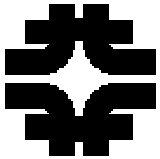
Fermilab Grid Computing – CDF, D0 and more..

Victoria A. White
Head, Computing Division, Fermilab
white@fnal.gov



First - Apologies

1) For not being in Korea with you
   2 DOE reviews in the US this week
2) For my substitute speaker, Don Petravick, having to cancel his trip at the last moment
3) For a poorly prepared, last-minute, talk as a result

Fermilab Ongoing Scientific Program, 2005

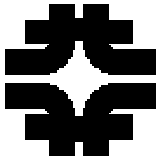
- CMS
- International Linear Collider
- LHC
- MiniBoone
- MINOS
- Pierre Auger, CDMS
- Run II (CDF, D0, Accelerator and potential upgrades)
- Simulation for Accelerators
- Sloan Digital Sky Survey
- Test Beams and MIPP
- Theory Programs and Lattice QCD computational program
- R&D for SNAP
- R&D for DES, Nova, Flare, Minerva
- R&D for Linear Collider detectors

* Grid related

Inclusive Worldwide Collaboration is essential for Particle Physics

What are we doing at Fermilab to help this ?

- Networks and Network research
- Grids - for CMS, for Run II, SDSS, Lattice QCD
  - Leadership role in Open Science Grid
  - Lots of technical work with LCG
- Guest Scientist program
- Education and Outreach program
- Experiment sociology and leadership – changes
- Videoconferencing
- Virtual Control Rooms
- Physics Analysis Center for CMS
- Public Relations



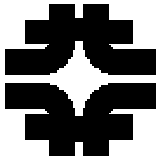
Grid Projects and Working groups

Fermilab is involved in numerous Grid projects and coordination bodies

- PPDG, GriPhyN and iVDGL - the Trillium US Grid projects
- LHC Computing Grid bodies and working groups on Security, Operations Center, Grid Deployment, etc.
- SRM – storage systems standards
- Global Grid Forum – Physics Research Area
- Global Grid Forum – Security Research Area
- **Open Science Grid**
- **Interoperability of Grids and Operations of Grids**
- Ultralight Network research
- Ultranet/Lambda-Station Network and Storage project
- And probably some I forgot…..

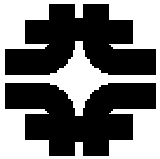
Fermilab Grid Strategy

- Strategy (from a Fermi-centric view)
  - Common approaches and technologies across our entire portfolio of experiments and projects
  - FermiGrid – so we can share all Fermilab resources
- Strategy (from a global HEP view)
  - Work towards common standards and interoperability
- At Fermilab we are working aggressively to put all of our Computational and Storage Fabric on "the Grid" and to contribute constructively to interoperability of various Grid infrastructures



CDF and D0 (and MINOS)

- Running experiments with working Monte Carlo, Data Processing and Analysis systems, handling tens of TBs of data per day
- CDF and D0 both now have global computing models
- All 3 experiments use SAM for Data Handling - metadata, reliable file transfer, data delivery, bookkeeping
- D0 use SAM-GRID (SAM + JIM workload manager) and have successfully reprocessed data offsite
- CDF use CAFs and dCAFs and SAM and will be taking the next steps towards Grid job submission



SAM-Grid

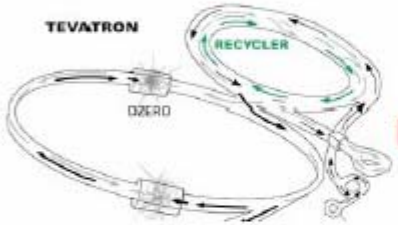
- SAM-Grid is fully functional distributed computing infrastructure in use by D0, CDF and MINOS
- ~30 SAM stations worldwide active for D0
- ~20 SAM stations worldwide active for CDF
- D0 successfully carried out reprocessing of data at 6 sites worldwide
  - http://www.fnal.gov/pub/ferminews/ferminews04-02-01/p1.html FermiNews article February
  - And in the latest reprocessing has already handled more events off-site
- SAM-GRID uses many common (and evolving!) Grid standards, will evolve further, and will soon run routinely on top of LCG computing environment and Open Science Grid

Computing Model

Remote Farms

Central Farms

Raw Data
RECO Data
RECO MC
User Data



Data Handling Services

Central Storage

Remote Analysis Systems

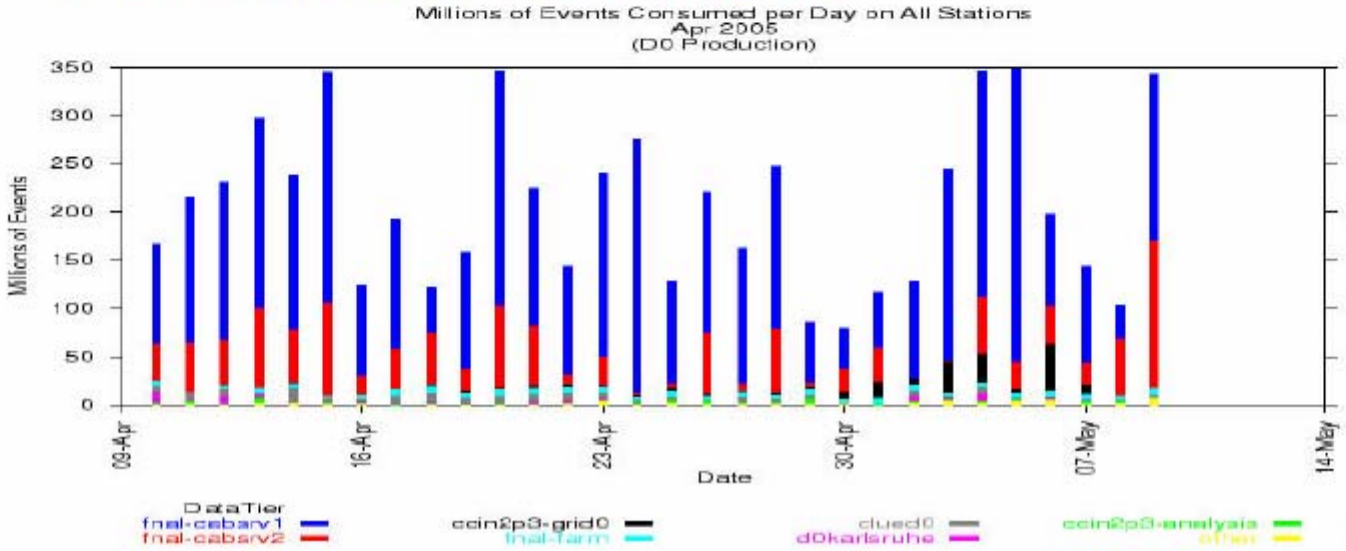
Central Analysis Systems

User Desktops

Amber Boehnlein, FNAL

SAM Data Handling

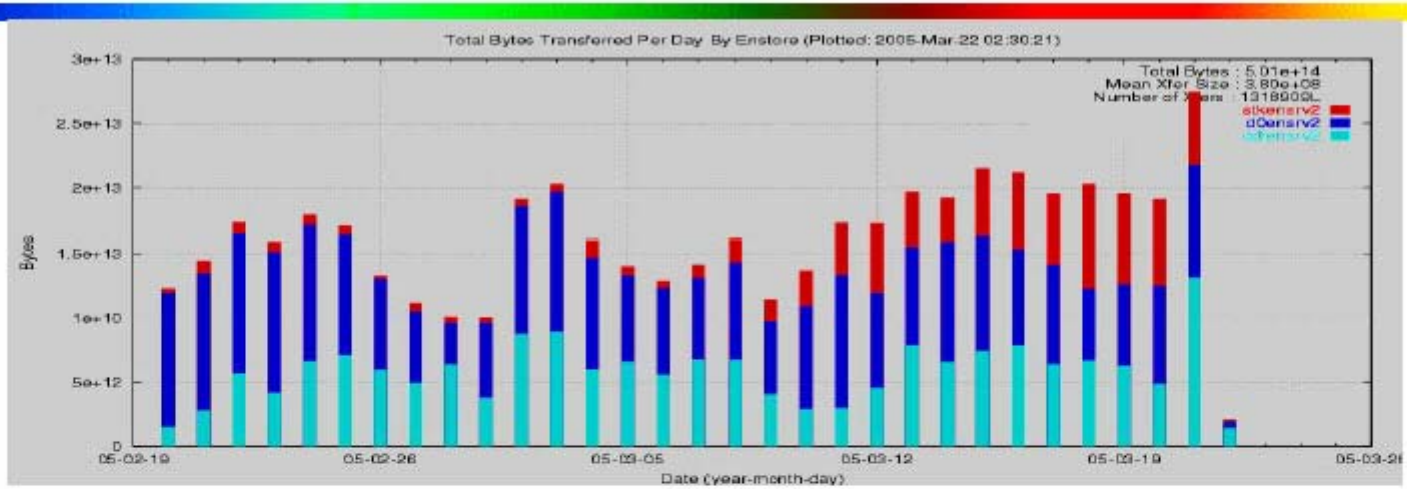
- 15M-25M Events logged per week
- Production capacity sized to keep up with data logging.
- Tape writes/reads
  - 7TB/week average writes
  - 30 TB/week reads
- Analysis requests at FNAL
  - 750 -1100 M events/week
  - 50 TB/week in 1000 requests



Boehnlein, FNAL

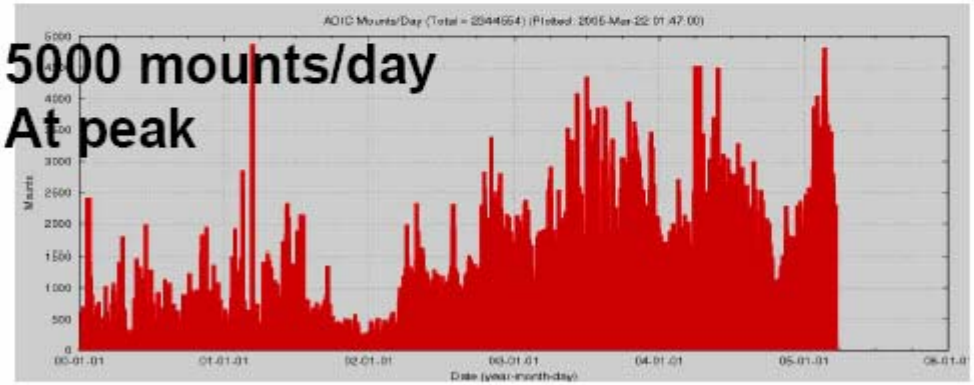
Central Robotics

30TB
At peak



Daily Enstore traffic for CDF, DO, and other users

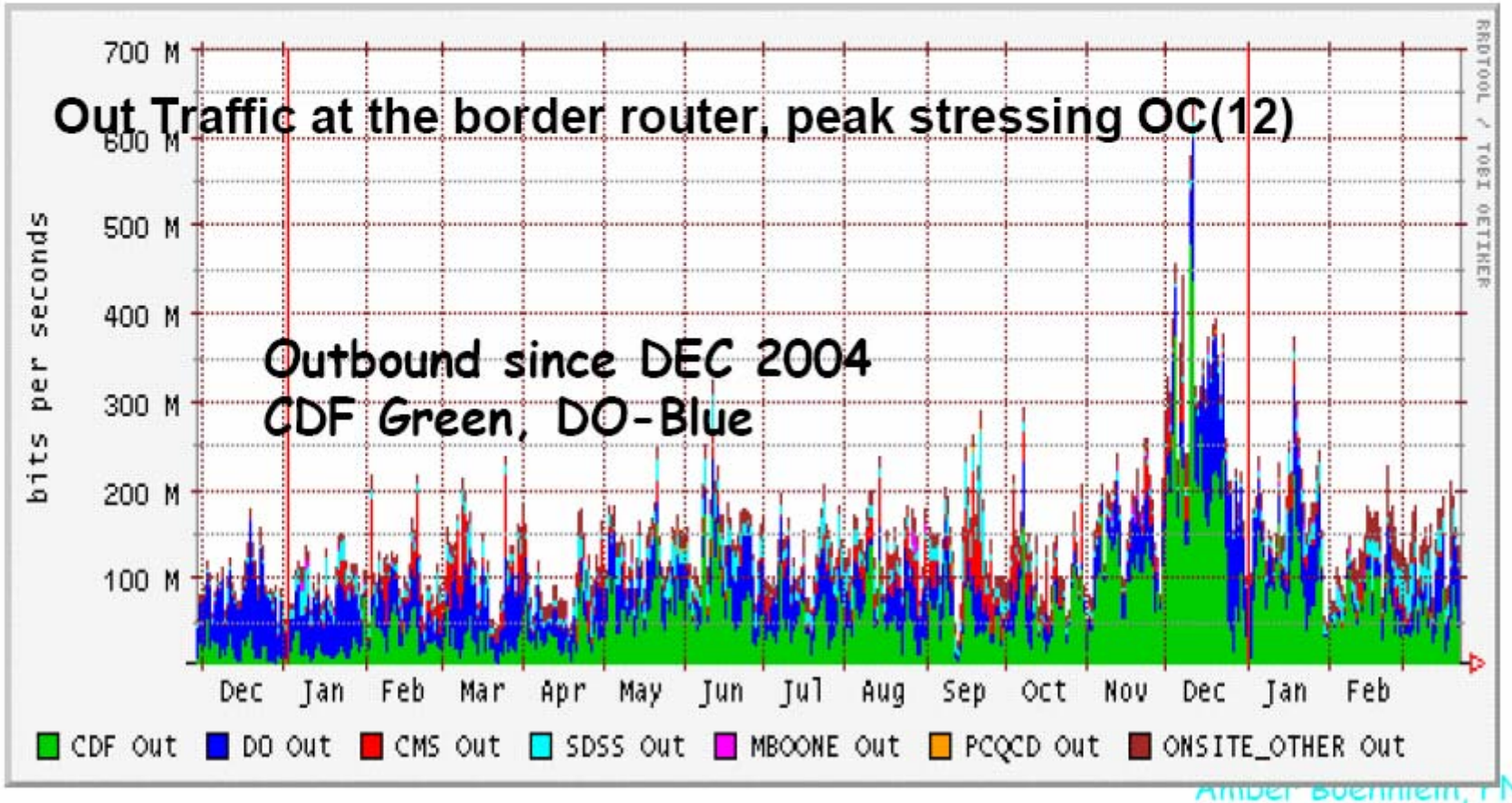
DO 9940	638 TB
DO LTOI	175 TB
DO LTOII	159 TB
Total	~1pb



Diversity of robotics/drives maintains flexibility

Wide Area Networking

- OC(12) to ESNET, filling production link, anticipate upgrade
- R&D: Fiber link to Starlight-used to support reprocessing for WestGrid



Monte Carlo Production

- Standardization reduces effort
  - Mc_runjob for work flow
  - Mcfarm enables small farms to appear as unified system
  - SAMGrid provides unified job submission
- In the past year, DO produced 160M Monte Carlo Events at 10 different sites
  - Southern Analysis Region (SAR)
    - SPRACE (Sao Paulo),
    - TATA
    - UTA, OUHEP, LTU, LUHEP
  - IN2P3 (Lyon), Nikhef, Prague, GridKA

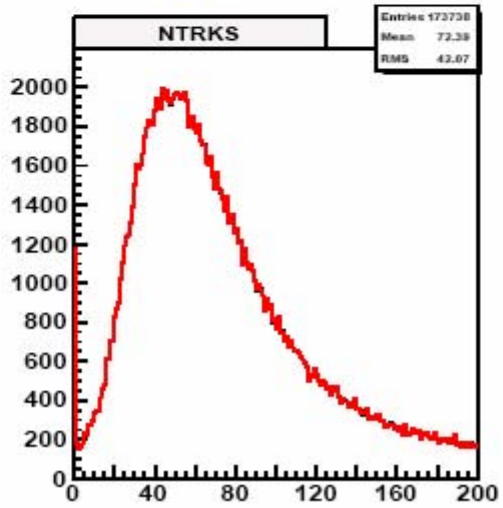
Reprocessing

	2003	2005
Luminosity	100 pb^{-1}	470 pb^{-1}
Events	0.5G	1G
Raw data	-	250 TB
DSTs (150kB/event)	45 TB	-
TMB (70 kB/event)	6 TB	70 TB
Time 50s/event (on standardized Computing unit)	2000 CPU for 3 months	3400 CPUs for 6 months
Remote processing	20 %	100 %

Deployment Effort ~ 3 dedicated people for a year, + site providers
<not including the SAMGrid development effort>

Site Certification

- Each center processes agreed datasets (100 files) for certification
- Unmerged and merged TMBs are compared
- Common set of events is compared between sites.
- Database changes caused some certification delay—good news-procedure worked

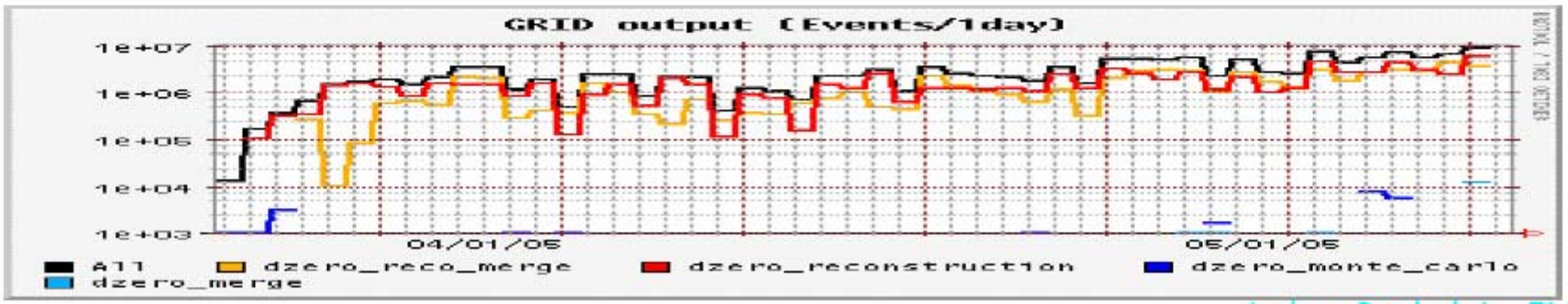


Reprocessing Resources

WestGrid	600 cpus		running
IN2P3	400 cpus		running
Wisc	30 cpus		certified
Prague	200 cpus		running
SAR(UTA)	230 cpus		certified
GridKA	500 cpus		certifying
CMS Farm	100 cpus	OSG	Certifying*
UK(4 sites)	750 cpus		1 running 1 certifying
External	~2800 cpus		~40% running
Target	3400 cpus		
FNAL Farm	1200 cpus		Data collection

Outlook

- Reprocessing
  - Working on efficiency
  - Adding robustness to operational scripts.
  - Install more sites-Sprace & Oscer running test jobs
  - T2-HEPGRID Brazil can close the gap!
  - Will need p17 MC!
- SAMGrid
  - Add brokering to decrease effort
  - Interface SAMGrid to LCG and OSG
    - Increase resources and knowledge



Amber Boehnlein, FNAL

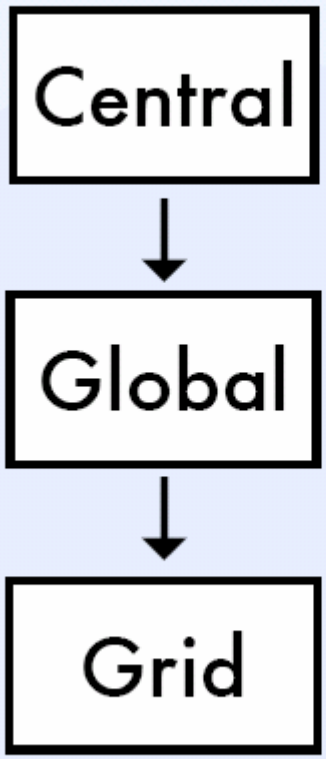
Virtual Center

- DO considers Remote computing essential. Use a model to attribute credit for computing contributed to the experiment
- For the value basis, determine the cost of the full computing system at FNAL costs, purchased in the yearly currency
  - Calculated based on the number of events, time/event, timescale on which to do the task with a nominal efficiency.
  - Performed as part of the yearly bottoms up budget request exercise
- Assign fractional value for remote contributions based on the number of events
  - Merit based assignment of value
  - Assigning equipment purchase cost as value ("Babar Model") doesn't take into account life cycle of equipment nor system efficiency or use or shared hardware
  - Computing planning board includes strong remote participation, representation to facilitate making requests for resources

Conclusions

- The DO computing is successful
- Reprocessing underway
  - T2-HEPGRID and Sprace are/will participate
- Use Virtual Center Concept to calculate the "value" that remote computing give the collaboration.
- DO continues to pursue a global vision for the best use of resources by moving towards interoperability with LCG and OSG

CDF approach to Grids

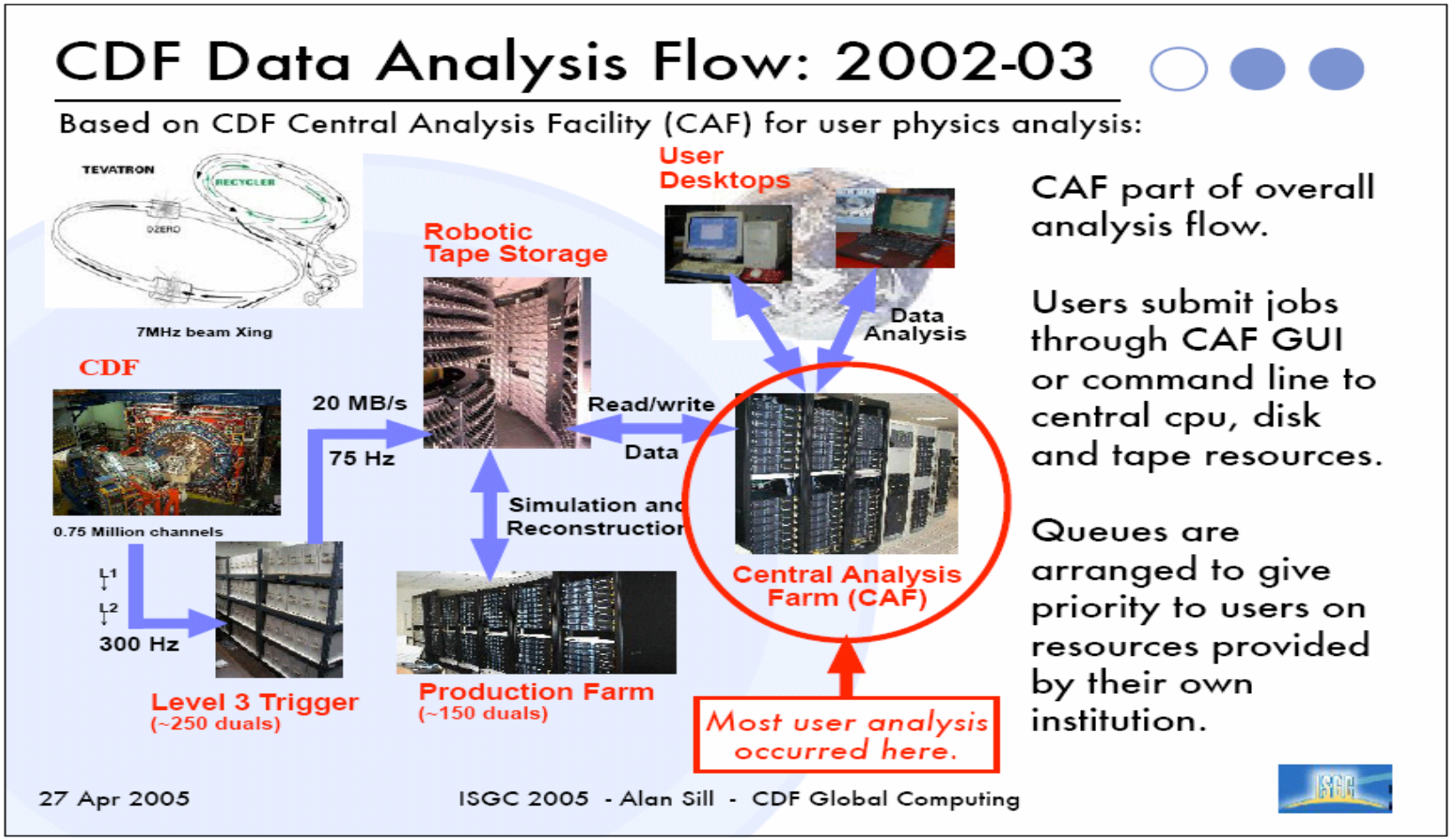


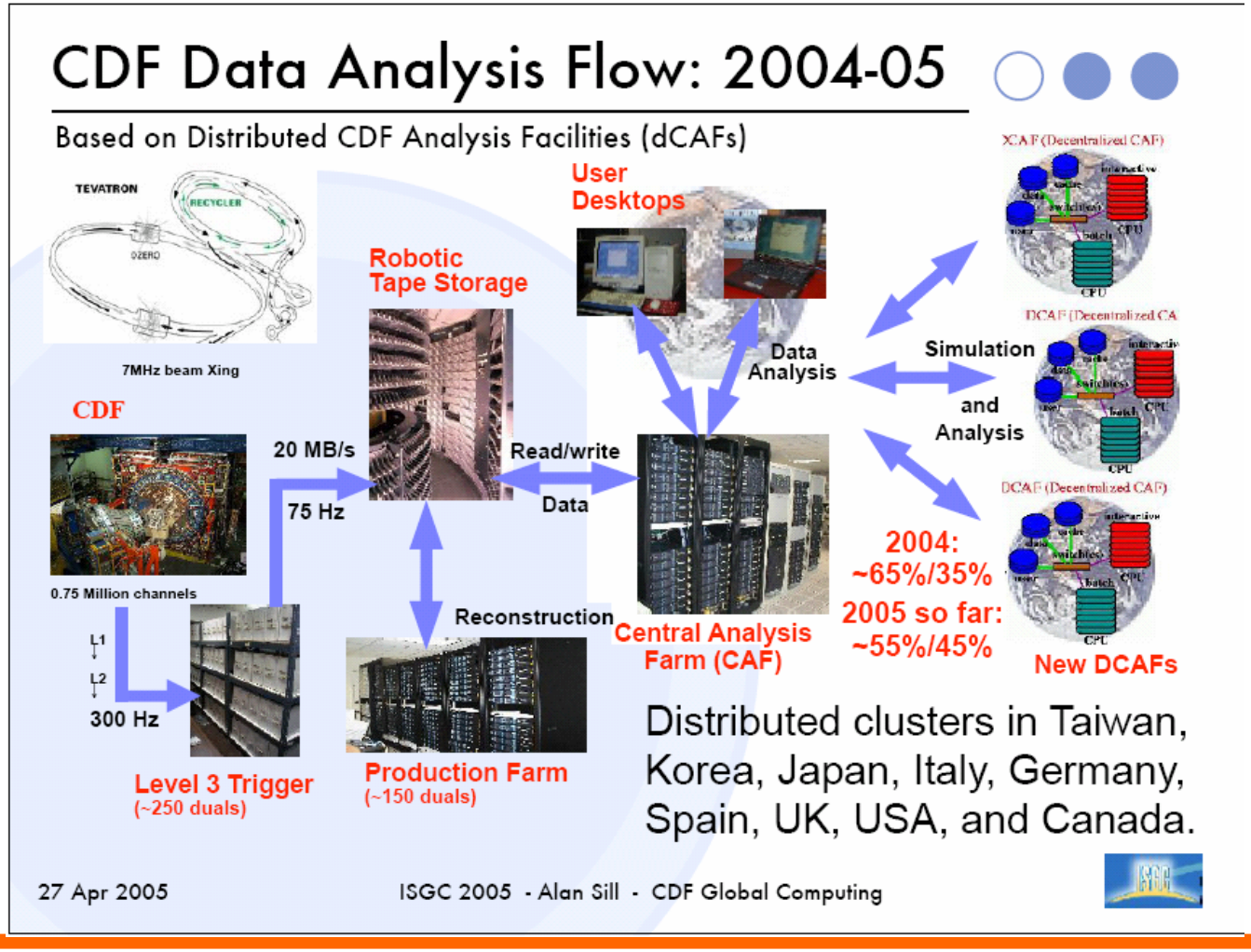
We developed a large central computing resource based on Linux farms with a simple job management scheme.

We extended the above model, including its command line interface and gui, to manage and work with remote resources.

We are now in the process of adapting and converting our work flow to the Grid.

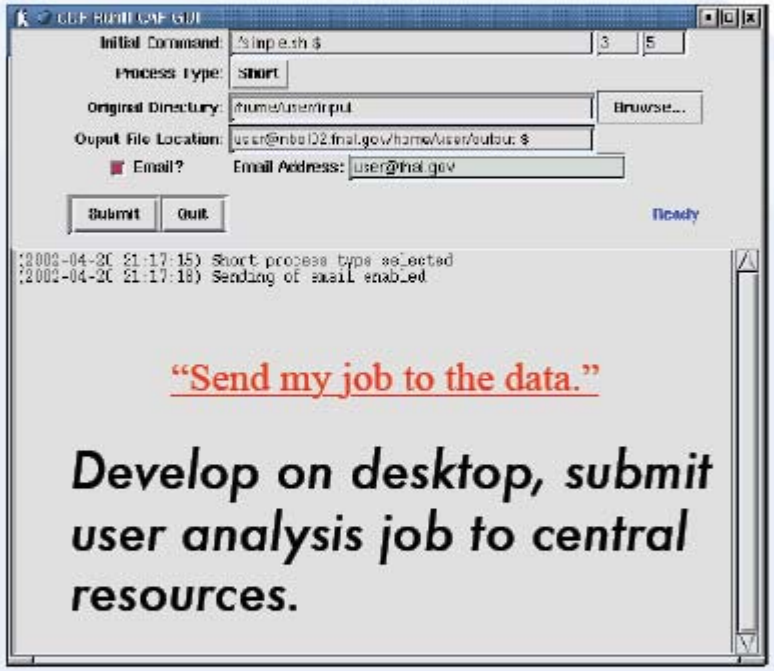
CAF – successful and GUI popular

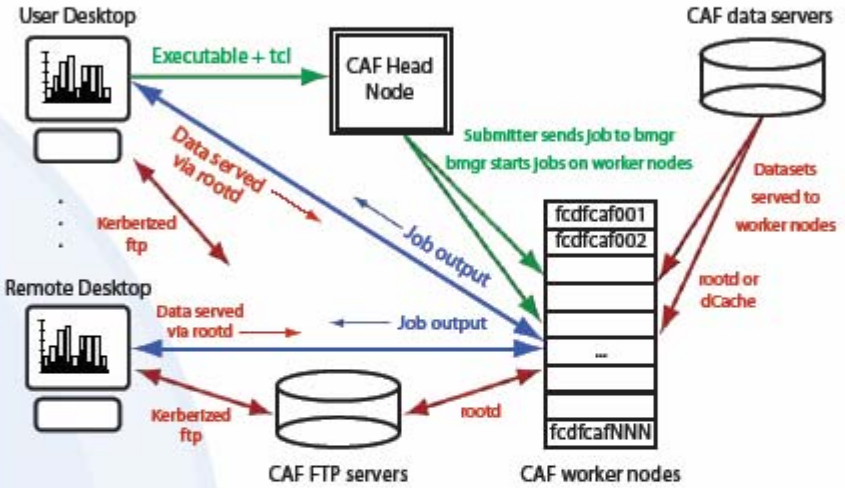



CDF Data Analysis Flow: 2004-05
Based on Distributed CDF Analysis Facilities (dCAFs)
TEVATRON
RECYCLER
0ZERO
7MHz beam Xing
CDF
0.75 Million channels
L1
L2
300 Hz
Level 3 Trigger
(~250 duals)
20 MB/s
75 Hz
Robotic
Tape Storage
Read/write
Data
Reconstruction
Production Farm
(~150 duals)
User
Desktops
Data
Analysis
Central Analysis
Farm (CAF)
Simulation
and
Analysis
2004:
~65%/35%
2005 so far:
~55%/45%
New DCAFs
Distributed clusters in Taiwan,
Korea, Japan, Italy, Germany,
Spain, UK, USA, and Canada.
27 Apr 2005
ISGC 2005 - Alan Sill - CDF Global Computing

CAF GUI * interface

* (Command line submission also possible.)





User submits job, which is tarred and sent to CAF cluster. Results packed up and sent back to or picked up by user.

Most active CDF physicists became familiar with operation of this system.

Environment on a CDF CAF

- All basic CDF software pre-installed on CAF.
- Authentication via Kerberos
  - Jobs are run via mapped accounts with authentication of actual user through special principal
  - Database, data handling remote user ID passed on through lookup of actual user via special principal (important for monitoring)
- User's analysis environment comes over in tarball - no need to pre-register or submit only certain jobs.
- Job returns results to user via secure ftp/rcp controlled by user script and principal.

Elements of current CDF approach

- Cluster technology (CAF = "CDF analysis farm") extended to remote sites (dCAFs = decentralized CDF Analysis Farms).
- Multiple batch systems supported: Converting from FBSNG system to Condor on all dCAFs.
- SAM data handling system required for offsite dCAFs:
  - Either use to pin pre-defined datasets at remote locations, or can use SAM's automatic cache management features.
  - SAM also used on CDF systems with non-CAF architectures.
- Fully functional for users => heavily in use.

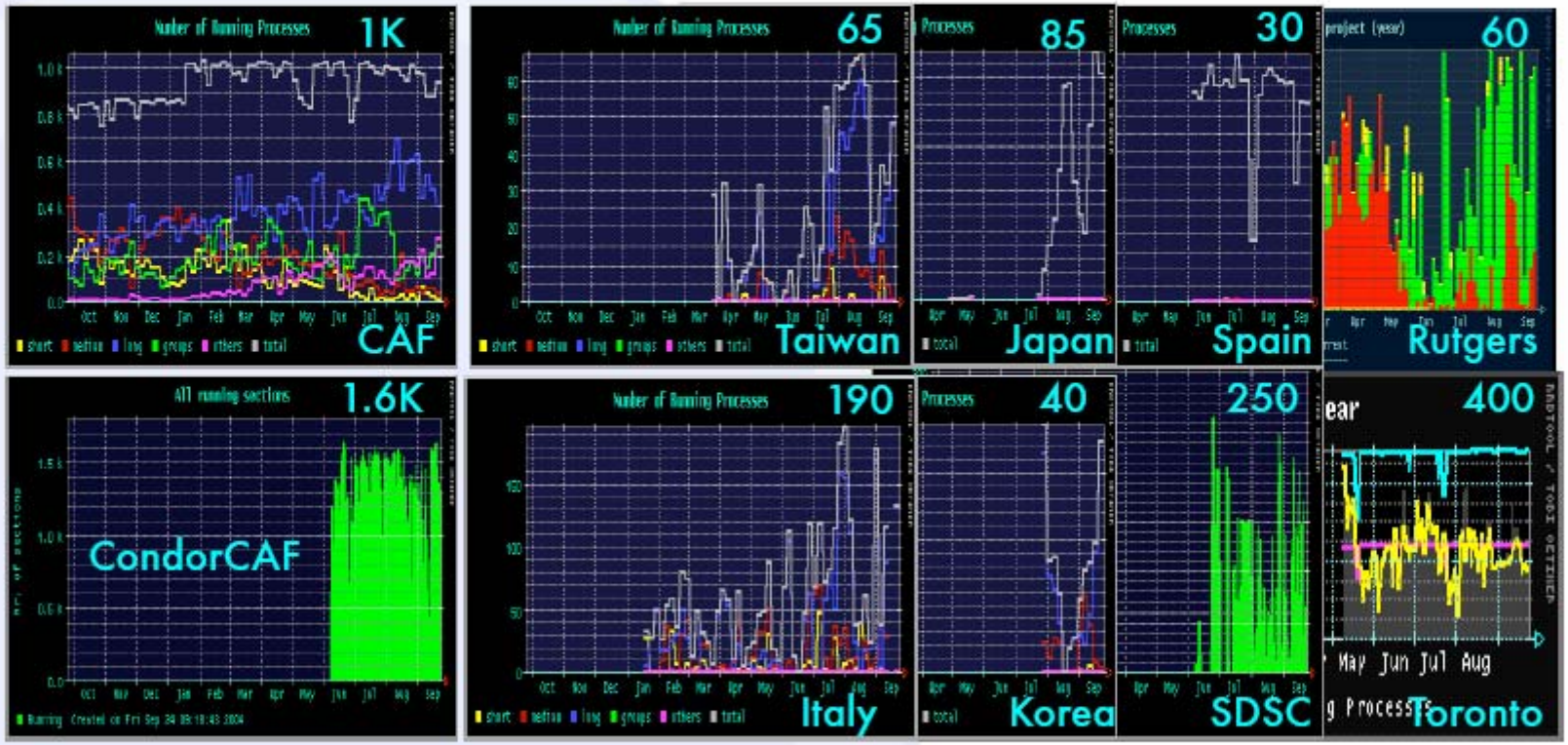
Current CDF Dedicated Resources:

Current Resources [*]			As of Apr. 15, 2004
Cluster Name and Home Page	Monitoring and Direct Information Links	CPU (GHz)	Disk space (TBytes)
Original FNAL CAF	queues, user history, analyze, ganglia, sam station, consumption	1000	370
FNAL CondorCAF (Fermilab)	queues, user history, analyze, ganglia, sam station, consumption	2200	(shared w/CAF)
CNAFCAF (Bologna, Italy)	queues, user history, analyze, resources, network, sam station, datasets, consumption	480	32
KORCAF (KNU, Korea)	queues, user history, ganglia, sam station, datasets, consumption	178	5.1
ASCAF (Academia Sinica, Taiwan)	queues, user history, ganglia, sam station, datasets, consumption	134	3.0
SDSC CondorCAF (San Diego)	queues, user history, analyze, ganglia, sam station, datasets, consumption	380	4.0
HEXCAF (Rutgers)	queues, cpu, sam station, datasets, consumption	100	4.0
TORCAF (Toronto CDF)	queues, user history, analyze, ganglia, disk status, sam station, datasets, consumption	576	10
JPCAF (Tsukuba, Japan)	queues, user history, ganglia, sam station, datasets, consumption	152	10
CANCAF (Cantabria, Spain)	queues, user history, ganglia, sam station	50	1.5
MIT (Boston, USA) (MC only)	queues, user history, analyze	322	3.2
	Current Totals [*]:	5572	448

** (Counts only resources openly available to all CDF users)*

Utilization jumped up quickly!

- Experiment-wide global aggregate monitoring not done yet (in progress).
- Off-site new dCAF usage jumped up quickly after deployment in each case.



Central ———————————— Off-Site ————————————

"Dedicated" needs to go to "Grid"

Current CDF Dedicated Resources:

Current Resources [*] — As of Apr. 15, 2004

Cluster Name and Home Page	Monitoring and Direct Information Links	CPU (GHz)	Disk space (TBytes)
Original FNAL CAF	queues, user history, analyze, ganglia, sam station, consumption	1000	370
FNAL CondorCAF (Fermilab)	queues,	2200	(shared w/CAF)
CNAFCAF (Bologna, Ita			32
KORCAF (KNU			5.1
ASCAF (Ac Taiwan)			0
SDSC Co Diego)			0
HEXCAF (Ru			4.0
TORCAF (Toronto			10
JPCAF (Tsukuba, Japan)			10
CANCAF (Cantabria, Spain)	queues, user history,	50	1.5
MIT (Boston, USA) (MC only)	queues, user history, analyze	322	3.2
	Current Totals [*]:	5572	448

2.3 of 5.6 THz now offsite in DCAFs (not counting special-purpose and "opportunistic" computing); other clusters and Grid access in progress.

* (Counts only resources openly available to all CDF users)

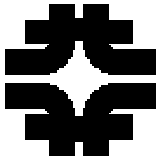
27 Apr 2005 ISGC 2005 - Alan Sill - CDF Global Computing

Towards Full Grid Methods

Testing various approaches to using Grid resources:

1. Adapt the CAF infrastructure to run on top of the Grid using Condor glide-ins
   - *Advantages: Grid neutral; works both on LCG and OSG, needs just a Globus gatekeeper; can opportunistically use idle resources. Disadvantage: Need a CDF head node near each CE*
2. Use direct submission via CAF interface to OSG and LCG
   - *Exploring differences between job submission models and resource broker environments; work in progress now.*
3. Use SAMGrid/JIM sandboxing as an alternate way to deliver experiment + user software.
4. Combine dCAFs w/ Grid resources (can be done at any time).

Global Grid Community



Federating Grids

As the US representative to the LHC Computing Grid Grid Deployment Board (LCG GDB) I have repeatedly pushed the case for the LHC Computing Grid to be a "Grid of Grids" – not a distributed Computing center

- Resources must be able to belong to >1 Grid
  - E.g. NorduGrid, UK Grid, Open Science Grid, TeraGrid, Brazil Grid?, University ABC Grid and FermiGrid!
- Grids infrastructures must learn to interoperate
- Governance, policies, security, resource usage and service level agreements must be considered at many levels including at the level of a Federation of Grids

› I recently called this a "Grid Bazaar" (ISGC2005)

Grid - its really about collaboration!

- It's about sharing and building a vision for the future
  - And it's about getting connected
- It's about the democratizati
  science

International Grid Workshop, Taiwan, May 2005