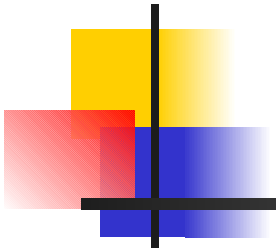
Metadata for the Common Physicist

Rick St. Denis, University of Glasgow
Wyatt Meritt, Julie Trumbo,Fermilab



- What is SAM?
- From 1 to 2, 2 to N – D0, CDF, MINOS, CMS
- Use Cases: reality, abstract, Abstract meets Reality
- Operations and Support
- Conclusions and Future

The SAM-Grid Team

Project Co-Leaders:
Wyatt Merritt (CD/Run II); Rick St. Denis (CDF/ U Glasgow)

Technical Co-Leaders:
Rob Kennedy (CD/CCF); Sinisa Veseli (CD/Run II)

Core Developers (SAM components):
Lauri Loebel Carpenter, Andrew Baranovski, Steve White, Carmenita Moore,* Adam Lyon, Petr Vokac,*** Mariano Zimmler***, Matt Leslie

Core Developers (JIM components):
Igor Terekhov,** Gabriele Garzoglio, Sankalp Jain,** Aditya Nishandar**

Support for CDF Migration:
Fedor Ratnikov, Randy Herber, Art Kreymer, Morag Burgon-Lyon,** Valeria Bartsch, Stefan Stonjek, Krzysztof Genser

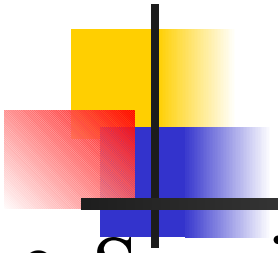
Database support:
Anil Kumar

Associated external projects:
PPDG, GridPP, SBIR II

Core Development Currently @ 7 FTEs

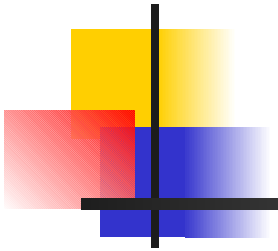
*** Deceased | ** Left project | *** Summer Students**

What is SAM?



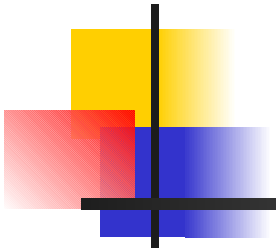
- Sam is a bundle of data handling services modelled on a relational database using Oracle and having interfaces to other services
- Services include
  - Routed file delivery with accounting
  - Caching and Replica location
  - User authentication and authorization
  - Dataset creation based on file and processing metadata
  - Local and remote monitoring capabilities

What is SAM?

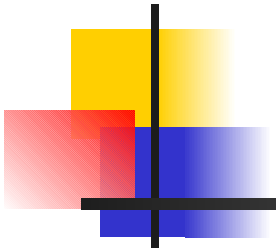


- Data handling system for Run II DØ and CDF
- SAM manages file storage (replica catalogs)
  - Data files are stored in tape systems at FNAL and elsewhere (most use ENSTORE at FNAL)
  - Files are cached around the world for fast access
- SAM manages file delivery
  - Users at FNAL and remote sites retrieve files out of file storage. SAM handles caching for efficiency
  - You don't care about file locations
- SAM manages file meta-data cataloging
  - SAM DB holds meta-data for each file. You don't need to know the file names to get data
- SAM manages analysis bookkeeping
  - SAM remembers what files you ran over, what files you processed successfully, what applications you ran, when you ran them and where
- Designed for PETABYTE (10^{15}) sized experiment datasets (that's us)!

SAM Terms and Concepts



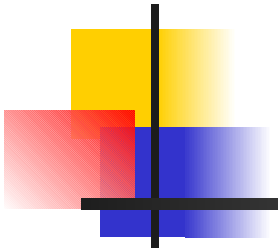
- A project runs on a station and requests delivery of a dataset to one or more consumers on that station.
- **Station**: Processing power + disk cache + (connection to tape storage) + network access to SAM catalog and other station caches
  Example: a linux analysis cluster at D0
- **Dataset**: metadata description which is resolved through a catalog query to a list of files. Datasets are named.
  Examples: (syntax not exact)
  - data_type physics and run_number 78904 and data_tier raw
  - request_id 5879 and data_tier thumbnail
- **Consumer**: User application (one or many exe instances)
  Examples: script to copy files; reconstruction job



Interfacing

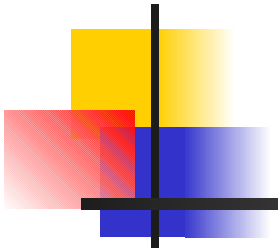
- Interfaces:
  - Batch system interaction
  - Experiment-specific metadata
  - Storage and use of external caching

SAMGrid & Grid Services



- Distributable sam_client provides access to:
  - VO storage service (sam store command, interfaced to sam_cp)
  - VO metadata service (sam translate constraints)
  - VO replica location service (sam get next file)
  - Process bookkeeping service

From 1 to 2 to many



- D0 started SAM: 40 active sites (9 @ FNAL)
- CDF uses SAM: 25 active sites (2 @ FNAL)
- MINOS committed
- CMS starting

CDF SAM Deployment

SAM Stations:

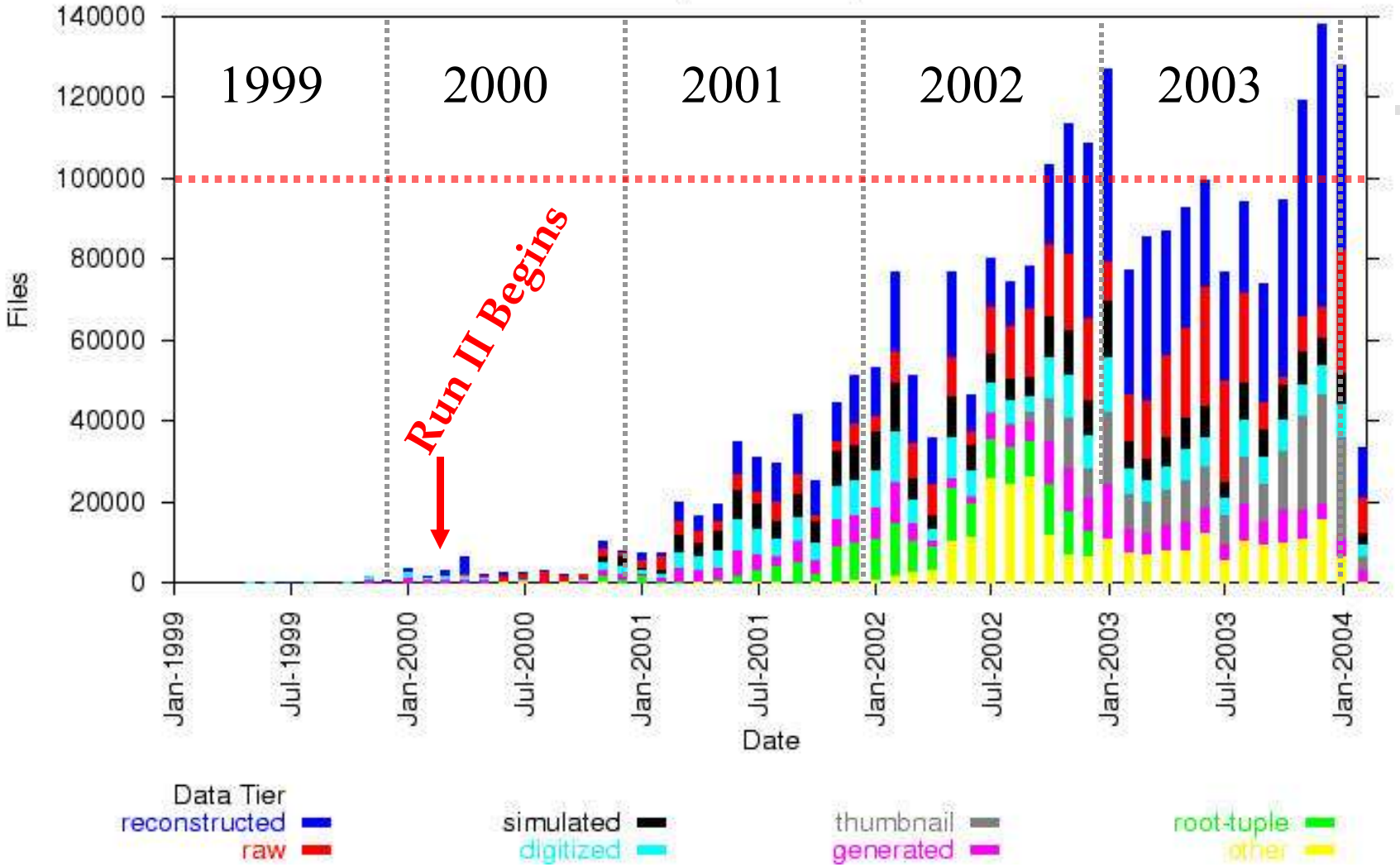
Monitor Level: Critical	Host	Version	Up Since
cdf-sam	fcdfdata016.fnal.gov	v4_2_1_69	19 Jul 2004 18:34:19

Monitor Level: High	Host	Version	Up Since
cdf-cnaf	cdfsam.cnaf.infn.it	v4_2_1_63	22 May 2004 07:19:01
cdf-fzkka	cdf.fzk.de	v4_2_1_72	23 Jul 2004 10:58:27
cdf-knu	cluster67.knu.ac.kr	v4_2_1_72	13 Jul 2004 03:38:41
cdf-oxford	matrix.physics.ox.ac.uk	v4_2_1_71	20 Jul 2004 11:34:23
cdf-rutgers	hexsam.rutgers.edu	v4_2_1_63	06 Jul 2004 18:16:06
cdf-sdsc	t2sam01.sdsc.edu	v4_2_1_72	22 Jul 2004 14:21:40
cdf-taiwan	ascaf.sinica.edu.tw	v4_2_1_72	20 Jul 2004 09:43:33
cdf-toronto	bigmac-cdf03.physics.utoronto.ca	v4_2_1_63	14 Jun 2004 10:57:44
cdf-trieste	pccdf2.ts.infn.it	v4_2_1_63	19 Jul 2004 13:31:03
cdf-ttu	pantheon.cs.ttu.edu	v4_2_1_63	23 Jul 2004 08:58:07

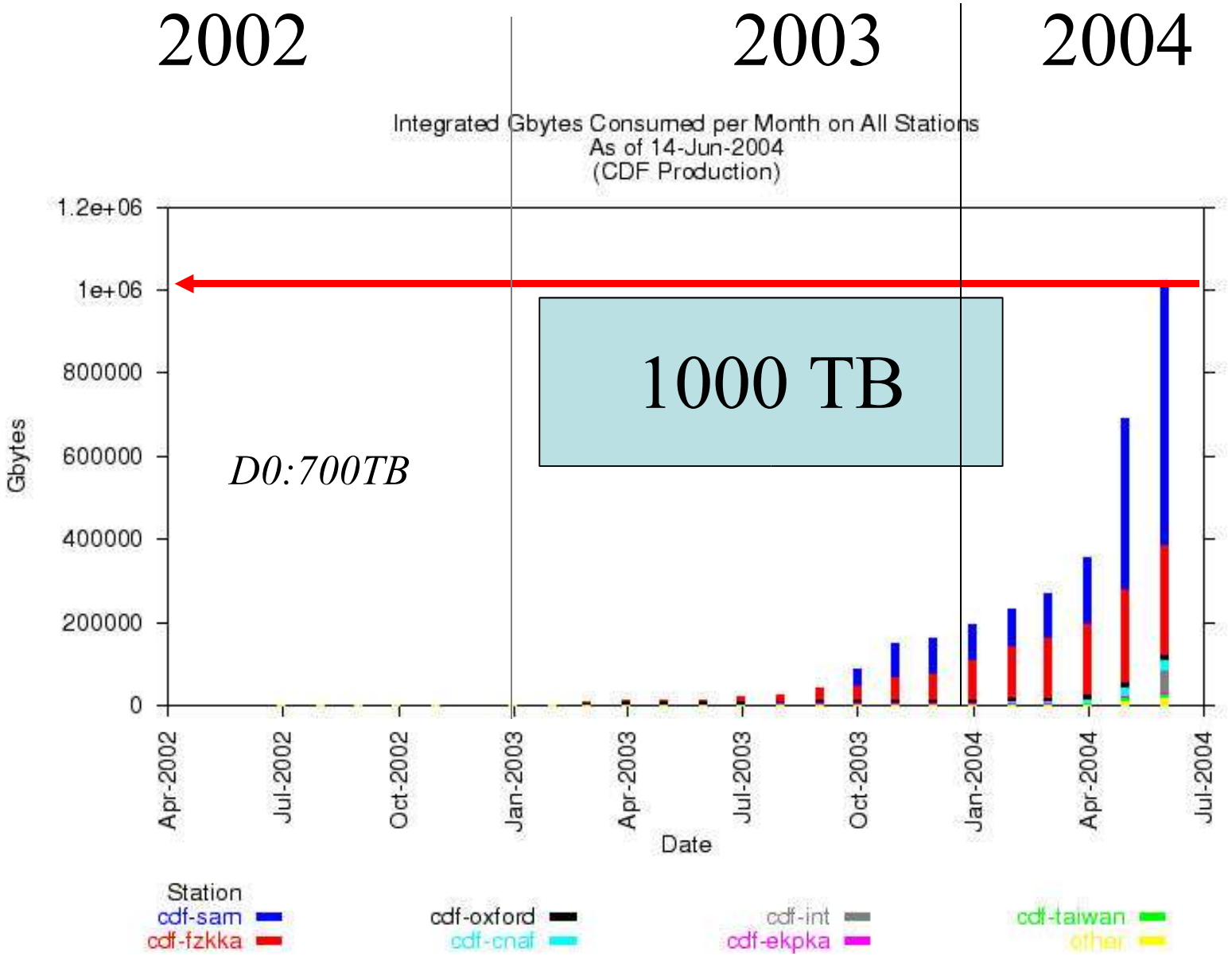
cdf-liverpool			
cdf-mit-1			
cdf-ncdf151			
cdf-ral	cdfa.rl.ac.uk	v4_2_1_63	08 Jul 2004 09:34:33
cdf-rdk-fnal-1			
cdf-sam2			
cdf-scotgrid			
cdf-scotgrid-2			
cdf-taiwan2			
cdf-test	cdfpcb.fnal.gov	v4_2_1_72	20 Jul 2004 01:01:02
cdf-ttu-hpcc			
cdf-ttu-phys	castor.phys.ttu.edu	v4_2_1_72	19 Jul 2004 22:36:10
cdf-tufts			
cdf-ucsd			
samadams	samadams.fnal.gov	v4_2_1_63	21 Jun 2004 16:20:32
samgfarm	samgfarm1.fnal.gov	v4_2_1_64	16 Jul 2004 17:08:40

Sam Statistics - DØ

Files delivered by month



Total CDF Files To User



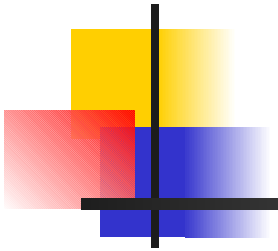
Sam in Operation

- Looking at SAM in operation -

  SAM TV @ DØ **SAM TV @ CDF**

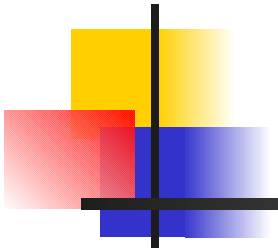
  - Currently created from log files
  - Version in development is created from MIS database, filled by new MIS server

The Impact of CDF requirements and nonrequirements



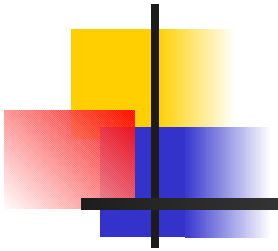
- Many runs in a file and hooks to separate luminosity bookkeeping
- Clean separation of file types: Generic, MC, processed
- Keep track of group responsible for file
- Require: format, size, crc type/value, file content status id
- Not Required: data tier, file partition, process id, stream, event count, first/last event number start/end times
- Removed: min bias number, type and physics process (left over from MC: put into generic metadata)

Introspection



- CDF participation drove a revisiting of the D0 Design
- An entirely new user community enforced validity of user requirements and forced designers and implementers to confront the difficulties of implementation
- Rules in Relational Database were moved in part to API to allow for experiment-specific needs.
- Boundries became more clearly defined

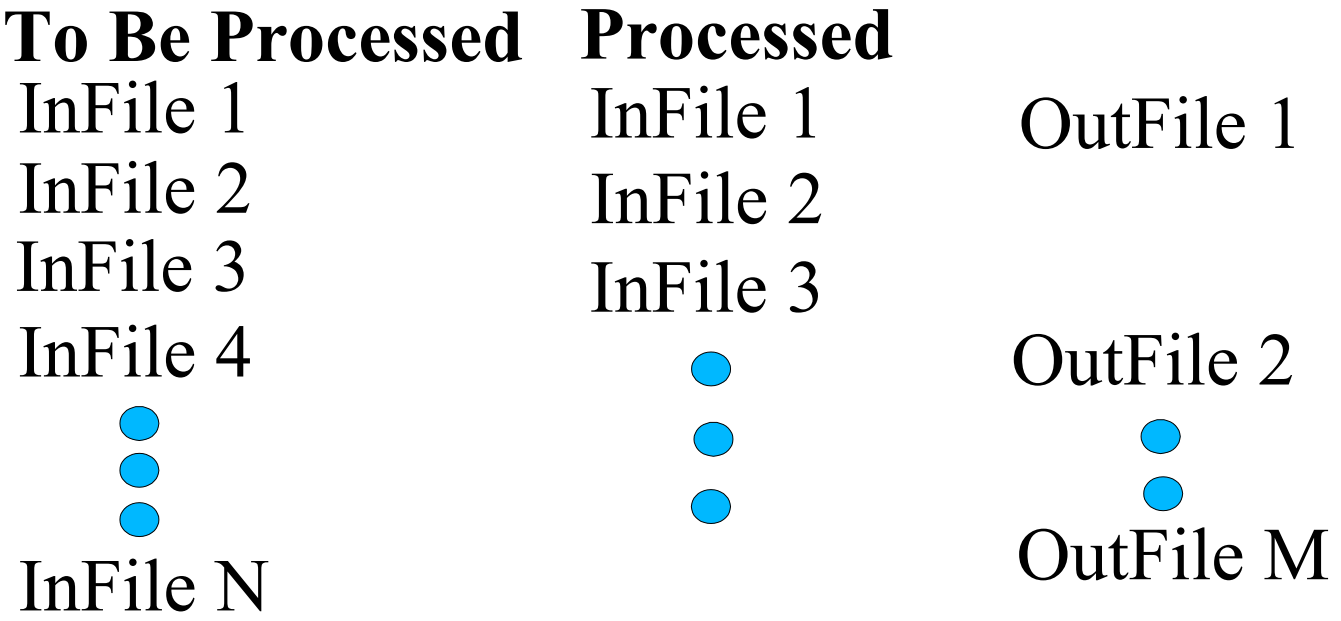
Implement Rules in API



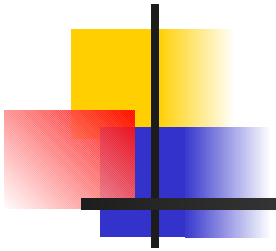
- physicsGeneric
  - **Data tier** is *unofficial reco* (D0)
- NonPhysicsGeneric
  - **File status** of *being imported* or *deleted (CDF)*
- Imported detector
  - **File status** of *available* with **Data tier** of *raw* and 17 characters.
- Imported Simulated and so on....

Valid Data Groups:Workflow-Data handling interaction

- Specifies the transitions in a workflow step
- Need to keep file operations as atomic when doing bookkeeping:



Requirements



- Data on computers in remote sites
- Data storage from remote sites
- Different modes:
  - MC production
  - User MC production
  - Reconstruction
  - User Reconstruction
  - Analysis
- CDF: 25K files, 1G; D0: 100K files, small

Current Resources

Cluster Name and Home Page	Monitoring and Direct Information Links	CPU (GHz)	Disk space (TBytes)
Original FNAL CAF	queues, user history, ganglia, sam station, consumption	1200	200
FNAL CondorCAF (Fermilab)	queues, user history, analyze, ganglia, sam station, consumption	2000	~(shared w/CAF)
CNAFCAF (Bologna, Italy)			7.5
KORCAF (KN			6
ASCAF (Ac Sinica, Tai			
SDSC Condo (San Diego)			.0
HEXCAF (Rutgers)	consu		4.0
TORCAF2 (Toronto CDF)	queues, ganglia, disk status, sam station, datasets, consumption	576	10
JPCAF (Tsukuba, Japan)	queues, user history, sam station, datasets, consumption	152	5.0
	Current Totals:	5012	234

1.8 of 5.0 THz is now offsite

Datasets Stored Locally on cdf-cnaf

DATASET ID	GBYTE		S	CACHED	LOCKED
xbhd0d	4096.0			408(1%)	325(1%)
xbhd0c	4096.00	140592895	5733	109(1%)	109(1%)
hbhd0d	4096.00	169644634	32359	351(1%)	351(1%)
jbot0h	649.09	3240403	690	1(0%)	none
gmbs09	1224.65	7676037	1330	17(1%)	17(1%)
bpel0d	138		2194	all	all
gpjj08			70	1(1%)	1(1%)
xpmm0d			2675	all	all
xpmm0c	524.53	16019542		all	all
jpmm08	575.70	27928			none

Expert Usage!

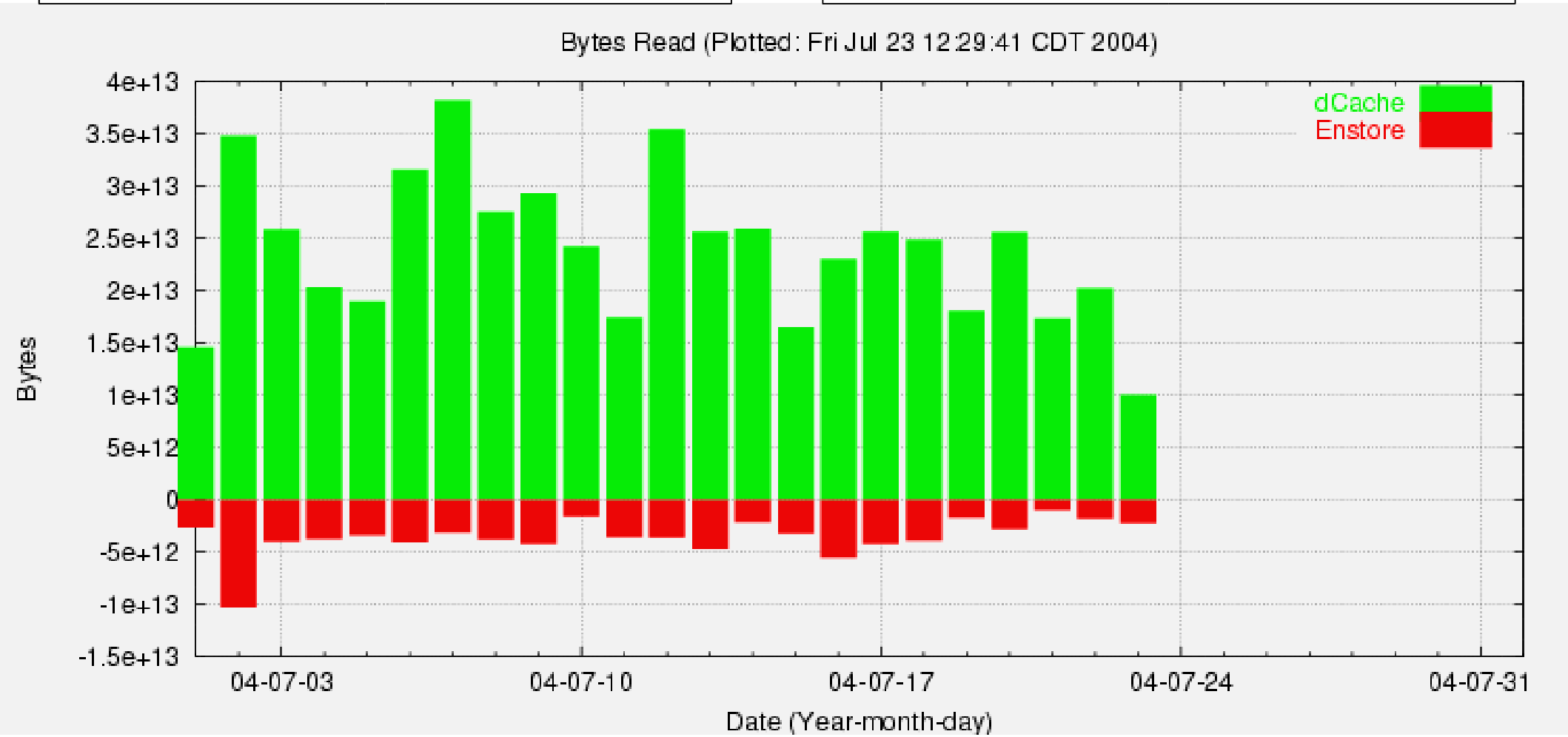
All in Cache

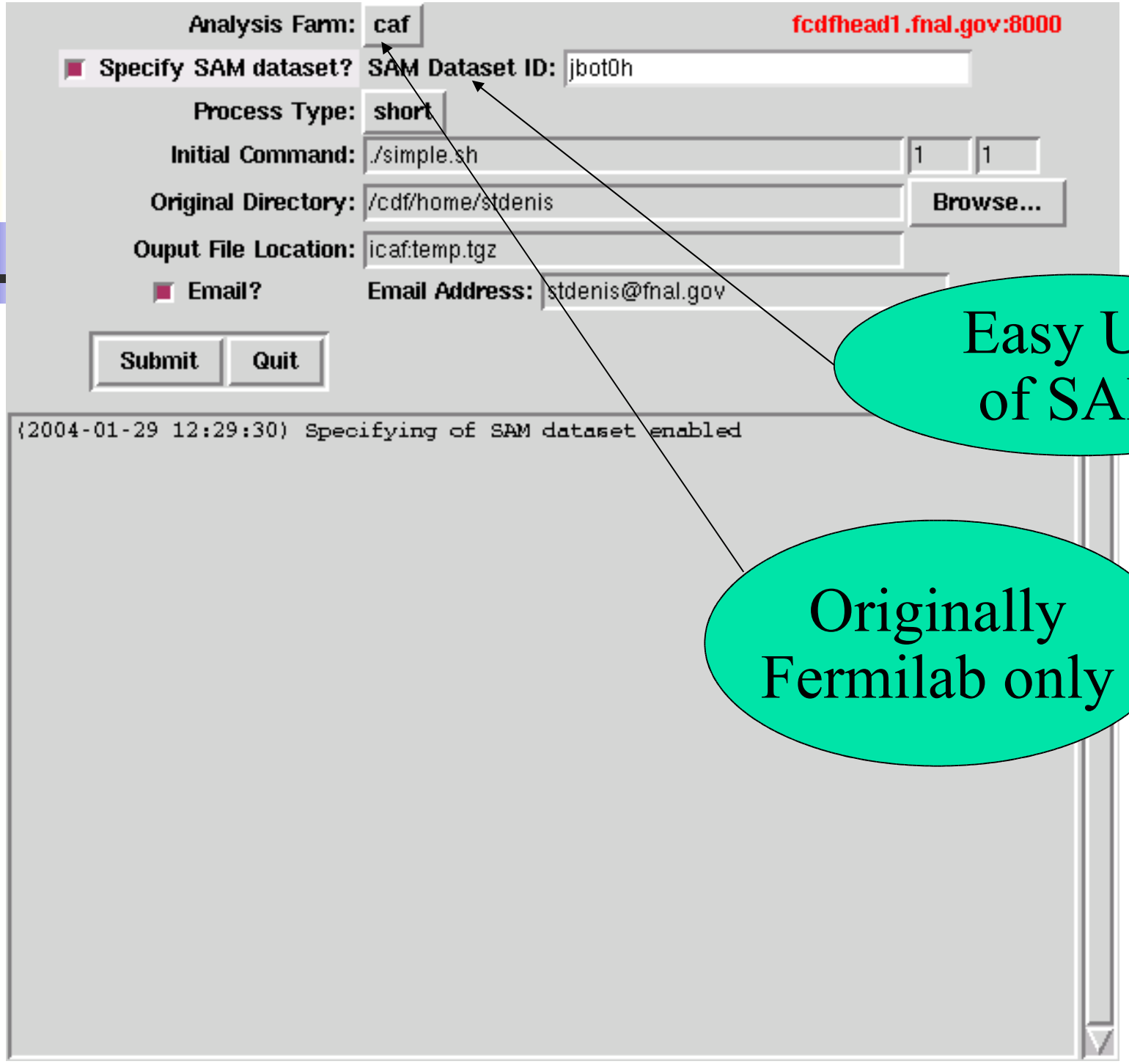
And Locked Via SAM

CDF Data Handling: Dcache on CAF

ALL CDF on CAF reads 25TB/Day

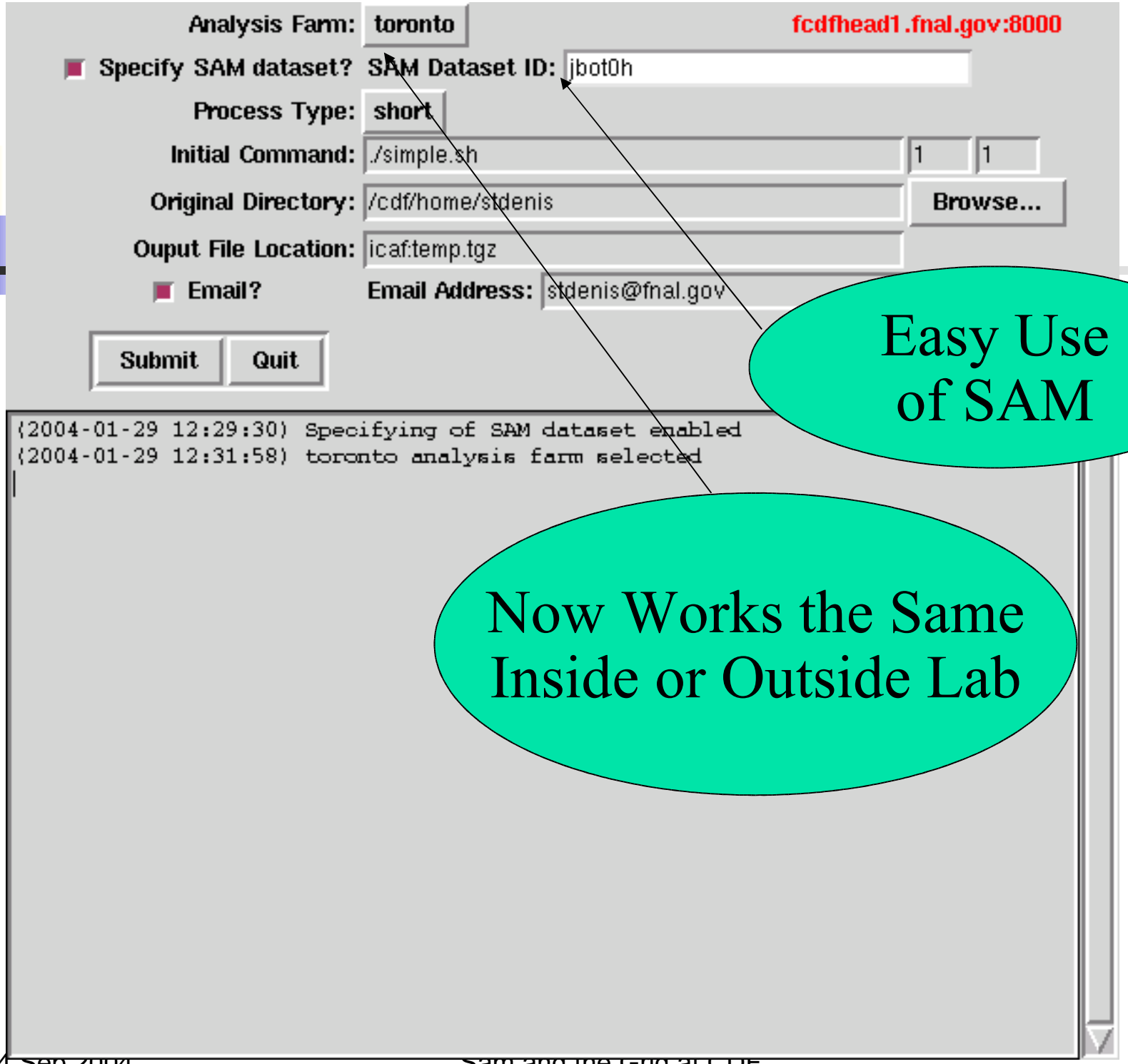
NonGrid Running

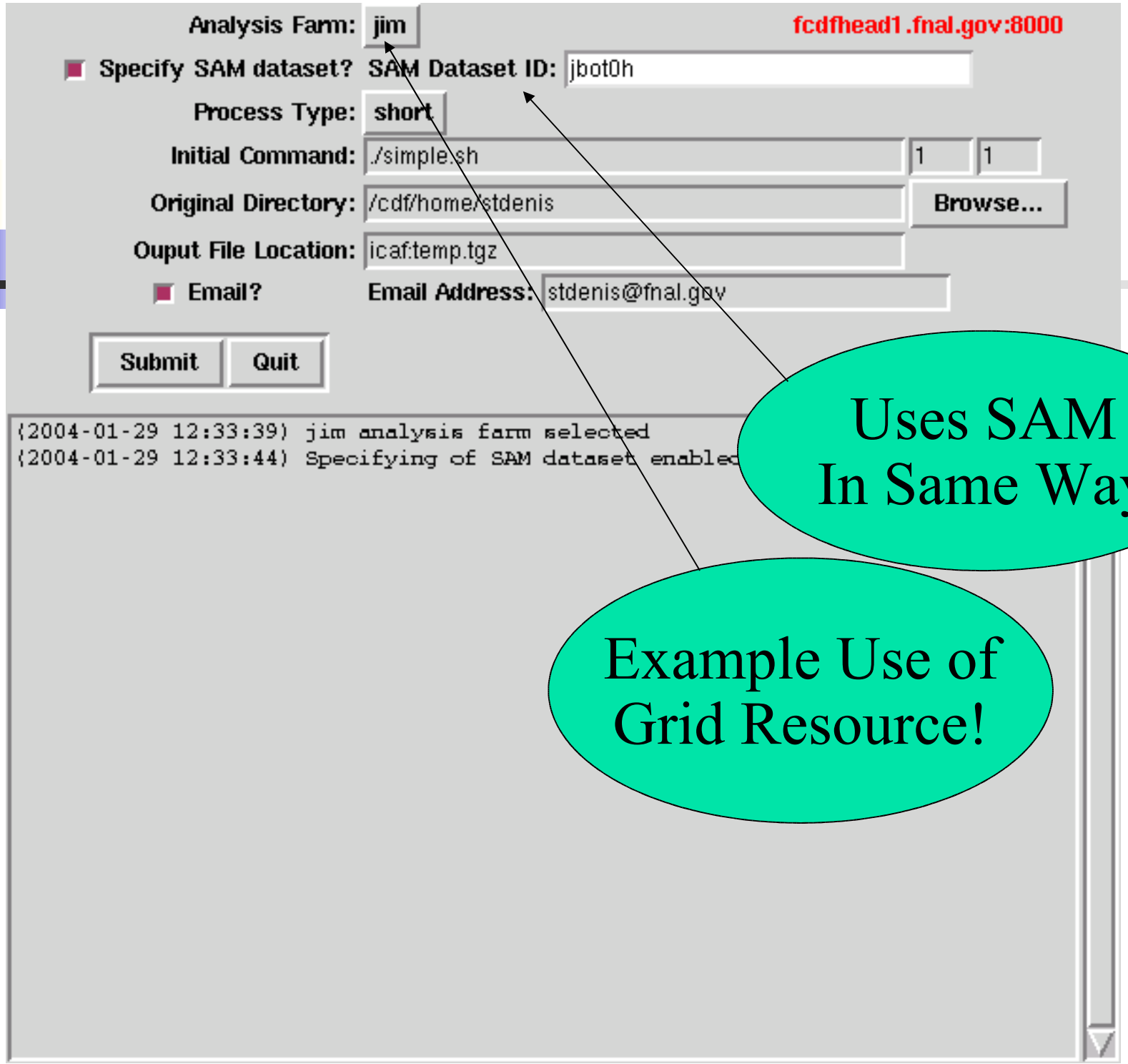




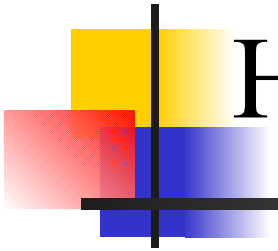
Easy Use of SAM

Originally Fermilab only


Analysis Farm:
toronto
fcdfhead1.fnal.gov:8000
Specify SAM dataset?
SAM Dataset ID:
jbot0h
Process Type:
short
Initial Command:
./simple.sh
1
1
Original Directory:
/cdf/home/stdenis
Browse...
Ouput File Location:
icaf:temp.tgz
Email?
Email Address:
stdenis@fnal.gov
Submit
Quit
(2004-01-29 12:29:30) Specifying of SAM dataset enabled
(2004-01-29 12:31:58) toronto analysis farm selected
Easy Use of SAM
Now Works the Same Inside or Outside Lab


Analysis Farm: jim
fcdfhead1.fnal.gov:8000
Specify SAM dataset?
SAM Dataset ID: jbot0h
Process Type: short
Initial Command: ./simple.sh 1 1
Original Directory: /cdf/home/stdenis
Browse...
Ouput File Location: icaf:temp.tgz
Email?
Email Address: stdenis@fnal.gov
Submit
Quit
(2004-01-29 12:33:39) jim analysis farm selected
(2004-01-29 12:33:44) Specifying of SAM dataset enable
Uses SAM
In Same Way
Example Use of
Grid Resource!

Use Cases Summary: HEPCAL,CDF,BABAR,ATLAS



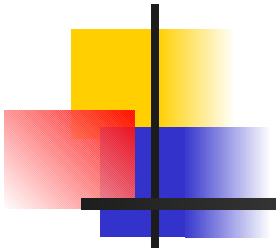
- Dataset Handling
- Analysis
- Job Handling

Dataset Handling I

- Specify a new dataset
- Read metadata for datasets
- Predefine metadata for a dataset to run
- Update and/or Add metadata for datasets
- Resolve physical data
- Download a dataset to a local disk
- Write experiment-specific metadata for the new dataset
- Access a Dataset

Dataset Handling II

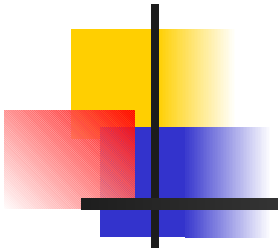
- Read all the visible metadata for a specified dataset
- Search for datasets whose metadata match a user query
- Publish private metadata
- Publish a private dataset
- Merge datasets
- Perform a transform on a dataset



Analysis

- Run a physics simulation
- Select a subset of data
- Run an algorithm over an input dataset

Job Handling



- Submit a job to a Grid
- Submit a job to the grid with predefined metadata
- Retrieve/Access the output of a job
- Estimate the system resource cost of running a job
- Monitor the progress of a job
- Repeat a previous job
- Recover failures in a previous job

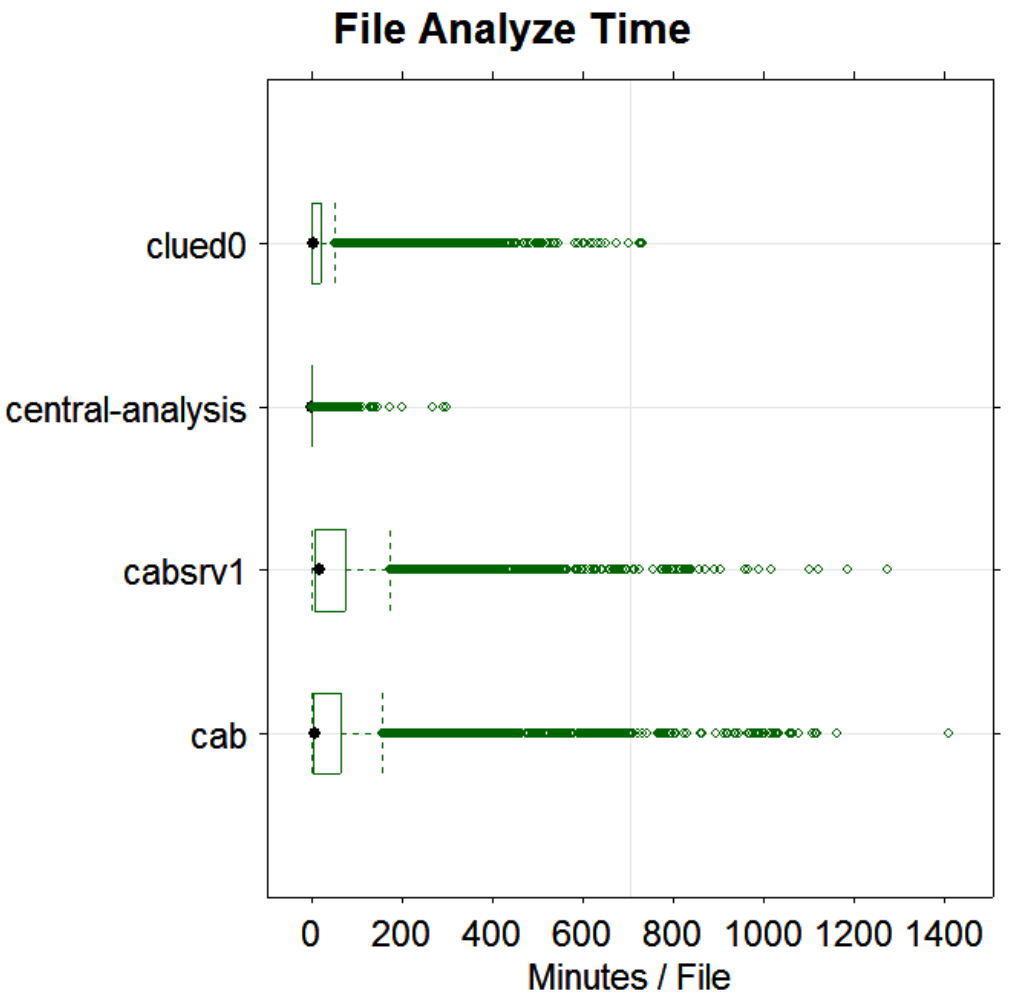
Operations

- Major operational issues Sep 03 – Sep 04
  - DØ – hardware problems with production database machine (central point of failure) Dec03-Jan04; 15% drop in file deliv

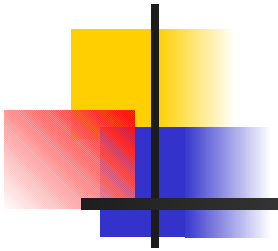
Problems Encountered/Solved/Unresolved

- CDF Contentious design issues Sep 03 – Sep 04
  - installation difficulties
  - file name as GUID no change to model
  - interface into experiment framework work in SAM
  - communication with dcache work in SAM, future work
  - use of dimensions and parameters proposed work in SAM
  - process bookkeeping future work in SAM
- MINOS – file delivery ordering & grouping no change to model

SAM Statistics - Operations Data



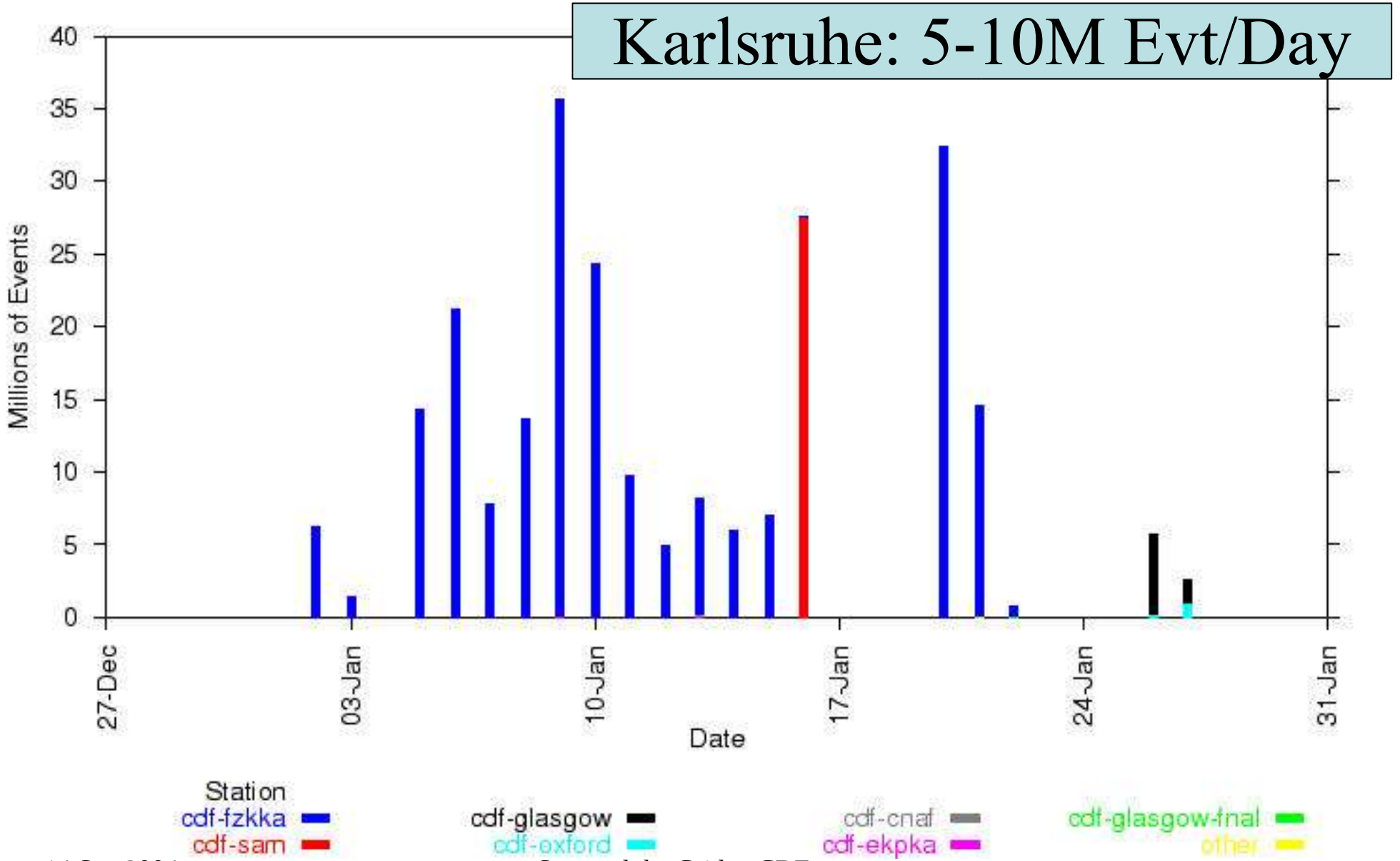
SAM: The work plan for the next 2 years



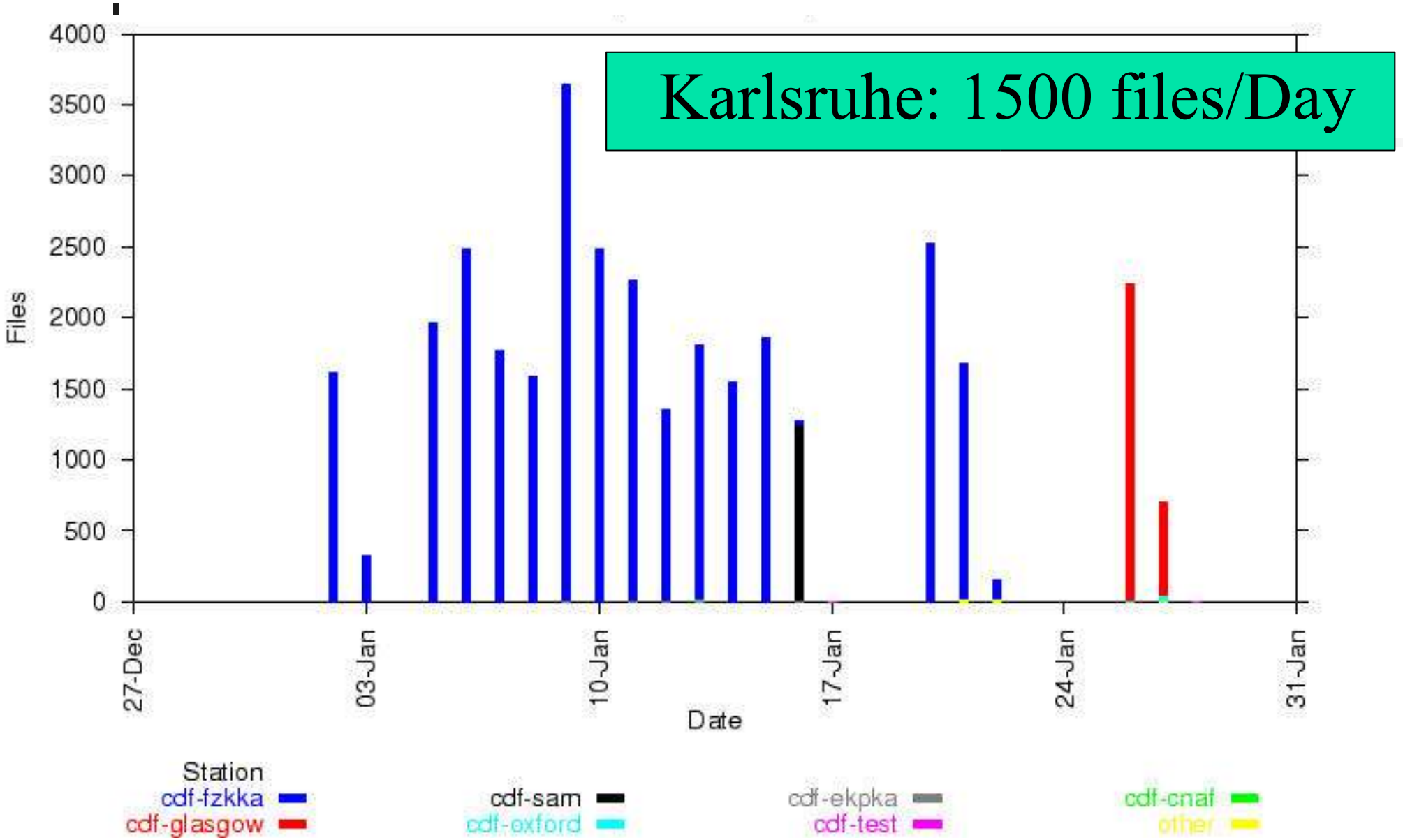
- Evaluate technology changes/upgrades
  - Improvements for installation/config management
  - CORBA to Web Services
  - XML based logging
  - Distributed database
  - Merge SAM catalog w/ other replica schemas
  - Working with SRM
  - Interaction of tools with data handling: Workflow, local and global job management
  - VO Organisation/security: file transfer

Extra Slides

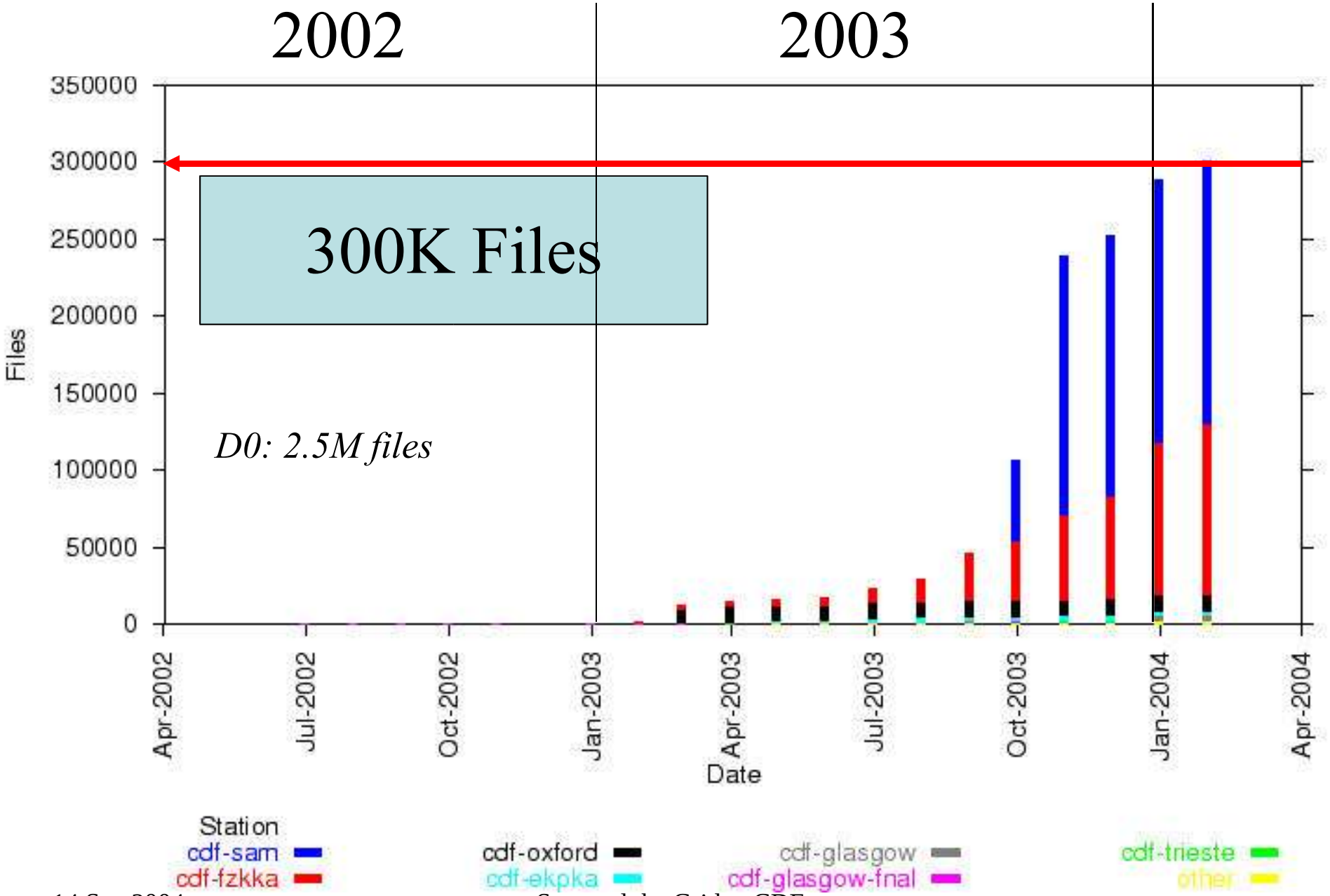
CDF Events Transferred per Month



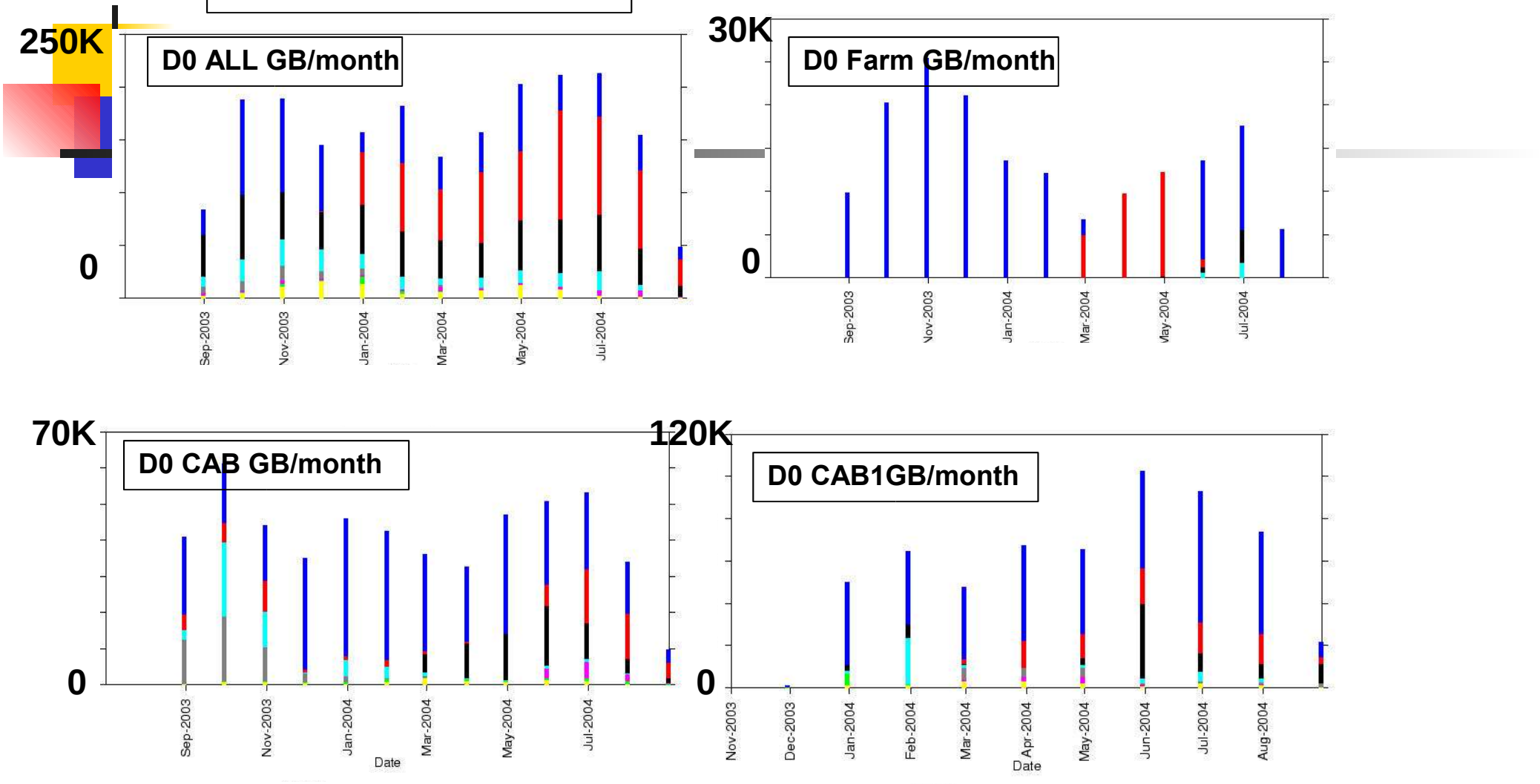
CDF Files in a Month



All CDF Files Moved by SAM

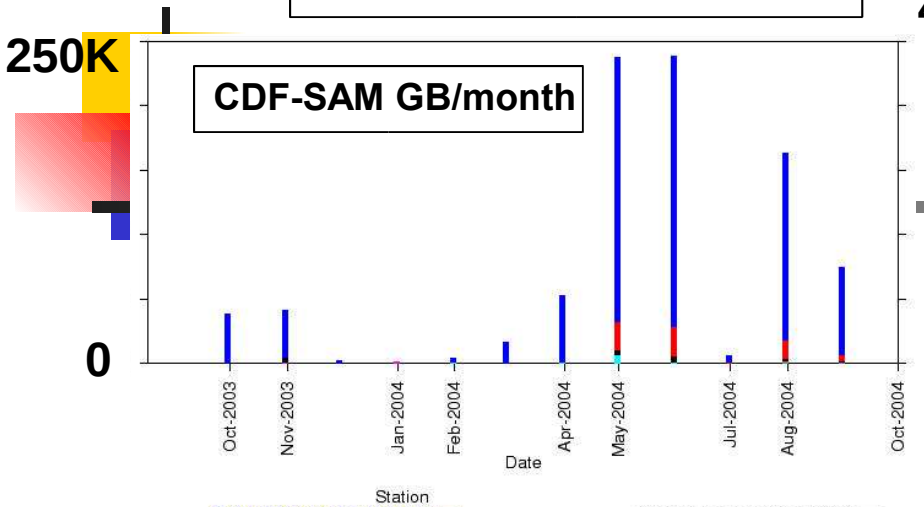


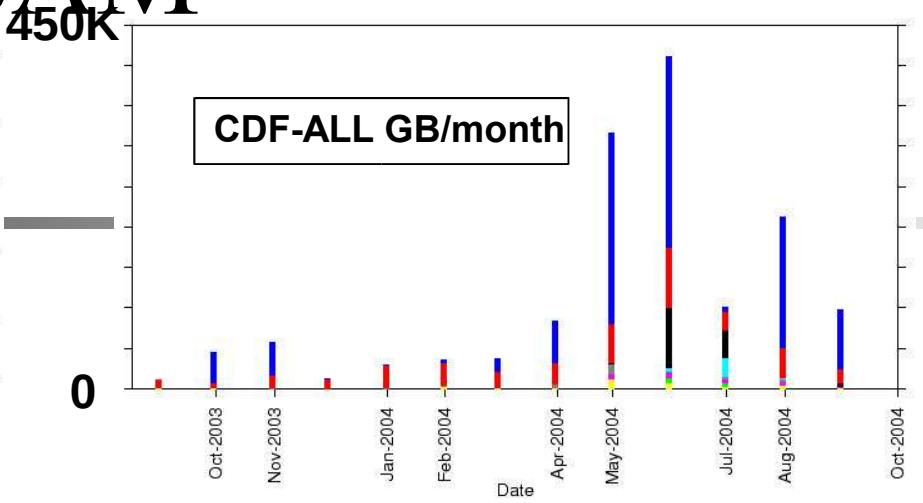
Usage Statistics for D0 SAM

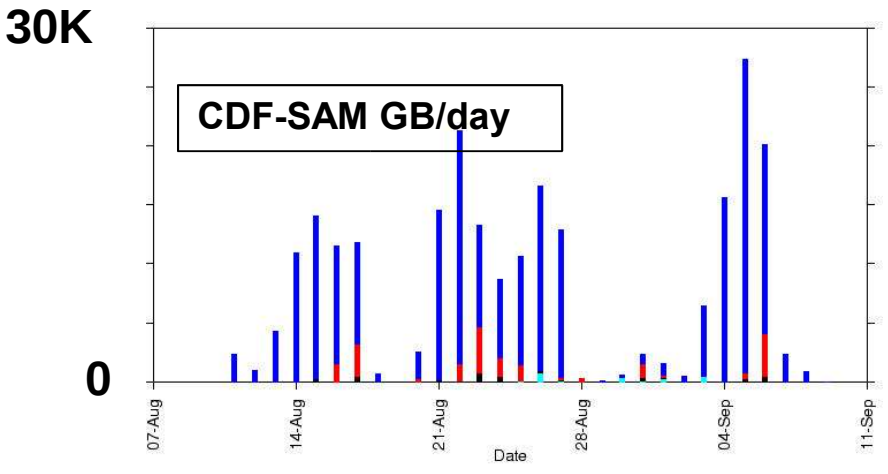


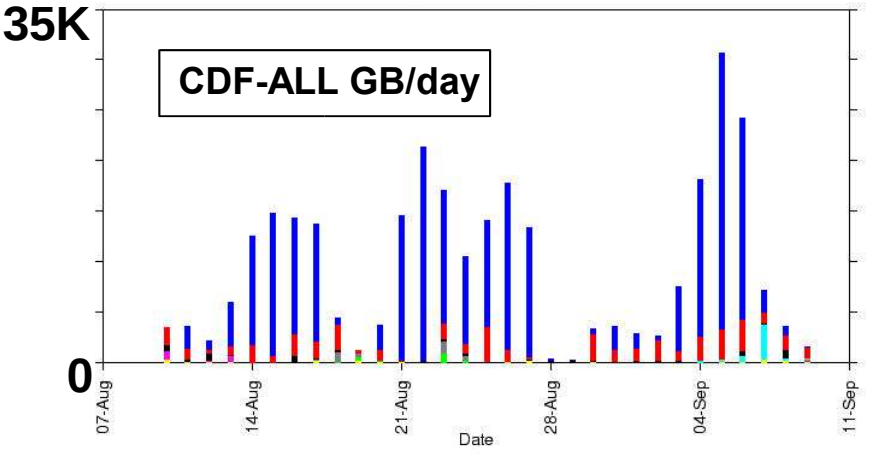
Usage Statistics for CDF SAM

Sum = 1.5 PB; 12B evts







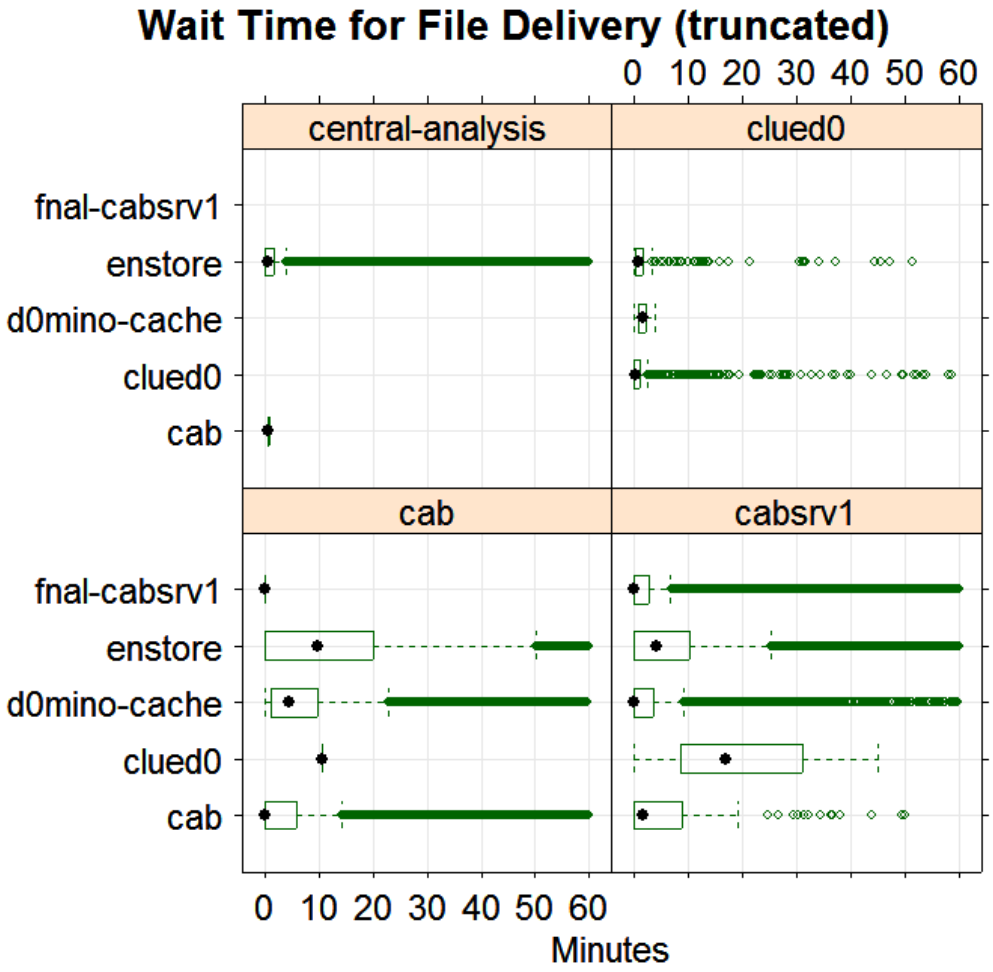


Scale of CDF Offsite Requirements

	THz	%offsite	CPU Speed	#duals offsite
FY04	3.7	25%	3GHz	150
FY05	9.0	50%	5GHz	360 more
FY06	16.5	50%	8GHz	220 more

6-7 sites, 100 duals each (or larger number of smaller sites), by 2006
+ equivalent capacity @FNAL

SAM Statistics - Operations Data



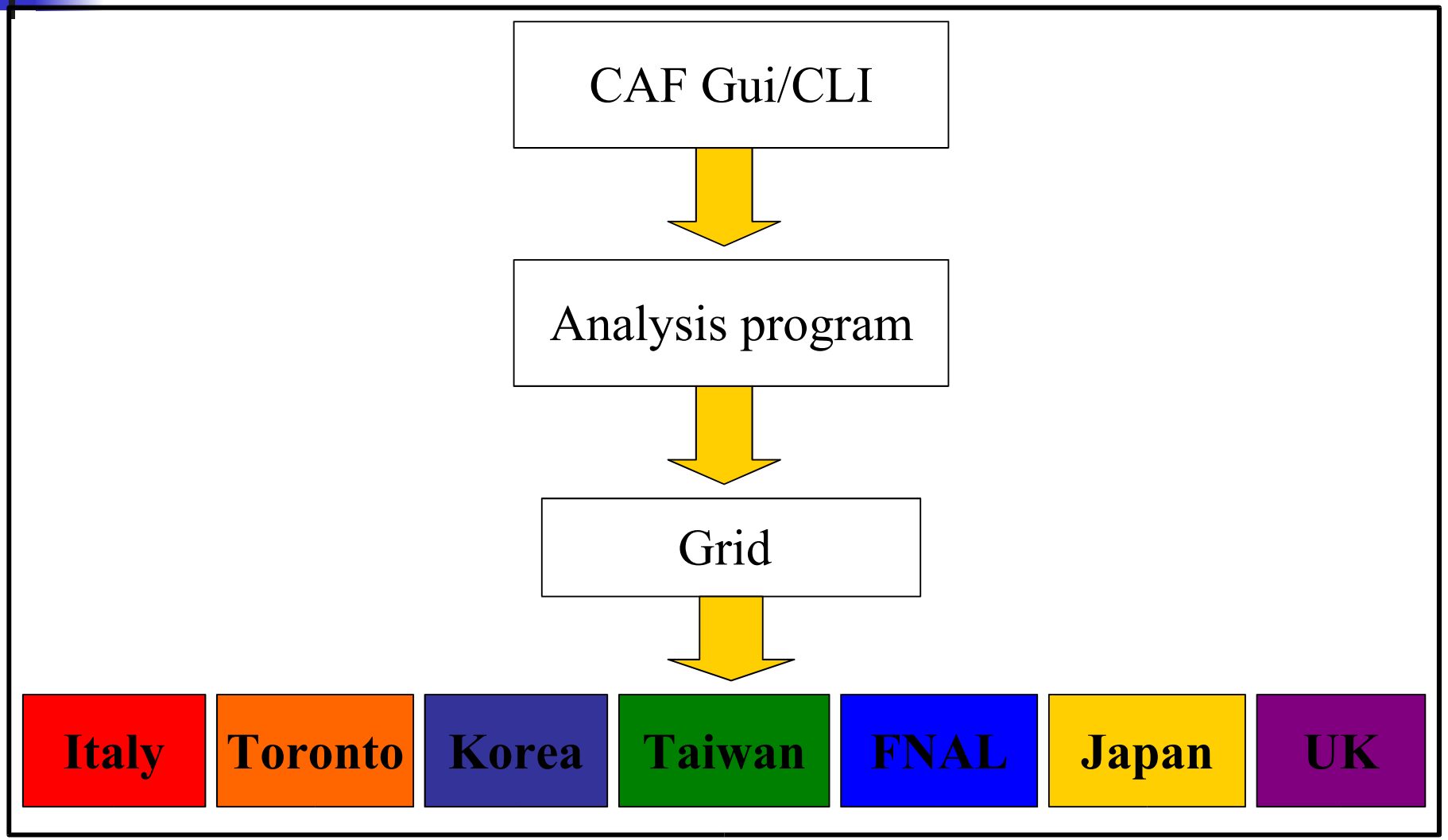
- Time between *Request Next File* and *Open File*
- For CAB and CABSRV1
  - 50% of enstore transfers occur within 10 minutes.
  - 75% within 20 minutes
  - 95% within 1 hour
- For CENTRAL-ANALYSIS and CLUED0
  - 95% of enstore transfers within 10 minutes

Station	CAB	CABSRV1	CLUED0	CA
% no wait	30%	40%	38%	18%

The Grid part of SAMGrid: JIM

- JIM components provide:
  - Job submission service via Globus Job Manager, augmented by some VO requirements
  - Job monitoring service from remote infrastructure
  - Authentication services

User Perspective: Task Submission & Execution



What can 20 duals and 6 TB do? (Example of Physics Datasets)

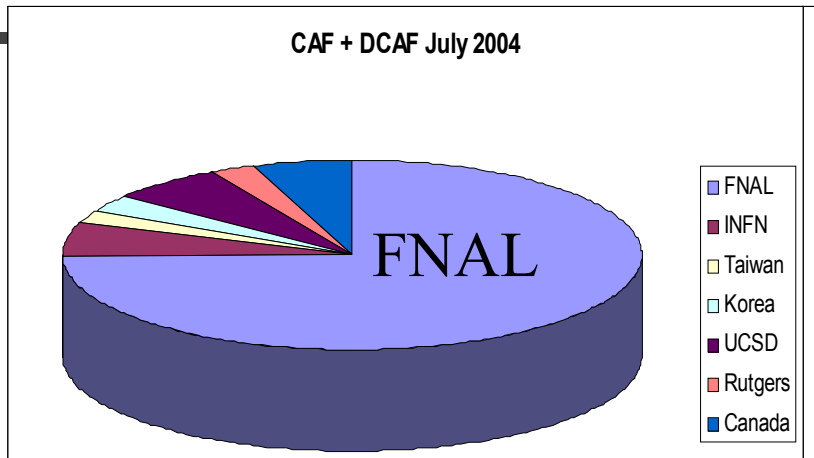
Stream	Events	Days	Input Size
Top,W/Z	20.5 M	10.3	4.5TB
Hadronic B and charm	156M	78.3	34.2TB

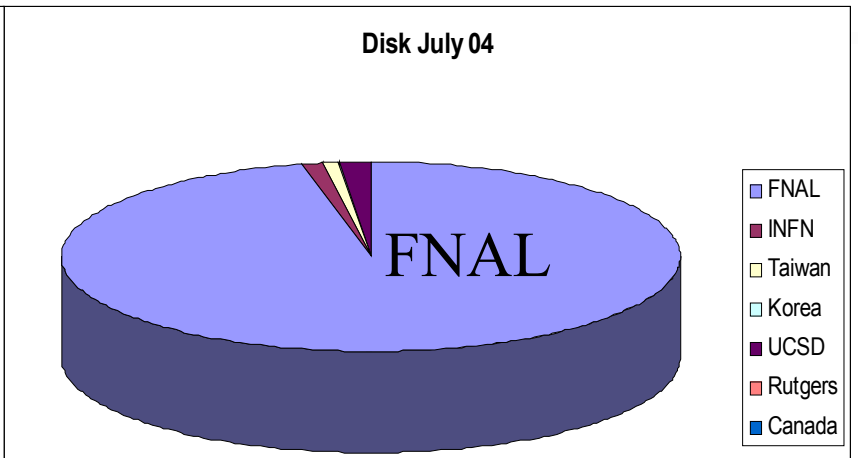
Need to transfer 0.6 GB/min or 1 TB/Day

CPU Growth OK, Disk Growth Slower: Need network and/or use offsite for MC

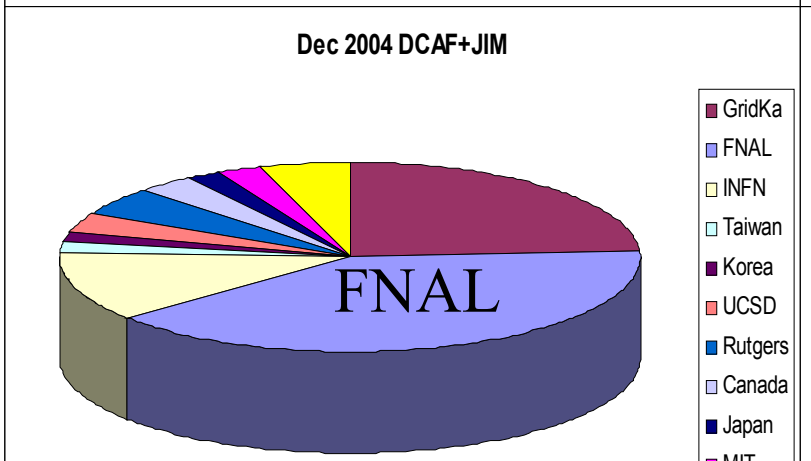
CPU **Disk**

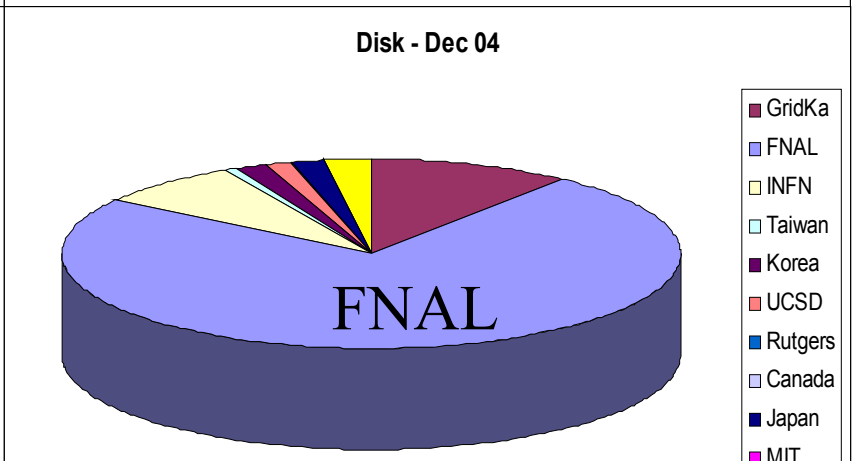
July 04





Dec 04

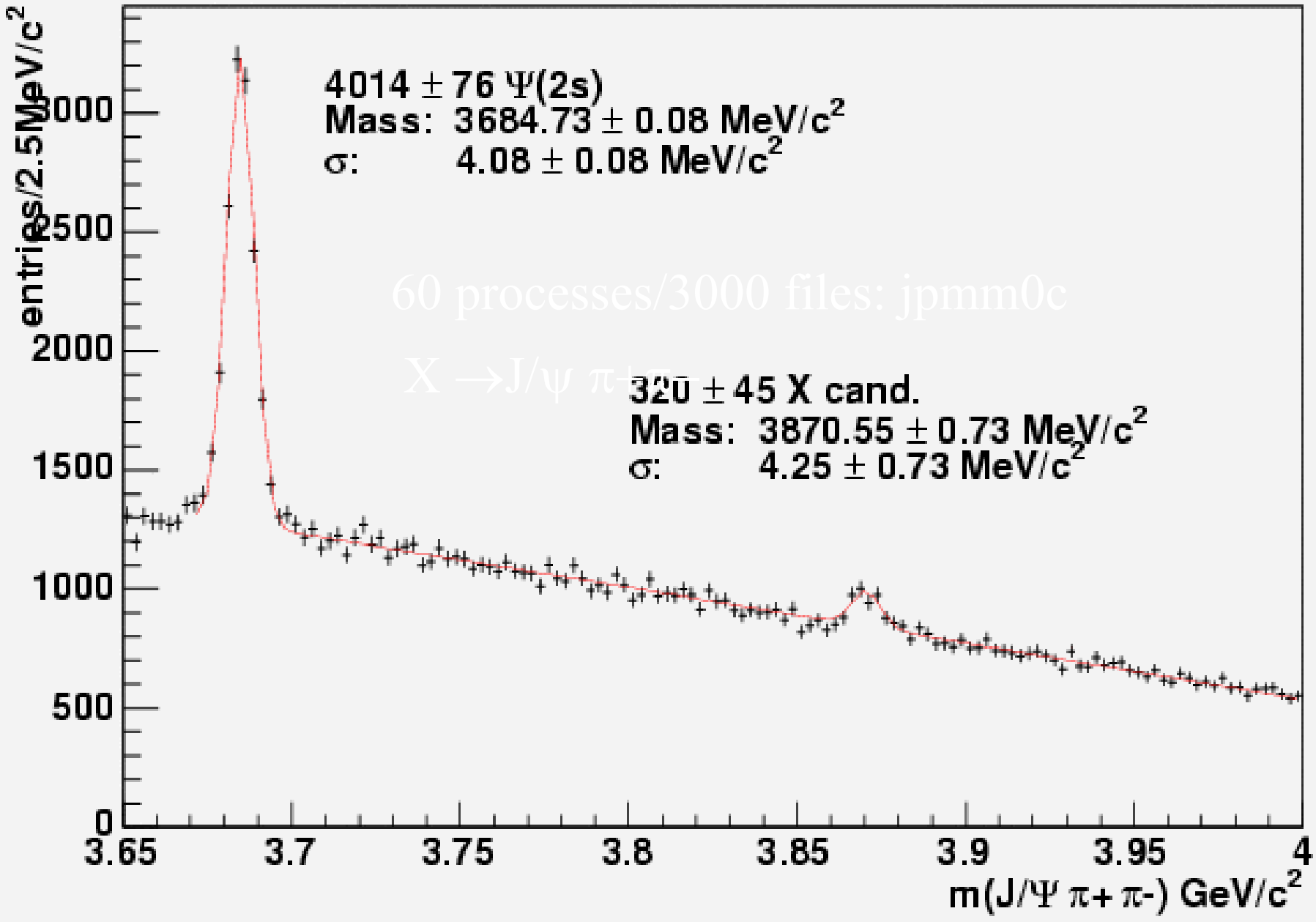




CPU from GridKa (Biggest present off-site SAM user)

Cluster not CDF-exclusive - Need Grid to make this resource available to full CDF collaboration!

- May 1-6: 650
- May 7-17:704
- May 18-27:604
- May 28-31:710
- May total 492,860 cpu hrs, 1THz roughly
- June 1-7: 740, 8-14 780, 15 power out, 16-30 700
- June total 507,360 cpuhrs, 1THz roughly

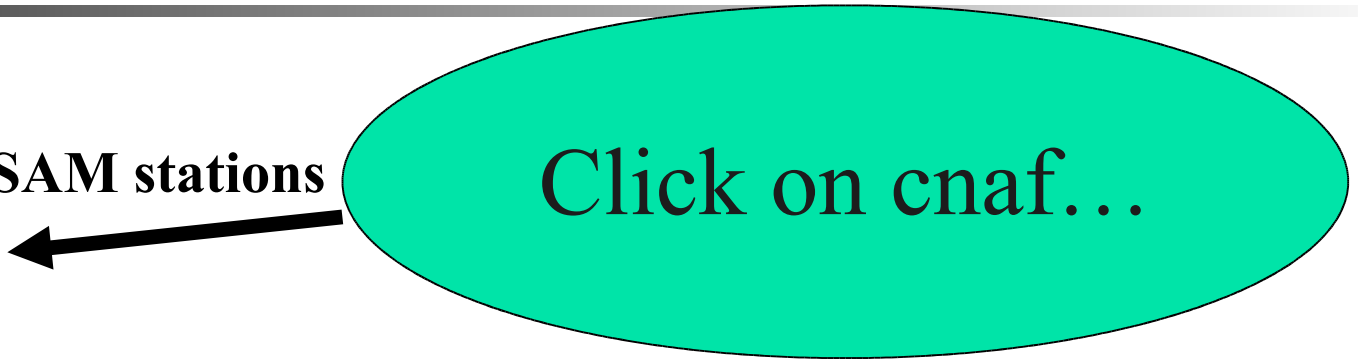

4014 ± 76 Ψ(2s)
Mass: 3684.73 ± 0.08 MeV/c²
σ: 4.08 ± 0.08 MeV/c²
60 processes/3000 files: jpmm0c
X →J/ψ π+π-
320 ± 45 X cand.
Mass: 3870.55 ± 0.73 MeV/c²
σ: 4.25 ± 0.73 MeV/c²
entries/2.5MeV/c²
m(J/Ψ π+ π-) GeV/c²

Screen Shot of Web page

http://hexfm1.rutgers.edu/DATA_INFO/sam_data/

CDF Datasets on SAM stations

- cdf-cnaf
- cdf-fzkka
- cdf-knu
- cdf-rutgers
- cdf-sdsc
- cdf-taiwan
- cdf-toronto
- cdf-ttu



Summer 2004 Goal: Expand Resources, More Efficient Operations

- ✓SAM on (D)CAFs
  - Reduce DH operations load: EMAIL/Fair Tape Share
- ✓Pin Datasets Remotely via SAM
- ✓MC Data Import:
  - Automate to reduce workload
  - Replace DFC with SAM
- 04 Goal was >25% offsite computing load
- Met this goal (35% of CDF collaboration-wide cpu capacity is now available offsite)

2004 Goals: Achievements So Far

- MC Data Import: will be in 5.3.4
- SAM on (D)CAF:
  - stress testing/fix bugs: need Beta Testers to do real analysis: used 20% of CAF reading golden Datasets (20TB/Day)
  - V6 schema adopted, product depoyment now underway
- Datasets Pinned and available
  - http://hexfm1.rutgers.edu/DATA_INFO/sam_data/
- DCAF utilization: few high-intensity users so far but no problems in principle
  - Provided useful cpu capacity for summer conferences
  - Now need next phase of data handling and grid submission

CDF Grid Strategy: Outlook and Goals

- Currently 35% of CDF collaboration-wide open computing capacity from external resources.
  - Utilizes only resources fully controlled by CDF so far: Kerberos/fbsng/CDF Condor dCAF
  - SAM used and available on ALL resources
- December 15, 2004: JIM/Grid3-OSG/LCG comparison ends (Mainly MC)
- By end of 2005: 50% of computing resources from external sources, broader use of Grid

Conclusions

- CDF making good progress toward providing increased off-site computing and DH capacity.
- Can capture many more resources using Grid to achieve physics mission.
- SAM is working now for CDF and will reduce operational loads, improve user experience.
- To make progress, add new software tools and move to capabilities like those supported for/by the LHC and other global grid efforts.