
- **Operating--\$150K/year/experiment**
  - primary source of budget for tape
  - COTS equipment requires additional operating funds and personnel relative to SGIs that are being retired.
- **Maintenance**
  - Have largely moved off the large SGIs
  - Robotics and Database machines

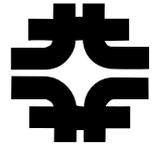
# Risks

---



- Increased calls on FNAL CD as migration of effort and equipment to LHC
- Declining equipment and operations budgets require choices.
- Scaling with data sample size might have unanticipated consequences
- Operational performance of new hardware elements, Moore's Law deviations, experiment code
- Longevity of hardware components and software applications

# DO Projection and History



DO	2003			2004	
	Projected	Projected(2)	Purchased	Projected	Purchased
FNAL Analysis CPU	\$505,400	\$500,000	\$470,400	\$339,000	\$277,000
FNAL Reconstruction	\$200,000	\$40,000	\$200,000	\$83,000	\$370,000
File Servers/disk	\$262,000	\$200,000	\$111,000	\$490,000	\$350,000
Mass Storage	\$460,000	\$285,000	\$280,000	\$230,000	\$254,700
Infrastructure	\$640,000	\$500,000	\$244,000	\$290,000	\$140,000
FNAL Total	\$2,067,400	\$1,525,000	\$1,305,400	\$1,432,000	\$1,391,700

Reconstruction costs underestimated-delayed deployment of adequate disk.

A data handling system that enables use of seamless offsite resources AND prestages data from tape AND robotic storage that out-performs expectation AND network capacity has enabled current budgets to provide sufficient computing for DO

# Summary

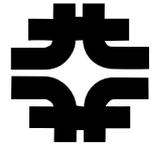
---



- Computing for Run II is performing well to meet the experiment needs
- Experiment and CD effort needed to cover the spectrum of tasks.
- Conscious effort towards streamlining operations.
- Living within the limited budgets with increasingly painful choices and increased risk.

# RUN II Department Roles

---



- Operations—Running the systems, standing pager rotations/shifts, researching latest technologies
  - purchasing and deploying equipment
  - tracking down and fixing problems
  - code management
- Development—exploring use cases, writing code, introducing new features, testing, documenting, exploring technologies
- Integration—testing, more testing, training users, transition from development to operations
- Planning—how best to use resources to meet stakeholder needs, facility issues
- Interfacing – Serve in experiment management roles, bridging the CD and the experiments, CD department to CD department, hosting guest scientists
- Participate in physics analysis as collaboration members -- 30% of department FTEs hold scientific positions

# Risks, expanded

---



- Increased calls on FNAL CD as migration of effort and equipment to LHC
- Declining equipment and operations budgets are already limiting the data collection rate.
  - Over time, limits in the equipment and operating budget will create delays
- Operational performance of user code
  - DO reconstruction code performance and release turn-around
  - CDF user code has caused inefficiencies on the CAF
- COTS Computing
  - Experiments need best price/performance, which introduces risk.
  - Moore's law
  - Have a good process in place for evaluation, purchase and acceptance.
  - Each purchase of worker nodes presents challenges
    - FNAL CD plays engineering/integrator role by default
  - Commodity file servers are maintenance intensive

# Risks, expanded

---



- **Data Handling**
  - SAM system, dCache, hardware working well
  - User patterns are still evolving, sometimes conflicts between wanting to get results out and using standard production.
  - Scaling with data sample size might have unanticipated consequences.
  - Count on next generation tape drives to mitigate tape costs
- **Longevity of hardware components and software applications**
  - Starting to use a 4 year replacement cycle for worker nodes so the equipment is off warranty the final year.
  - 5 year life cycle on major components, replacement needed again around 2010 when budget for Run II will be extremely limited.
  - Migrating either experiment from existing mode of operation or user interfaces would be time intensive and costly.