Metadata for the Common Physicist

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

- Goals of the Presentation
- Use Cases
- SAM in light of use cases
- SAM from 1 to 2, 2 to N – D0, CDF, MINOS, CMS
- Lessons from CDF merger
- Conclusions and Future

Goals

- Introduce:SAM Team, Metadata Working Group
- Describe the Many Faces of Metadata
- Examine metadata HEP Use Cases
- Greater understanding: Benefits of multiple experiment usage (sample)
- What SAM is and the SAM Schema
- Commonality with LHC expressed through use cases
- Support structure for migration: it can be done
- Keyword/Value pairs as a first step in common

The SAM-Grid Team and the Metadata Working Group

SAMGrid Project Co-Leaders: ***Wyatt Merritt, Rick St. Denis***

SAMGrid Technical Co-Leaders: Rob Kennedy, ***Sinisa Veseli***

SAMGrid Core Developers: Lauri Loebel Carpenter, Andrew Baranovski, Steve White, Carmenita Moore,* Adam Lyon, Petr Vokac,*** Mariano Zimmler***, Matt Leslie, **Lee Lueking, 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, Fedor Ratnikov, Alan Sill, Stefano Belforte,Ulrich Kerzel, Robert Illingworth

Database support: Anil Kumar, ***Julie Trumbo***

Metadata Working Group: ***Tony Doyle, Carmine Cioffi, Steven Hanlon, Caitriana Nicholson, Gavin Mccance, Solveig Albrand, Paul Millar, Tim Barrass***

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

Outline

- Goals of the Presentation
- Use Cases
- SAM in light of use cases
- SAM from 1 to 2, 2 to N – D0, CDF, MINOS, CMS
- Lessons from CDF merger
- Conclusions and Future

Use Cases Summary: HEPCAL,CDF,BABAR,ATLAS

3 Categories

Analysis

Job Handling

Dataset Handling

Analysis

Run a physics simulation

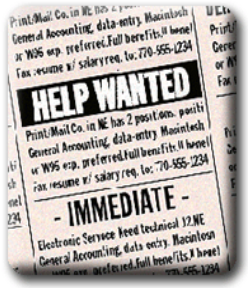
Select a subset of data

Run an algorithm over an input dataset

Ask for File

Analyze File

Output File



Job Handling

Estimate the system resource cost

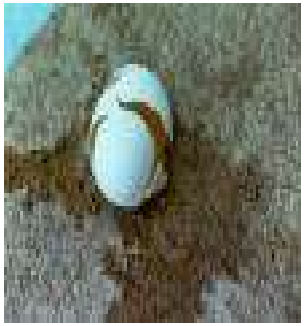
Monitor the progress of a job

Submit a job to a Grid

Retrieve/Access the output of a job

... with predefined metadata

Repeat a previous job



Recover failures in a previous job

Dataset Handling I

Read metadata for datasets

Update and/or Add metadata for datasets

Resolve physical data

Download a dataset to a local disk

Specify a new dataset

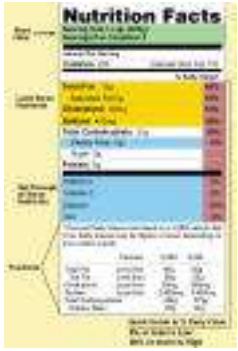
Access a Dataset

Predefine metadata for output dataset

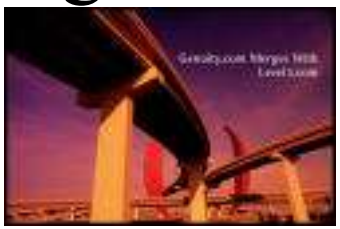
Write experiment-specific metadata for the new dataset

Dataset Handling II

Read all the visible metadata for a specified dataset



Merge dataset



Publish a private dataset

"Surely you were aware when you accepted the position, Professor, that it was publish or perish."

Perform a transform on a dataset

Search for datasets whose metadata match a user query

Publish private metadata

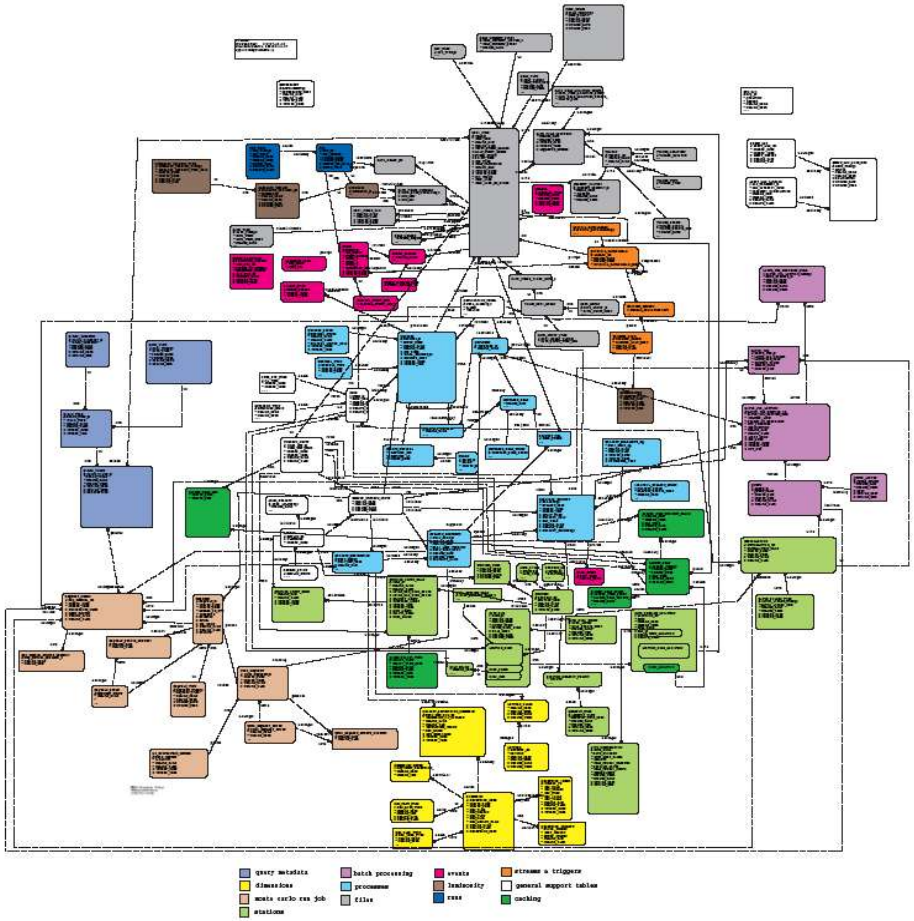
Outline

- Goals of the Presentation
- Use Cases
- SAM in light of use cases
- SAM from 1 to 2, 2 to N – D0, CDF, MINOS, CMS
- Lessons from CDF merger
- Conclusions and Future

The SAM Paradigm

- A project runs on a station and requests delivery of a dataset to one or more consumer processes associated with that station. Consumers perform a transformation on the dataset and output files to store with metadata. Services control optimal delivery and storage.
- File delivery is stateful and a permanent record of data handling is kept for a project.

Implemented on Relational Database



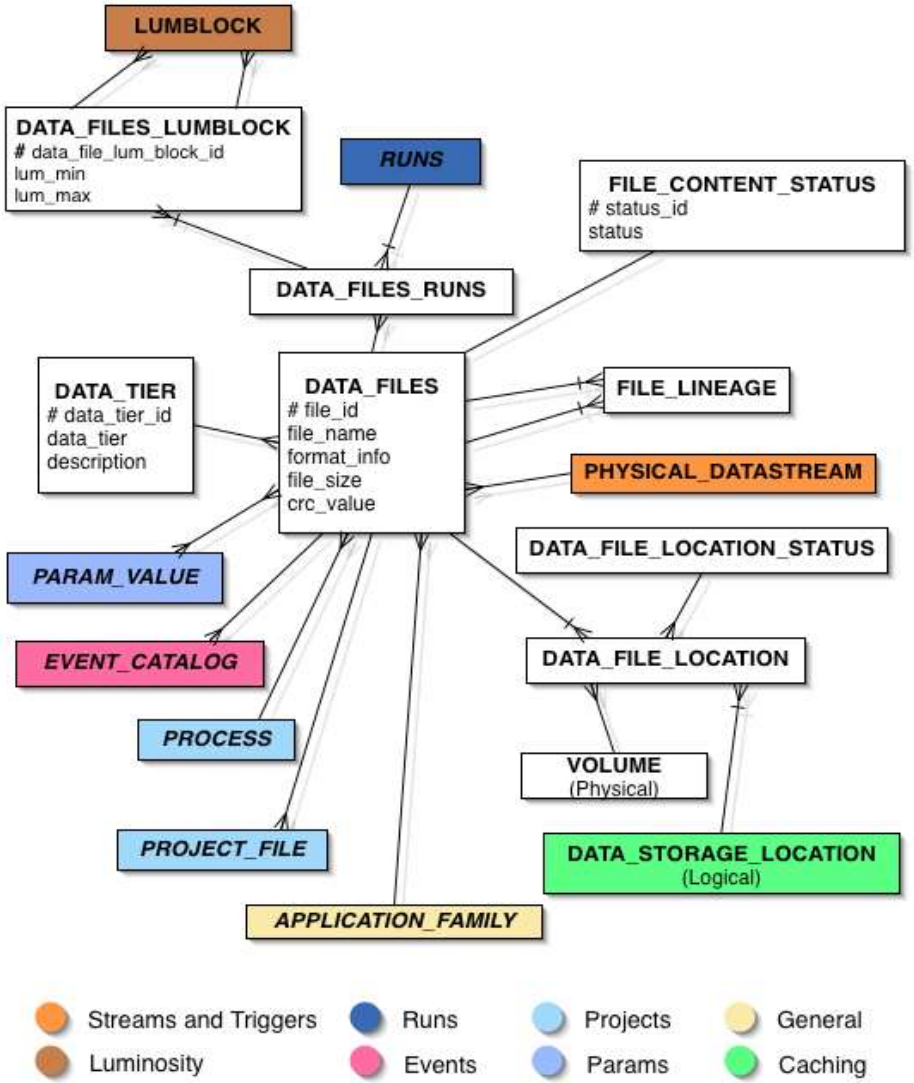
- DØ, CDF, and MINOS use the <u>same</u> DB Schema
- Relational
  - Matches metadata
- Monolithic
  - Efficient (>360 File/min)
- Flexible
  - Schema updateable *in a controlled fashion*

File Metadata

- SAM manages file storage (replica catalogs)

  - Data files are stored in tape systems at FNAL and elsewhere around the world for fast access
- SAM manages file meta-data cataloging
  - SAM DB holds meta-data for each file.

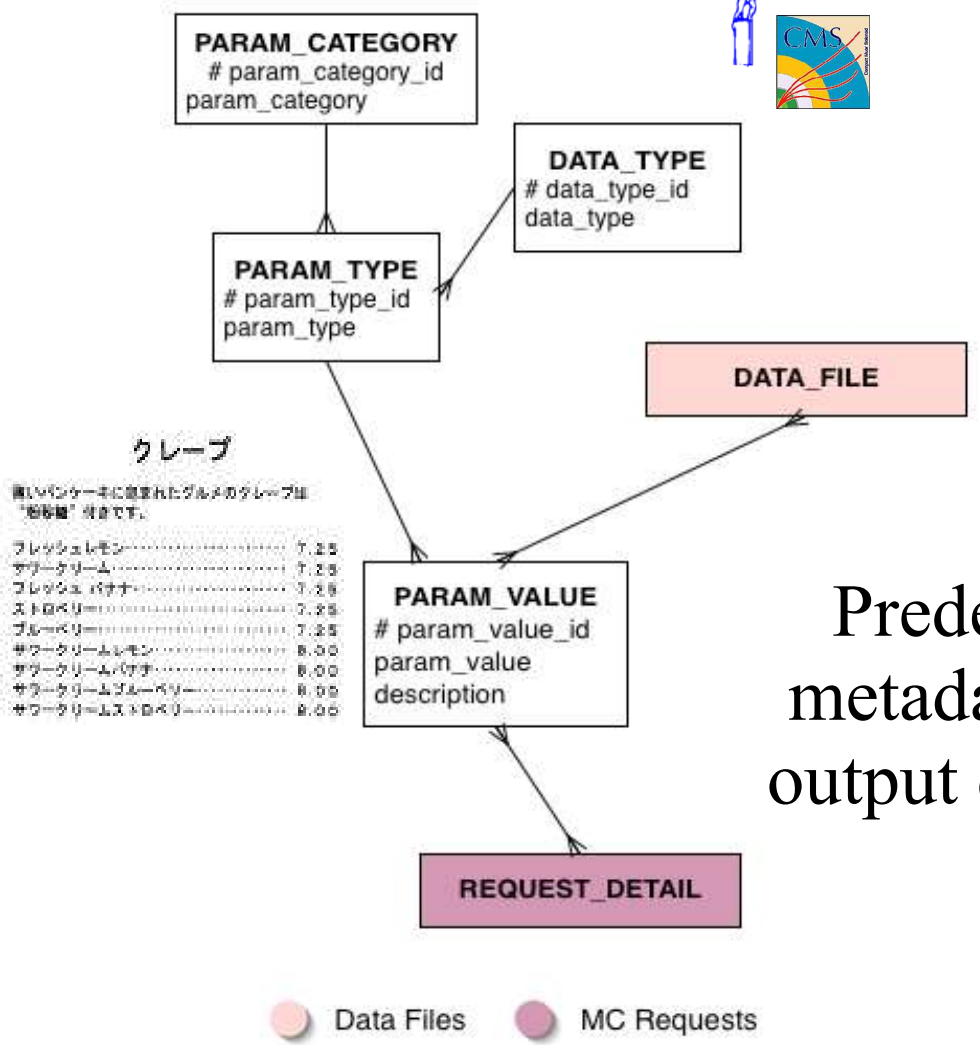
Data Files Metadata



- Data Files: The heart of SamGrid

- Fixed metadata
  - File name, size, crc
  - Production group
  - Data Tier (Raw, Reconstructed …)
  - Application, Locations
  - Detector, Runs,Event info
  - Project/Process, Luminosity
  - Stream/Trigger
- Connection to free metadata (Params) …

Params (Free file metadata):A common element with ATLAS, LHCB

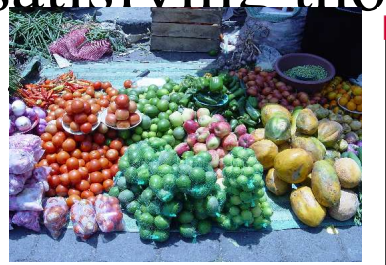


Predefine metadata for output dataset

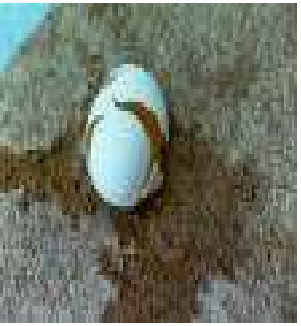
- Fixed metadata allows easy and performant querying
- Free metadata for application specific items
  - Categories group parameters (pythia, isajet, …)
  - Types are the keywords (decayfile, topmass, …)
  - Values
  - Queries are more difficult

Metadata Definitions

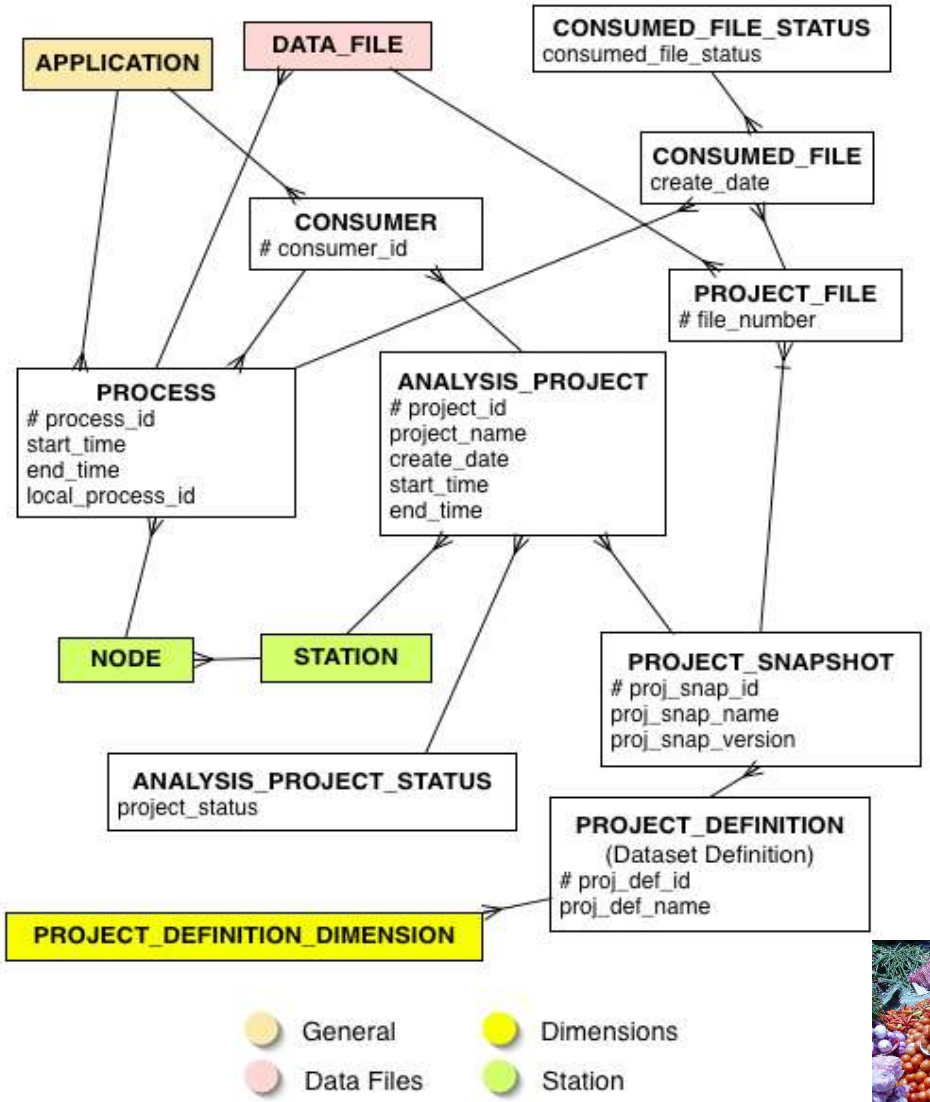
- SAM manages definitions of datasets based on metadata
  - SAM DB stores definitions based on metadata by group and user. These are resolved to lists of files satisfying those definitions when a user chooses to run a



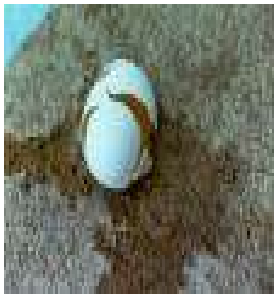
  - "data_type physics and run_number 78904"
- SAM manages analysis bookkeeping
  - SAM remembers what files you ran over, what files you processed successfully, what applications you ran, when you ran them and where. Hence it is possible to recover from errors and repeat runs.



Project Metadata



- Projects run by a user in a group on a dataset Snapshot with nodes from a SAMGrid station
- A Project has one or more Consumers (usually one)
- A Consumer has one or more Processes
- A Process is a job on a node. Keeps track of consumed files

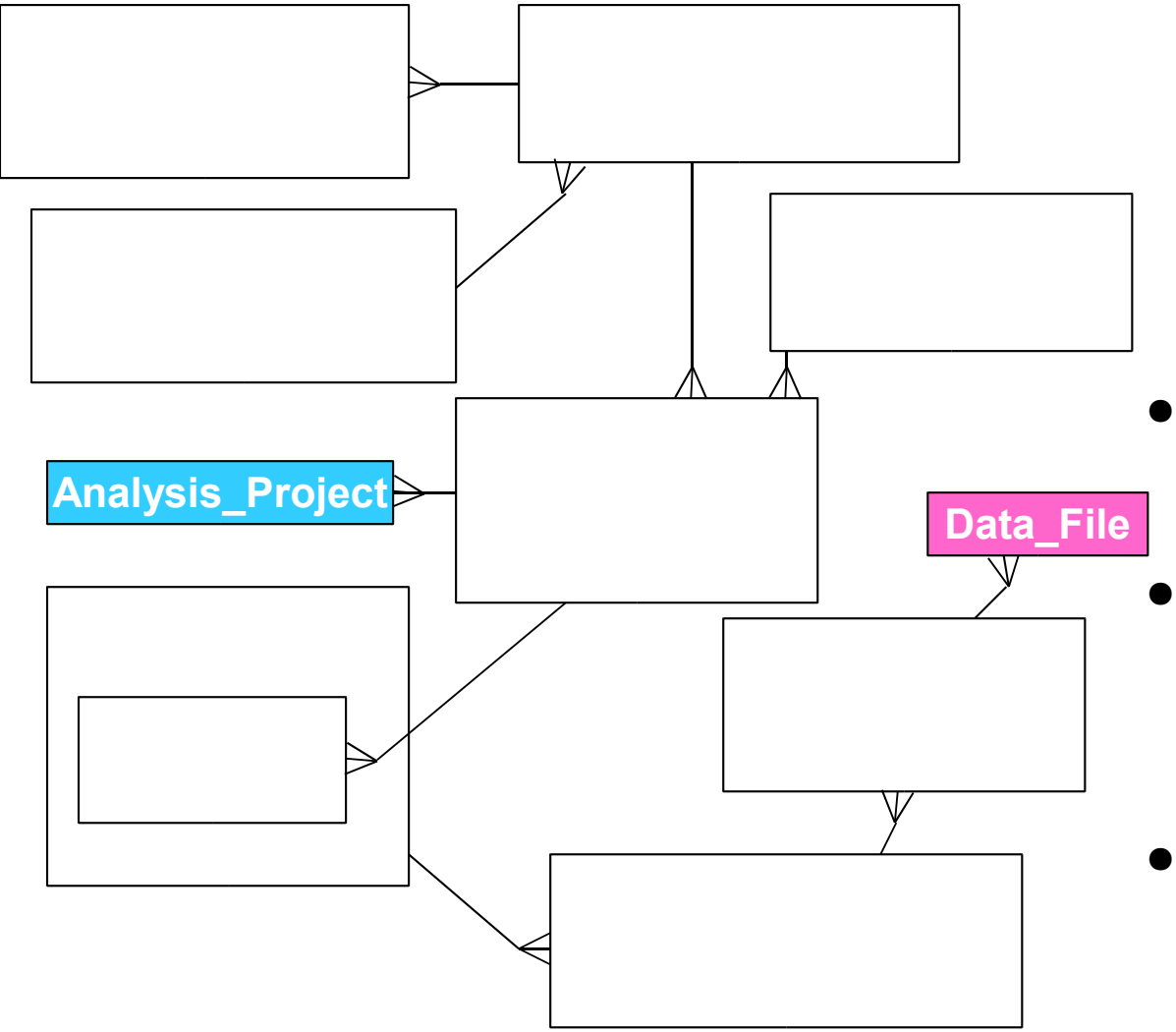


File Delivery

- SAM manages file delivery by dataset
  - Users at FNAL and remote sites retrieve files out of file storage. SAM handles caching or can interface to other cache systems (See Rob Kennedy's Talk)

  - You don't care about file locations

File Delivery:Station and Cache



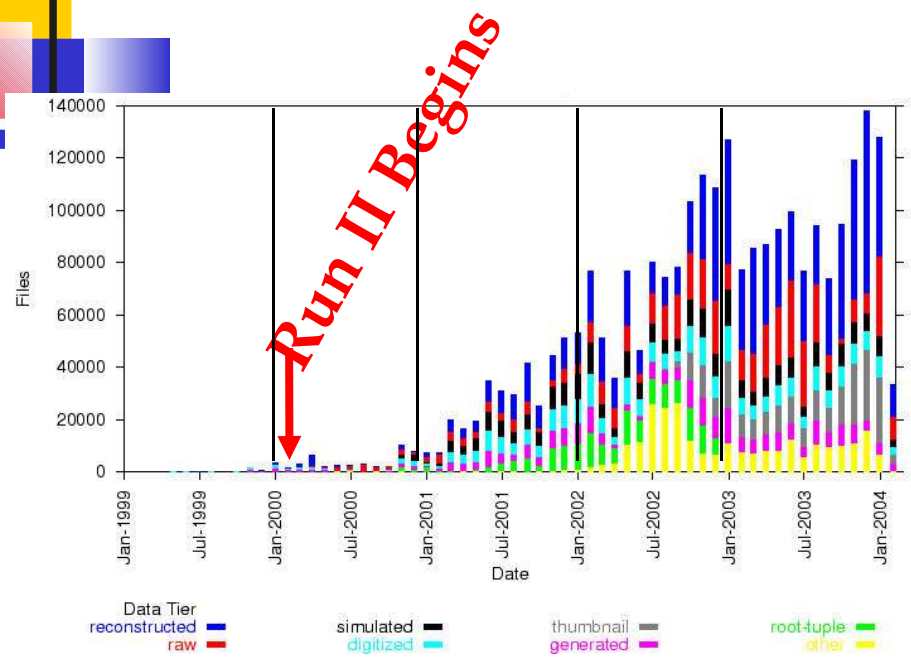
- The project master, a services to coordinate delivery of files from a storage element, runs on a station.
- A station uses CORBA for communication
- The station keeps track of the files it has been requested to send.
- The station may manage a cache or dispatch URL's to a cache

Outline

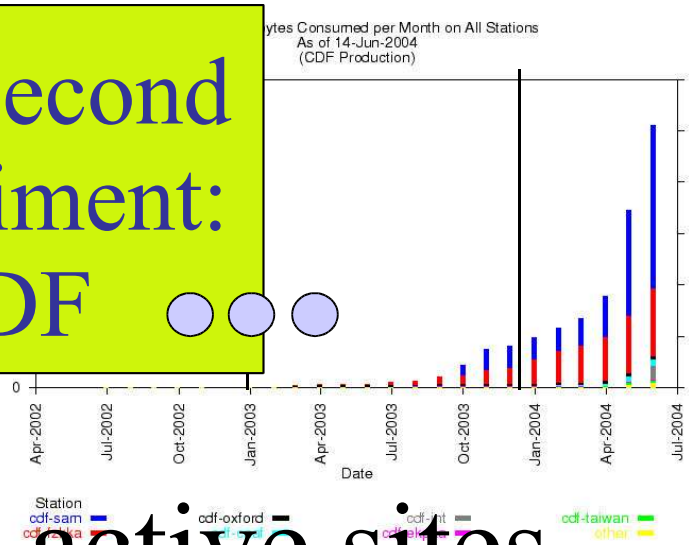
- Goals of the Presentation
- Use Cases
- SAM in light of use cases
- SAM from 1 to 2, 2 to N – D0, CDF, MINOS, CMS
- Lessons from CDF merger
- Conclusions and Future

SAM: from One Experiment: DØ

40 active sites



To a second experiment: CDF

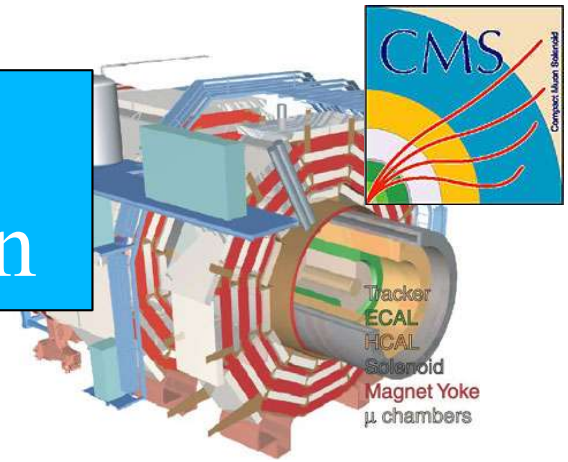


25 active sites

To MINOS

2 sites @fnal

CMS Evaluation



Outline

- Goals of the Presentation
- Use Cases
- SAM in light of use cases
- SAM from 1 to 2, 2 to N – D0, CDF, MINOS, CMS
- Lessons from CDF merger
- Conclusions and Future

Database and Schema Evolution

- Centralized Oracle Database at FNAL
- Three tier system ensures DB integrity
  - Development - Newest schema with artificial or special data. Used for testing
  - Integration – Dress rehearsal for modifying schema using a copy production data upon which a test harnesss is run.
  - Production - The real thing

Overview of Impact of CDF Involvement

- CDF participation drove a revisiting of the D0 Design: Requirements became clearer as the experiment went through different phases: commissioning, MC, stable running.
- An entirely new user community provided the trigger for a second generation design, the need for which was recognized by the original users.
- Boundries became more clearly defined and natural separation into services occurred.

Impact of CDF: A Shopping List

- Many runs in a file; **separate luminosity bookkeeping**
- Clean separation of **file types**: Generic, MC, processed
- Keep track of group responsible for file
- Require at DB Level: format, size, crc type/value, file content status id
- Not Required at DB Level: data tier, file partition, **process id**, stream, event count, first/last event number start/end times
- Removed: MC - min bias no. & type, physics process

Three Examples: Deeper Implications

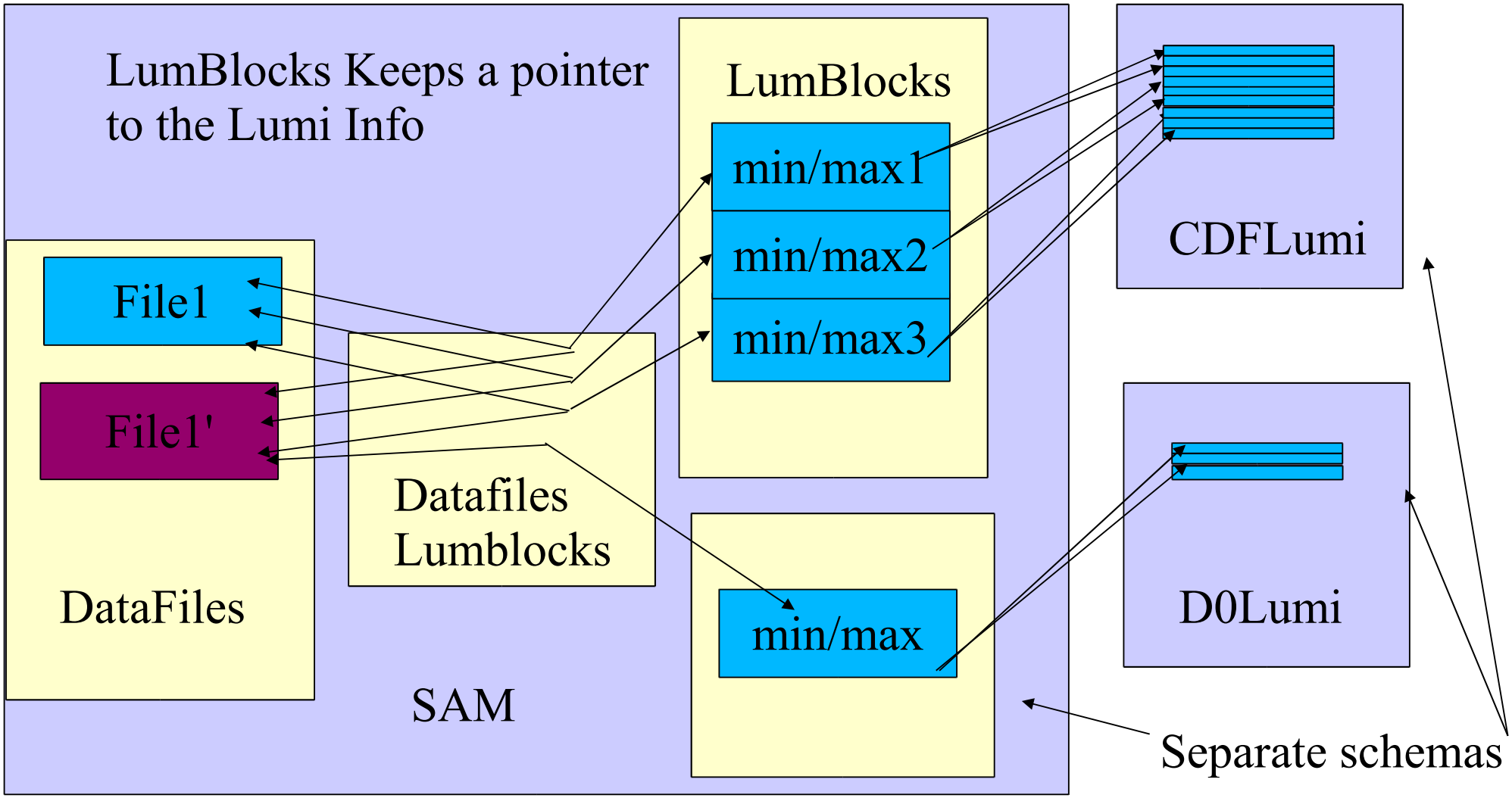
- Process ID:
  - Change in Paradigm
- Separate Luminosity bookkeeping:
  - Illustration of how to link different database schemas
- File Type:
  - Change in location of business rules

Process ID

- Sam Assumes
  - A process produces a file.
  - You ALWAYS want a process for a file
  - Therefore ProcessID is required
- Reality says
  - Sometimes files are imported from users not running with SAM to get input and keep track of files

The Process ID cannot be required

Linking Schema: Luminosity Bookkeeping



File Type: Change of location of business rules - Implement Rules in API

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

Conclusions

- Metadata: Workflow Processing, File/physics, Authorization, Quota
- Greater understanding: Experimental Lifecycle maturation, need for sharp boundaries, natural demarcation of services when experiments join: benefits to both.
- SAM is a system of data handling and work flow services described by metadata modelled on a relational database
- SAM implements the HEPCAL Metadata use cases.
- Migration of schema with running experiments is inevitable and can be accomplished
- Detailed schema and API implementations can be shared across HEP experiments.

Extra Slides

Interfacing

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

Valid Data Groups:Workflow-Data handling interaction

Ask for File
Analyze File
Output File

Merge dataset

- Workflow Step Transition
- File operations atomic
- Metadata for workflow
- Born of CDF/D0 Joint Effort

Perform a transform on a dataset

To Be Processed	Processed	
InFile 1	InFile 1	OutFile 1
InFile 2	InFile 2	
InFile 3	InFile 3	
InFile 4		OutFile 2
InFile N		OutFile M

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

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

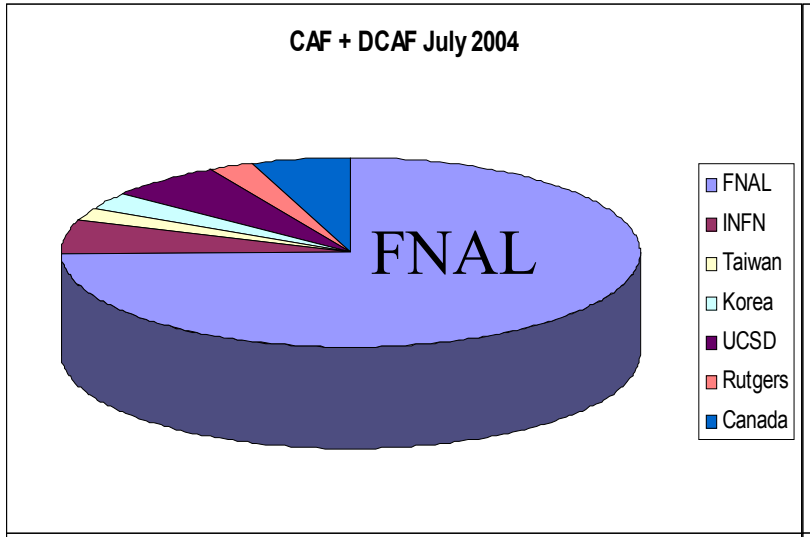
Monitor Level: High	Host	Version	Up Since
cdf-cnaf	cdfsam.cnaf.infn.it	v4_2_1_63	22 May 2004 07:19:01
cdf-fzkka	cdf.fzk.de	v4_2_1_72	23 Jul 2004 10:58:27
cdf-knu	cluster67.knu.ac.kr	v4_2_1_72	13 Jul 2004 03:38:41
cdf-oxford	matrix.physics.ox.ac.uk	v4_2_1_71	20 Jul 2004 11:34:23
cdf-rutgers	hexsam.rutgers.edu	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

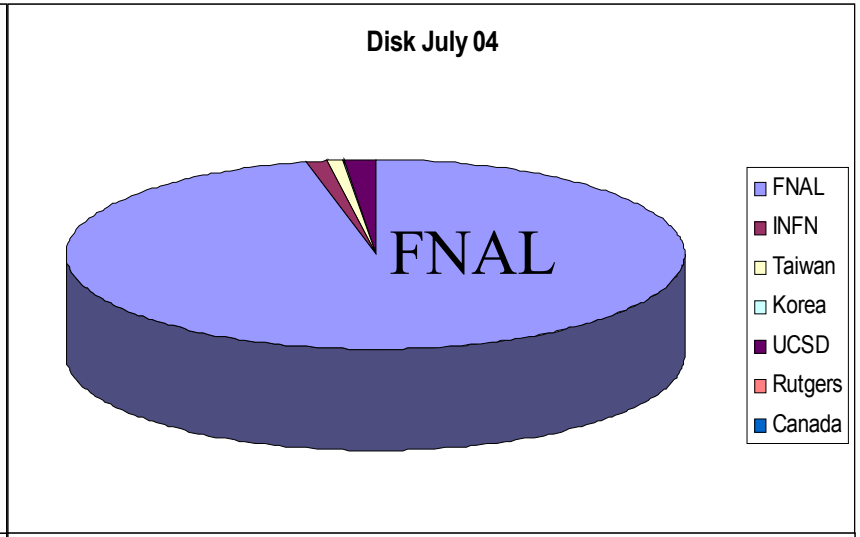
cdf-ral	cdfa.rl.ac.uk	v4_2_1_63	08 Jul 2004 09:34:33
cdf-rdk-fnal-1			
cdf-sam2			
cdf-scotgrid			
cdf-scotgrid-2			
cdf-taiwan2			
cdf-test	cdfpcb.fnal.gov	v4_2_1_72	20 Jul 2004 01:01:02
cdf-ttu-hpcc			
cdf-ttu-phys	castor.phys.ttu.edu	v4_2_1_72	19 Jul 2004 22:36:10
cdf-tufts			
cdf-ucsd			
samadams	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

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

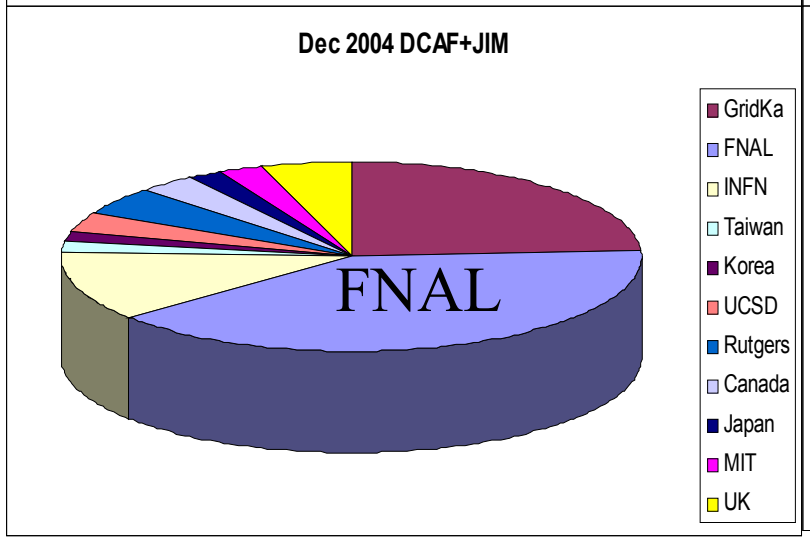
CPU **Disk**

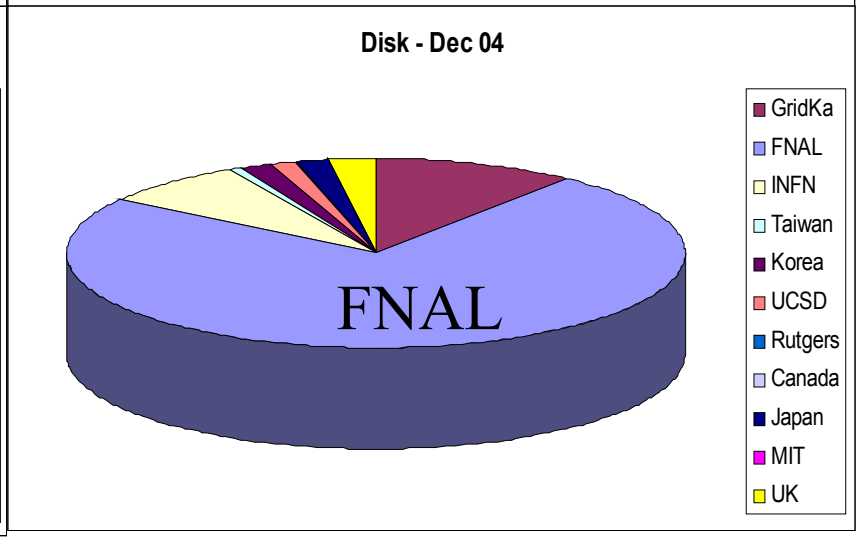
July 04





Dec 04





See http://cdfkits.fnal.gov/DIST/doc/DCAF/

CDF Global
Task Submission & Execution

Run a physics simulation

Select a subset of data

Run an algorithm over an input dataset

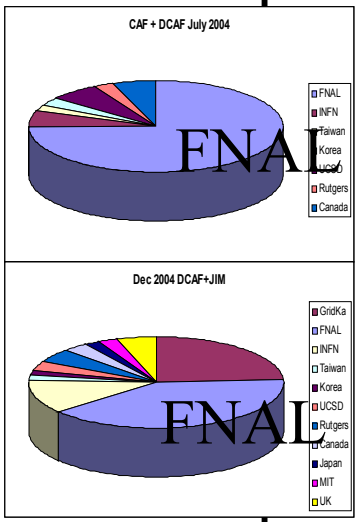
DCAF Gui/CLI

Analysis program

DCAF

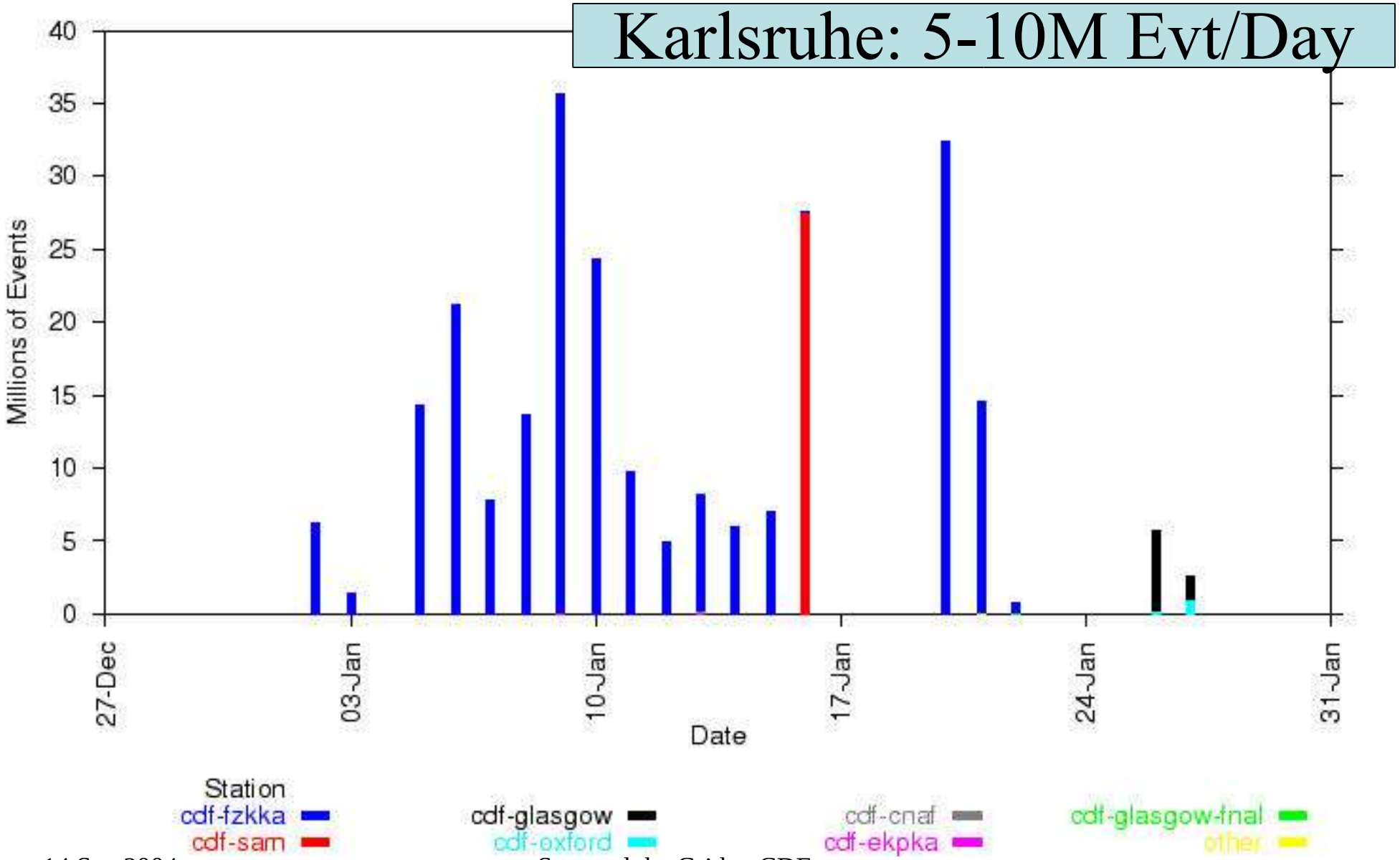
DCAF: 200GHz farm

Sam services on head node

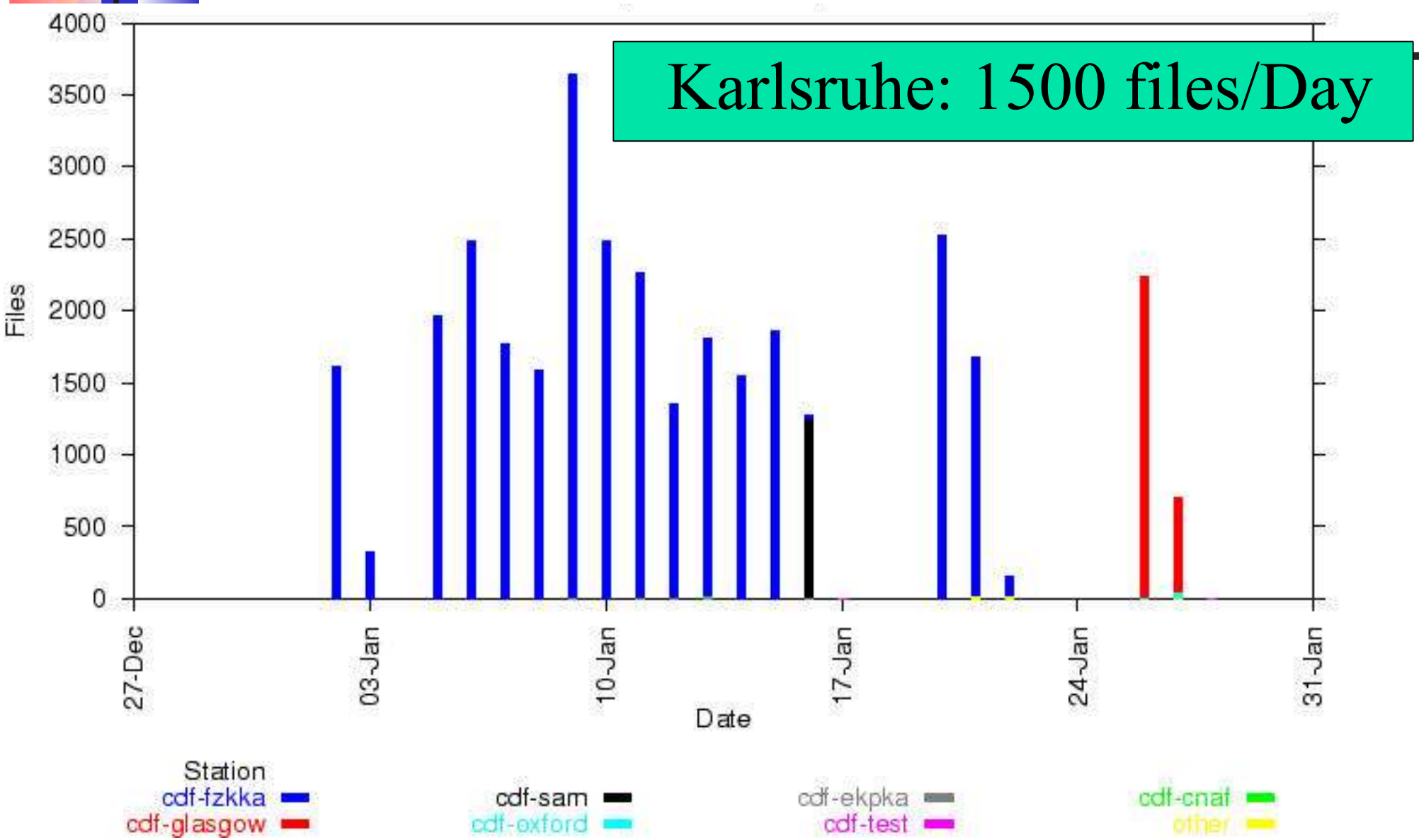


Italy	Toronto	Korea	Taiwan	FNAL	Japan	UK

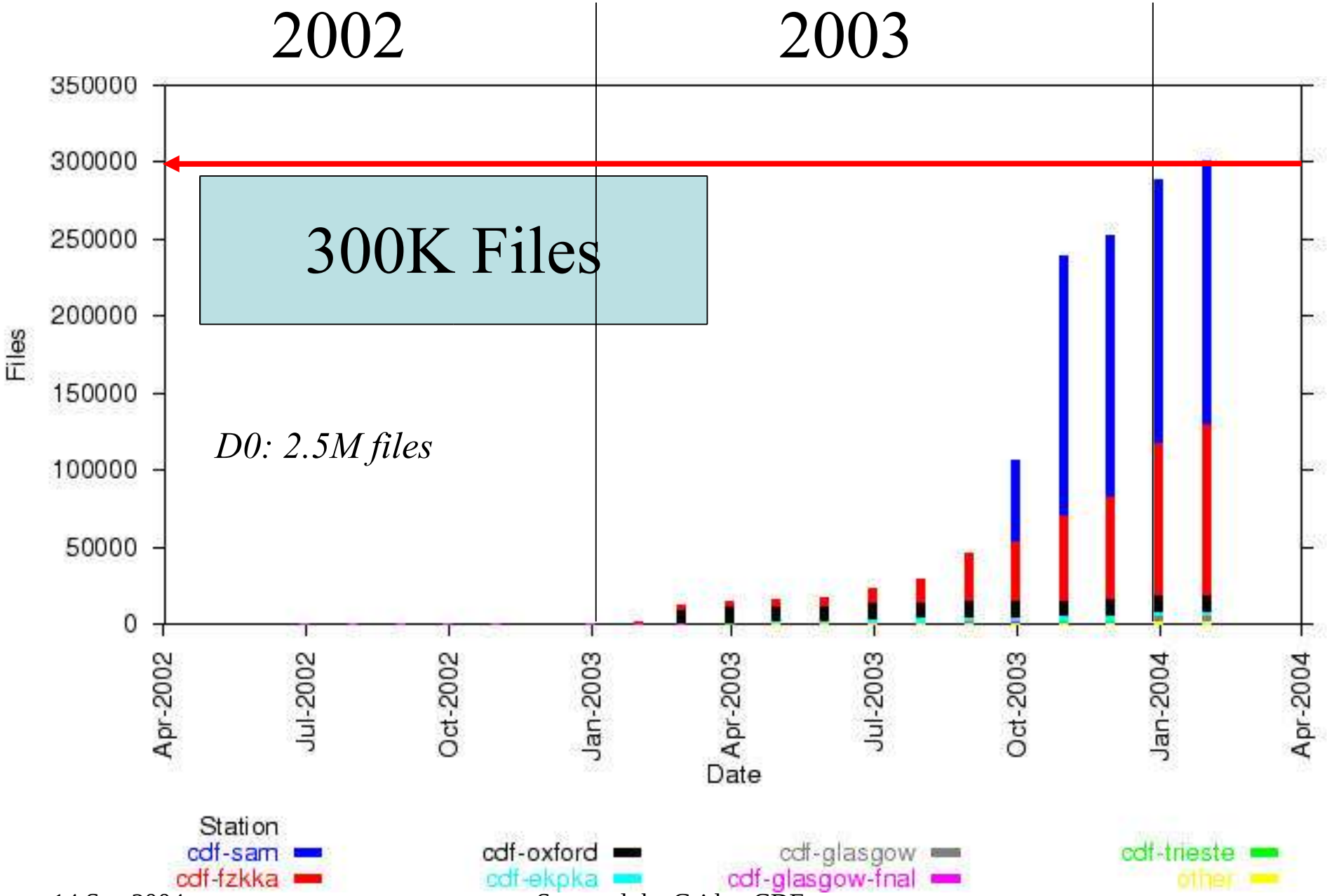
CDF Events Transferred per Month



CDF Files in a Month



All CDF Files Moved by SAM

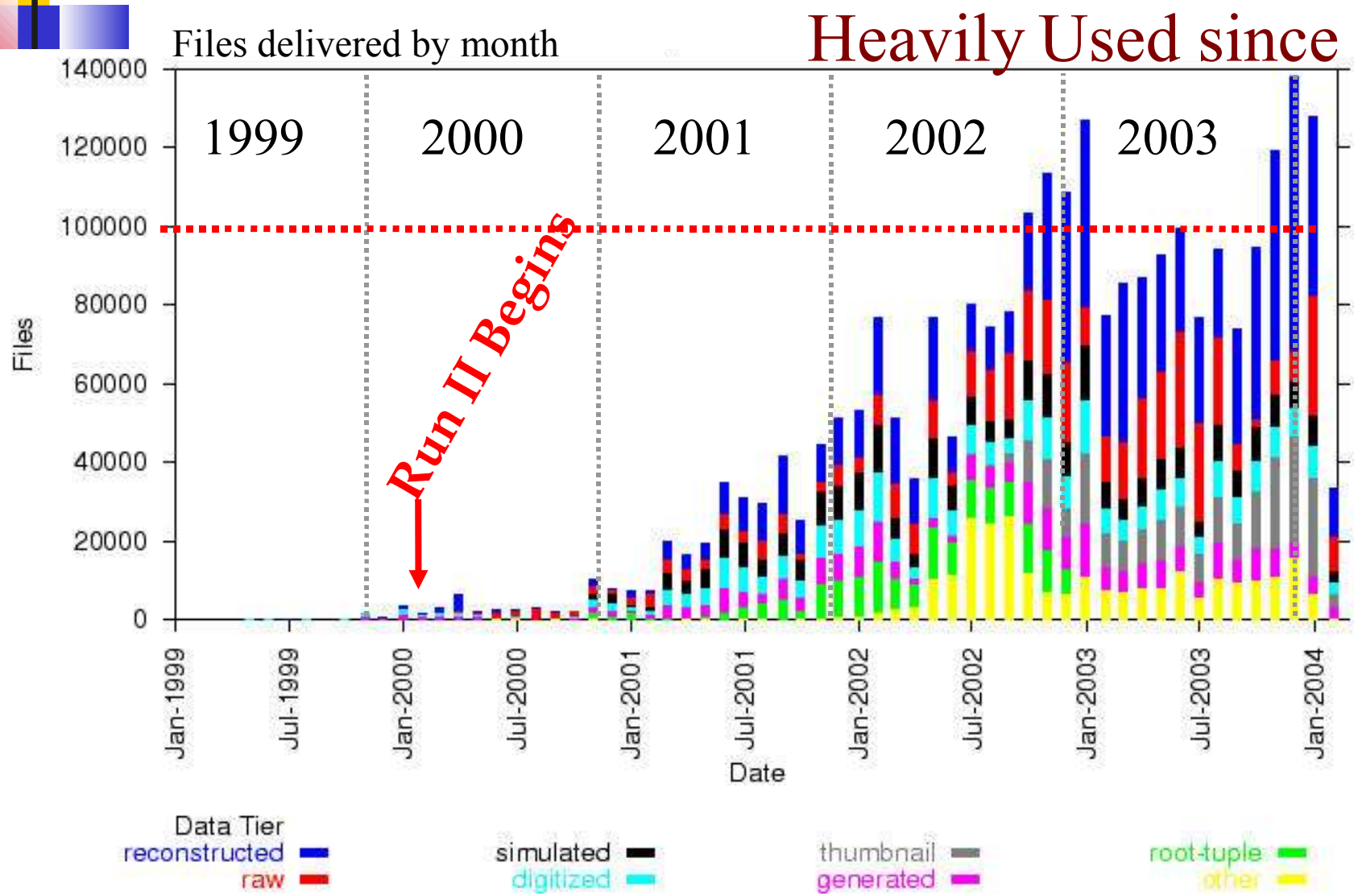


Q: What is SAM?
A:Data handling system for Run II DØ, CDF and MINOS

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

Designed for PETABYTE (10^{15}) sized experiment datasets

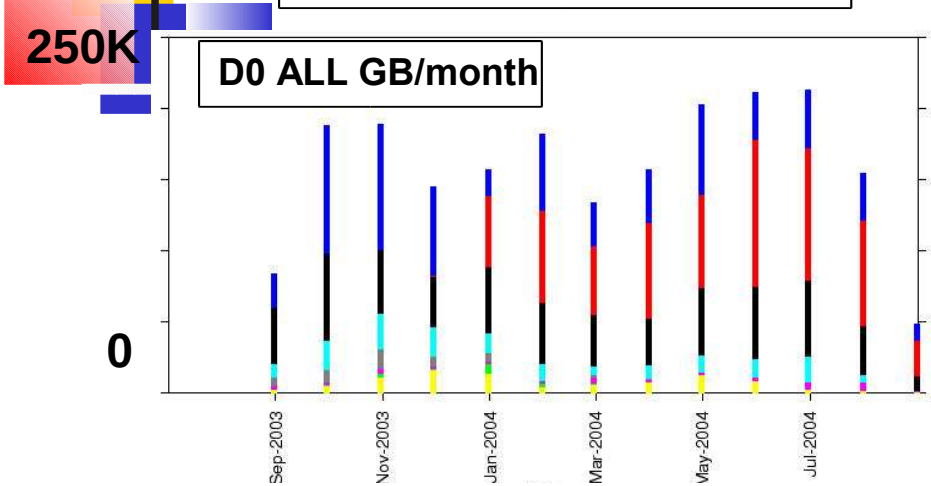
SAM goes from One Experiment: DØ

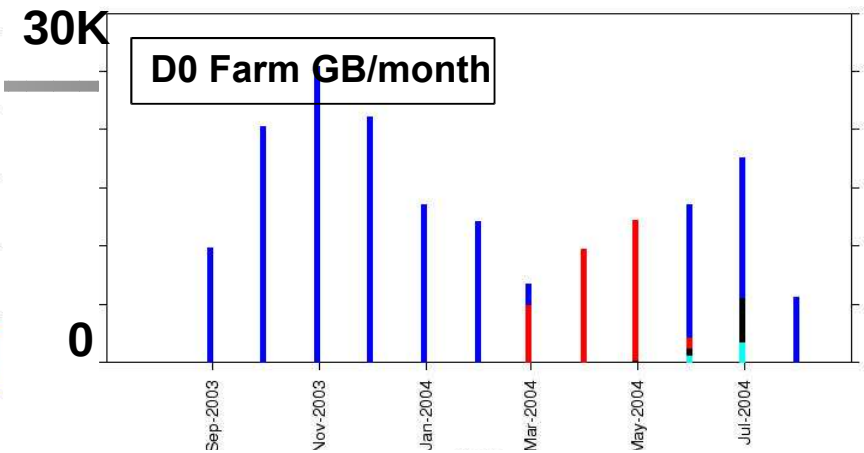


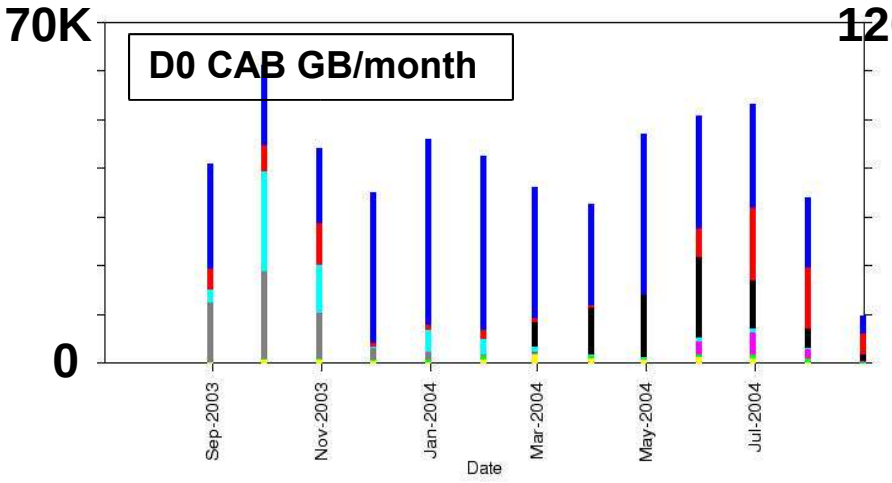
DØ -40 active sites, 9@FNAL

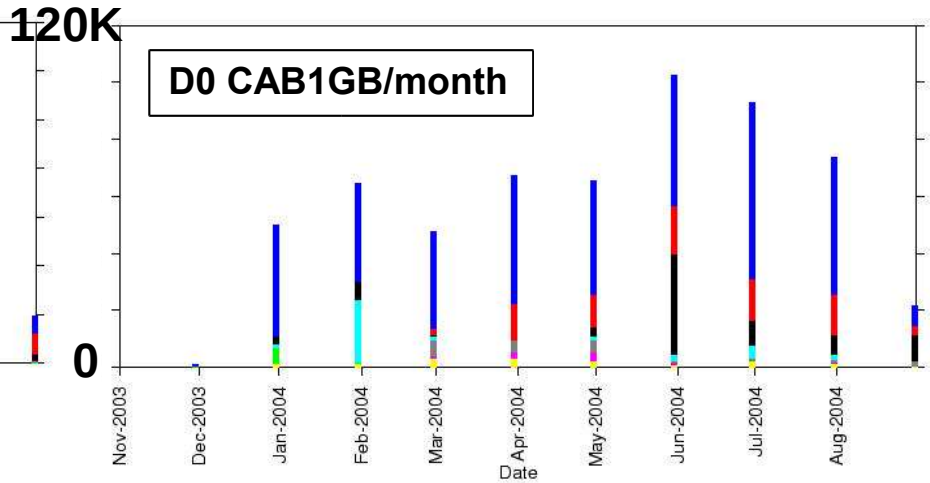
Usage Statistics for D0 SAM

Sum = 2.1 PB; 50B evts



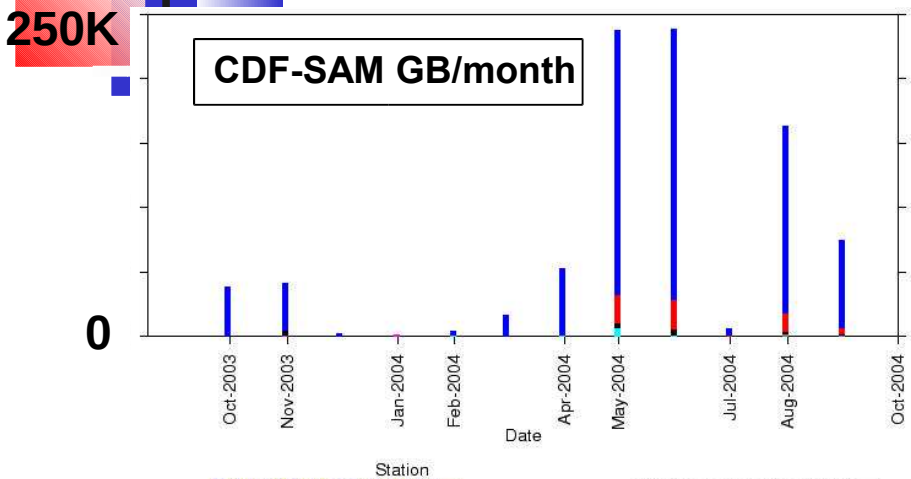


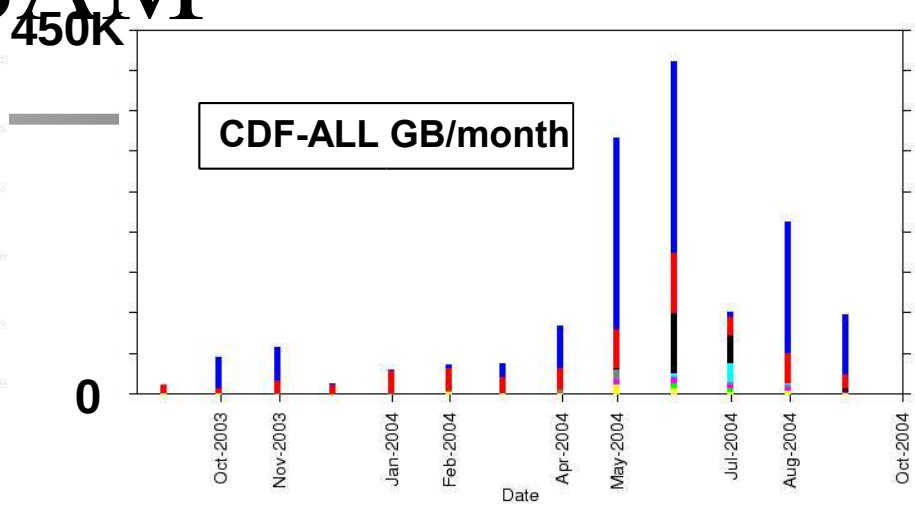


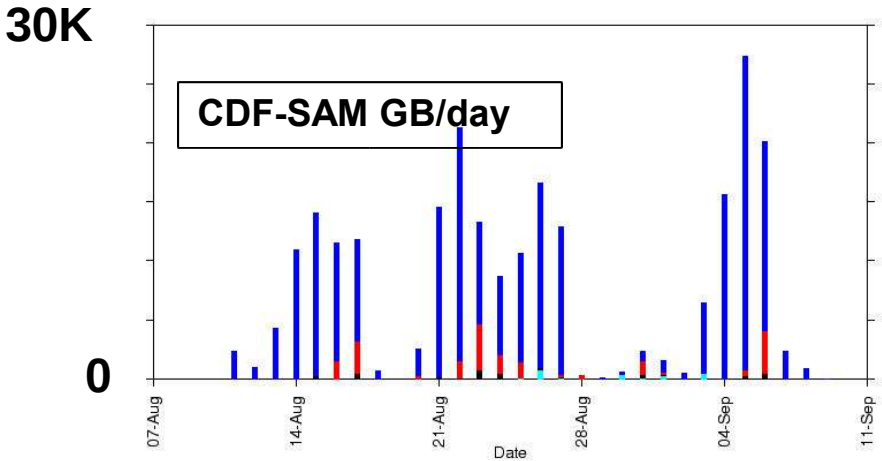


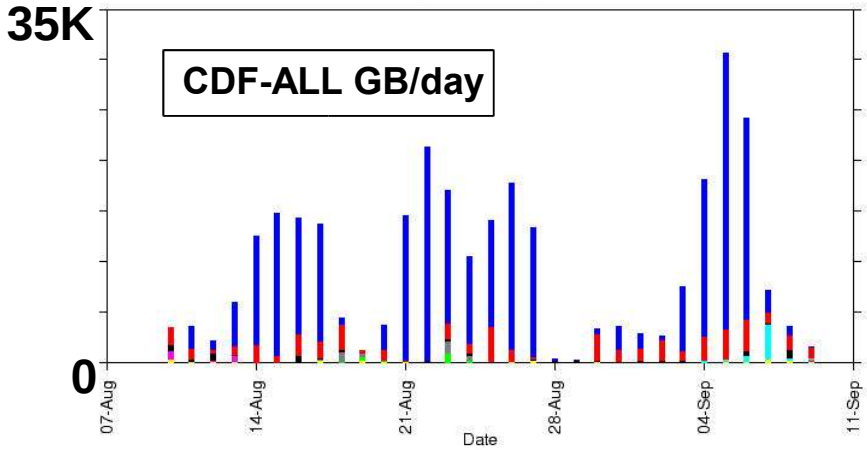
Usage Statistics for CDF SAM

Sum = 1.5 PB; 12B evts





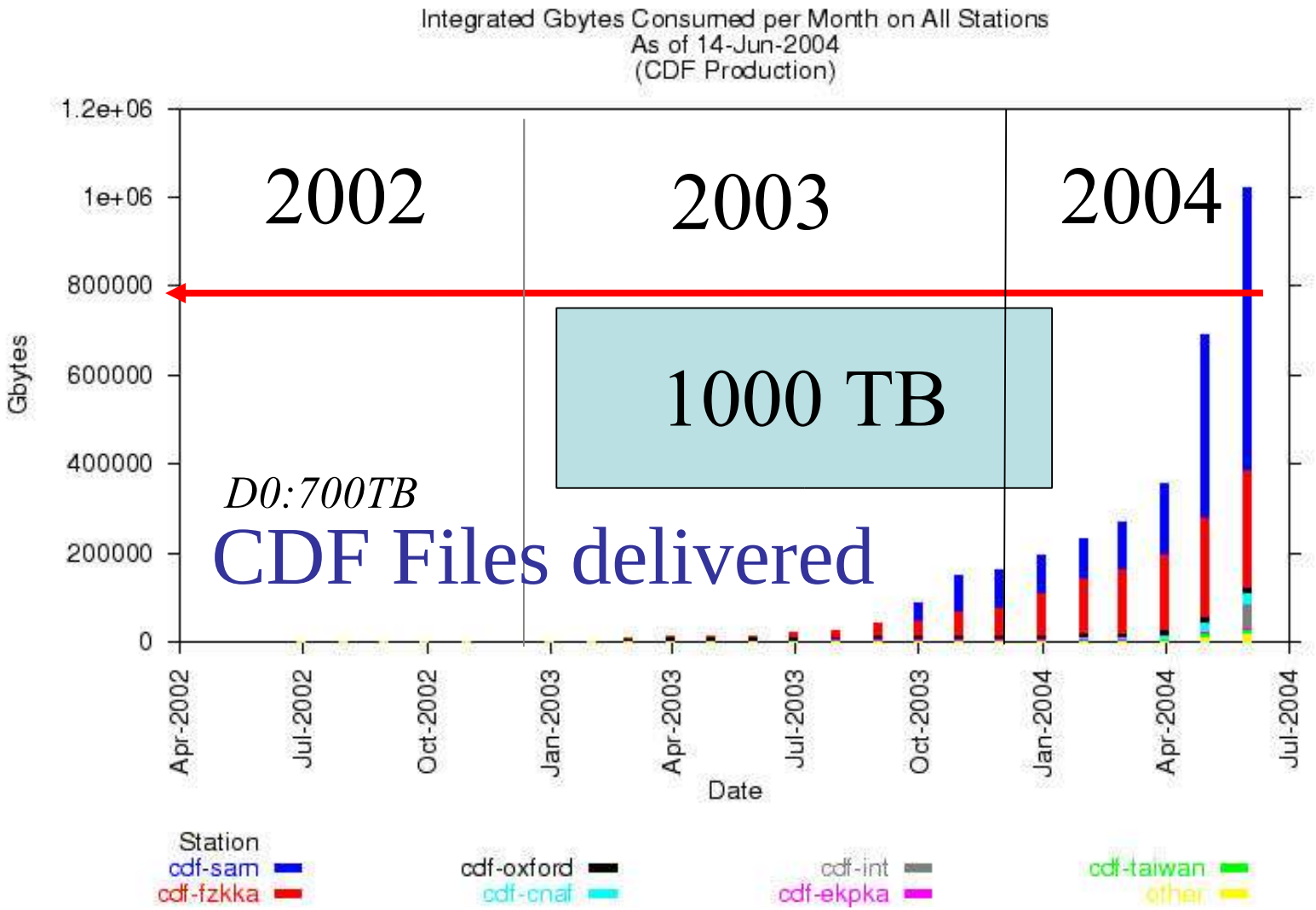




Current Resources			
Cluster Name and Home Page	Monitoring and Direct Information Links	CPU (GHz)	Disk space (TBytes)
Original FNAL CAF	queues, user history, ganglia, sam station, consumption	1200	200
FNAL CondorCAF (Fermilab)	queues, user history, analyze, ganglia, sam station, consumption	2000	~(shared w/CAF)
CNAFCAF (Bologna, Italy)	qu		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

To a second experiment: CDF

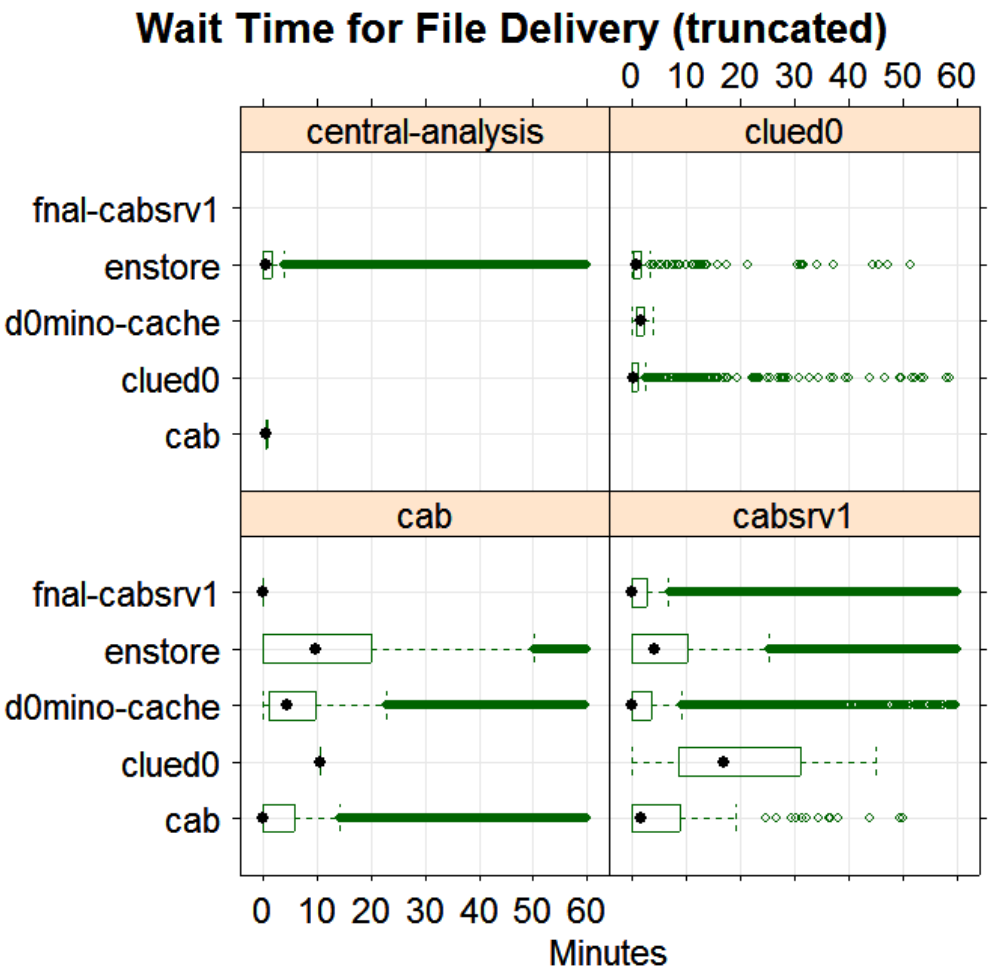


Sam Deployed Later to CDF: 25 active sites (2 @ FNAL)

SAM Terms

- **Station**: Permanent and transient services that monitor file consumption and make requests to storage resources for more files.
- **Project:** Delivers files to processes and keeps permanent record: sam get project summary
- **Dataset**: "data_type physics and run_number 78904"
- **Consumer**: User application that consumes and produces data (one or many exe instances)
  Examples: script to copy files; reconstruction job

SAM Statistics - Operations Data



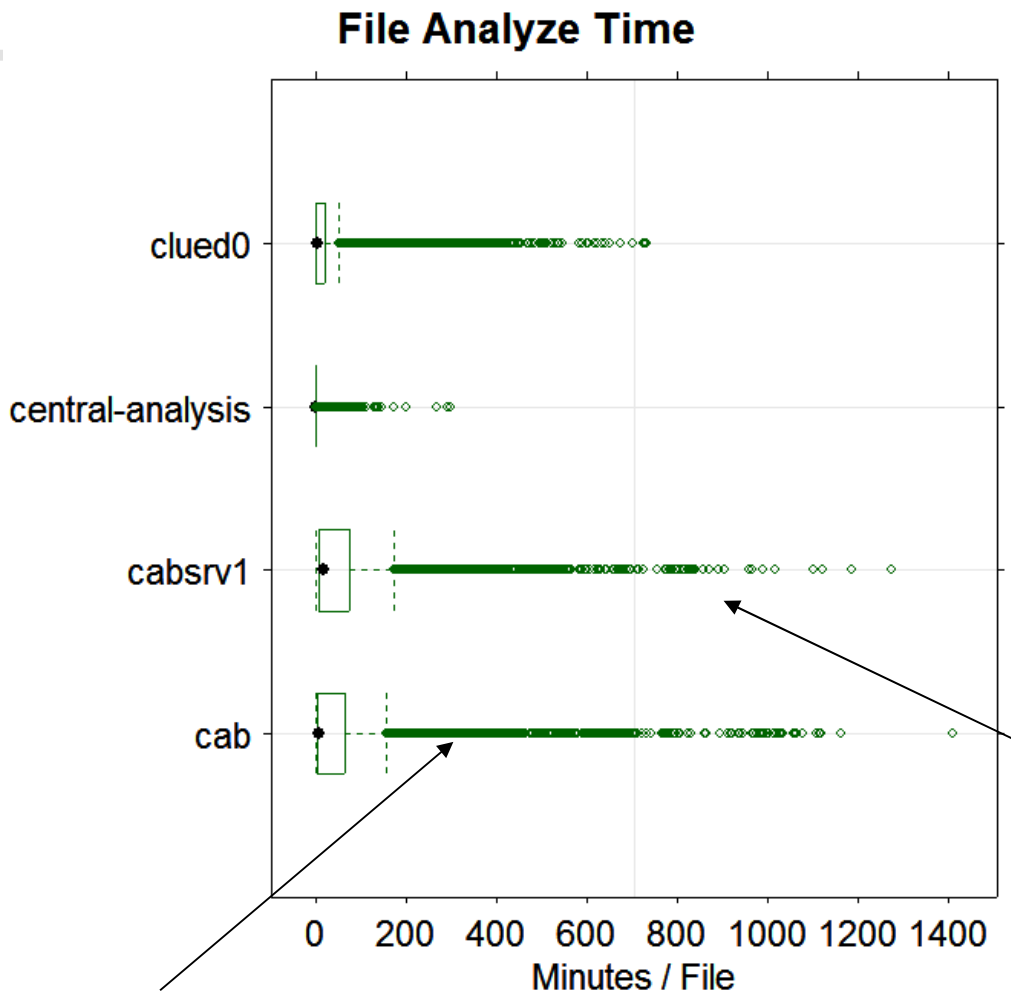
- Time between *Request Next File* and *Open File*
- For CAB and CABSRV1
  - 50% of enstore transfers occur within 10 minutes.
  - 75% within 20 minutes
  - 95% within 1 hour
- For CENTRAL-ANALYSIS and CLUED0
  - 95% of enstore transfers within 10 minutes

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

SAM Statistics - Operations Data



Files from tape come later

Cached Files delivered first and fast

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

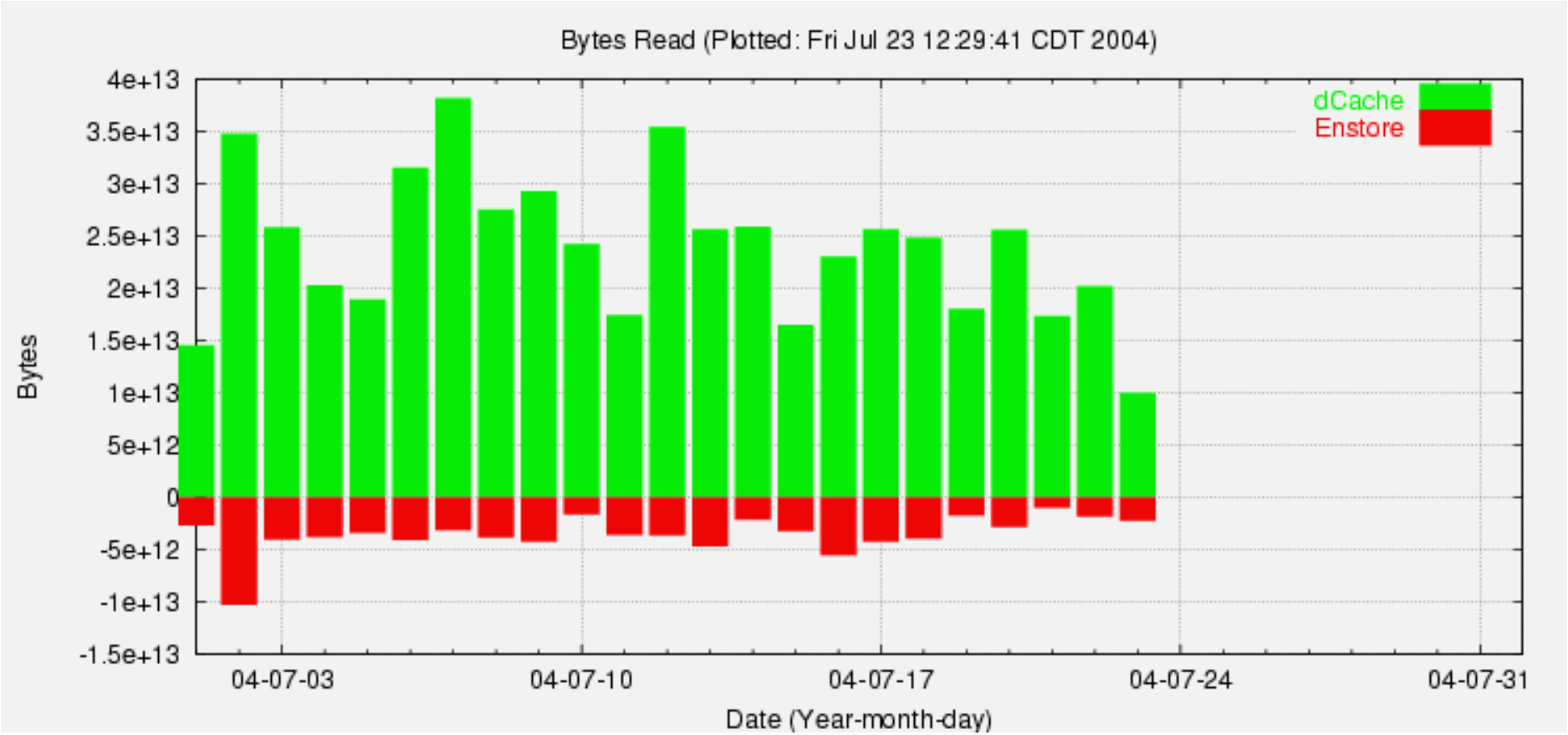
Cluster not CDF-exclusive - Need Grid to make this resource available to full CDF collaboration!

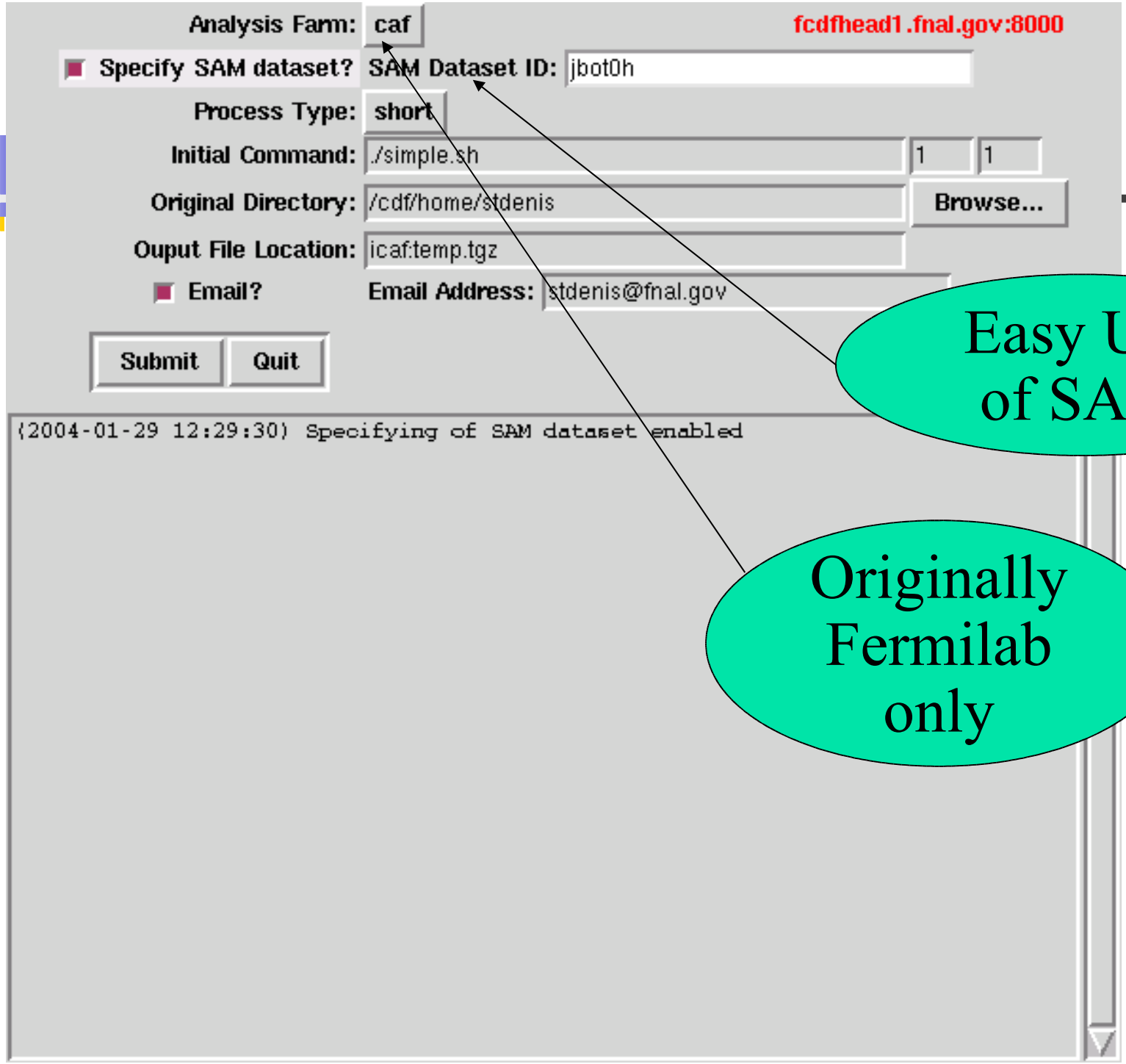
- May 1-6: 650
- May 7-17:704
- May 18-27:604
- May 28-31:710
- May total 492,860 cpu hrs, 1THz roughly
- June 1-7: 740, 8-14 780, 15 power out, 16-30 700
- June total 507,360 cpuhrs, 1THz roughly

CDF Data Handling: Dcache on CAF

CDF Reads 25 TB/Day on CAF

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

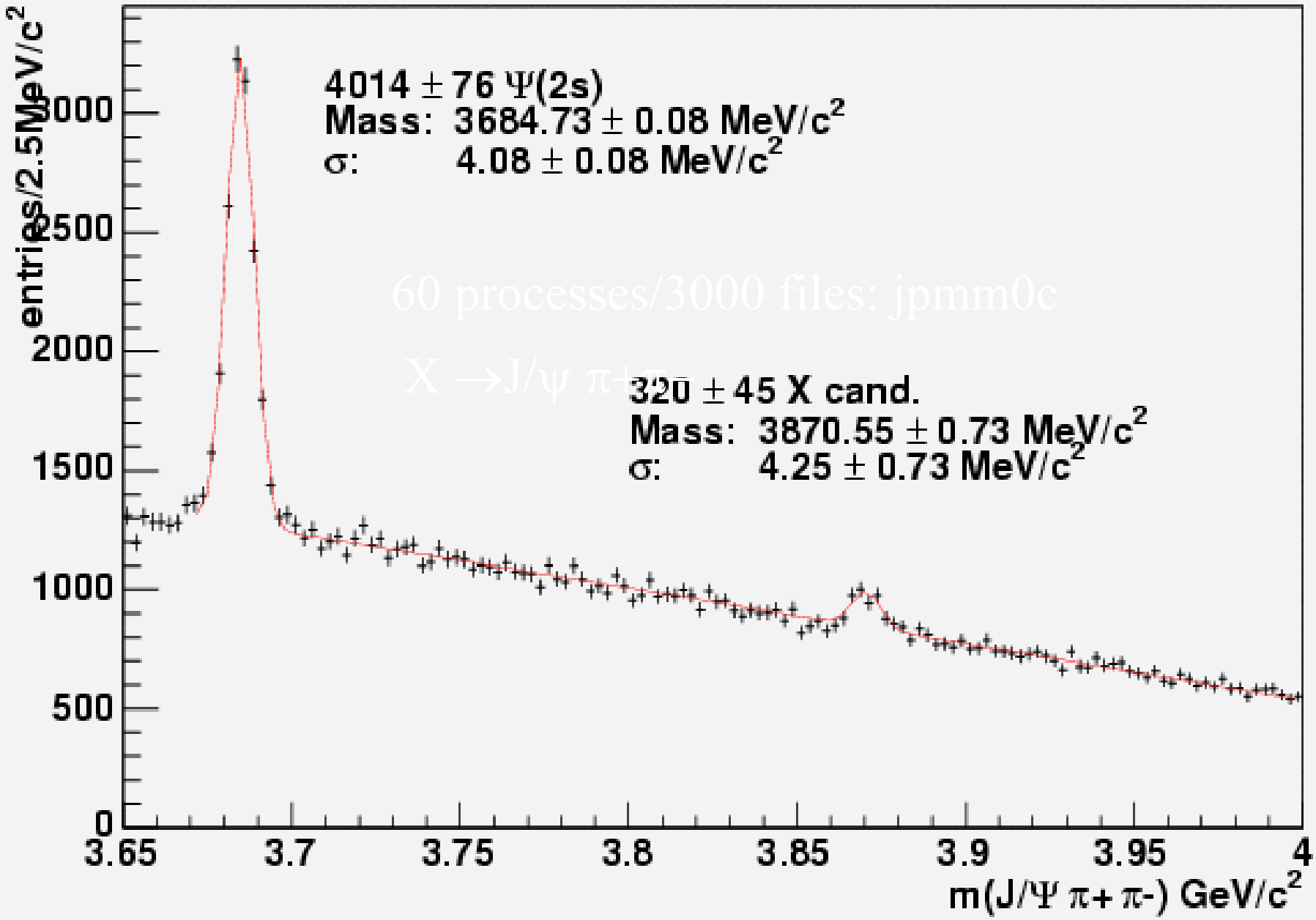
Email? Email Address: stdenis@fnal.gov

Submit Quit

(2004-01-29 12:33:39) jim analysis farm selected
(2004-01-29 12:33:44) Specifying of SAM dataset enabled

Uses SAM In Same Way

Example Use of Grid Resource!


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
0
500
1000
1500
2000
2500
3000
3.65
3.7
3.75
3.8
3.85
3.9
3.95
4

Screen Shot of Web page

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

CDF Datasets on SAM stations

Click on cnaf…

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

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

Datasets Stored Locally on cdf-cnaf: Locked (Still testing dynamic movement of files)

DATASET ID	GBYTES	EVENTS	FILES	CACHED	LOCKED
xbhd0d	[illegible]	[illegible]	27848	408(1%)	325(1%)
[illegible]	[illegible]	[illegible]	[illegible]	109(1%)	109(1%)
hbhd0[illegible]	[illegible]	[illegible]	52359	351(1%)	351(1%)
jbot0h	649.09	3240403	690	1(0%)	none
gmbs09	1224.65	7676037	1330	17(1%)	17(1%)
bpel0d	138[illegible]	[illegible]	2194	all	all
gpjj08	[illegible]	[illegible]	70	1(1%)	1(1%)
xpmm0d	[illegible]	[illegible]	2675	all	all
xpmm0c	524.53	16019542	[illegible]	all	all
jpmm08	575.70	27928[illegible]	[illegible]	[illegible])	none

Expert Usage! (testing dynamic movement

All in Cache

And Locked Via SAM

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.

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

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