# CDF experience with OSG

Rick Snider
*Fermilab*
on behalf of the CDF Offline

OSG User's Meeting
Brookhaven National Laboratory
June 16, 2008

# CDF

- Experiment studying collisions of protons and anti-protons at the Tevatron collider at Fermilab
- Each year, the experiment produces about:
  - 250 TB of raw data
  - 400 TB of reconstructed data
  - 120 TB of reduced datasets
  - 300 TB of MC data datasets

# CDF computing model

- Major processing steps
  - Raw data reconstruction
    - Performed at Fermilab
  - Data reduction and analysis
    - Performed at Fermilab
  - MC simulations
    - Detector simulations and “pseudo-experiments” data
    - Target off-site resources
  - Other CPU intensive computing
    - Event kinematic and topology probabilities (matrix element methods)

# CDF computing model

- Computing performed on a combination of
  - OSG resources at Fermilab
    - Some owned by CDF, some not
  - Remote OSG resources
    - Access resources around the Pacific Rim via OSG portals
  - LCG resources across Europe
  - Some legacy dedicated pools both at Fermilab and at collaborating institutions.

# CDF on the Open Science Grid

- Users submit jobs to two distinct portals for US/OSG-based resources
  - “FermigridCAF”:
    - Nodes hosted by Fermilab, operated within Ferimigrid/OSG
      - Submits primarily to four CE's
        - *FNAL_CDFOSG1 – FNAL_CDFOSG4*
      - Can in principle submit to any CE within Fermigrid
    - Have “local” access to data handling system and CDF offline code
  - “NAMCAF”:
    - Submits OSG sites in North America, including Fermigrid
      - Submits mainly to CE's at collaborating institutions (by agreement)
      - Intended to have only opportunistic access to Fermigrid CE's
    - Do not generally have access to data handling system or CDF offline code

This split between available functionality reflects history of experiment

- Have conducted large scale distributed computing for over four years
- Data is not distributed – not a large demand for off-site data access
- Migration to the Grid has been an evolution as technologies matured

# CDF on the Open Science Grid

- Target different computing problems to different sites
  - Direct processing that is event data intensive to on-site CE's
    - Raw data reconstruction
    - Data reduction and analysis
  - Send processing that does not require large scale data access to off-site CE's
    - MC simulations
      - Generated data is shipped back to Fermilab
    - Calculations for matrix element analyses

# Basic infrastructure

- Job submission, workflow management
  (see talk at 2008 Paradyne/Condor Week by D. Benjamin for details)
  - All access to OSG CE's is via Condor glide-in
    - Pilot jobs submitted to available CE's
    - Pilot job registers as a member of a Condor virtual pool
    - Wrappered user job is sent to the virtual pool member for execution
- Authentication
  - Pilot jobs run under service certificate
  - Users authenticate to submission portal via Kerberos 5
  - Fermigrid requires that user jobs run under user's ID
    - User's Kerberos credentials used to generate kx509 certificates
    - Use gLExec program to complete authorization for the user on the worker node, and allow jobs started as pilot to run with user's ID and certificate

# Basic infrastructure

- Data transport and storage
  - CE's at Fermilab use central data dandling system as a local resource
    - Based on SAM + dCache
  - Output data buffered on local disk
  - Output data transport via “fcp”
    - Provides queuing layer for underlying transport protocol
    - Currently using rcp/scp
    - Introduces transfer latency on the worker nodes

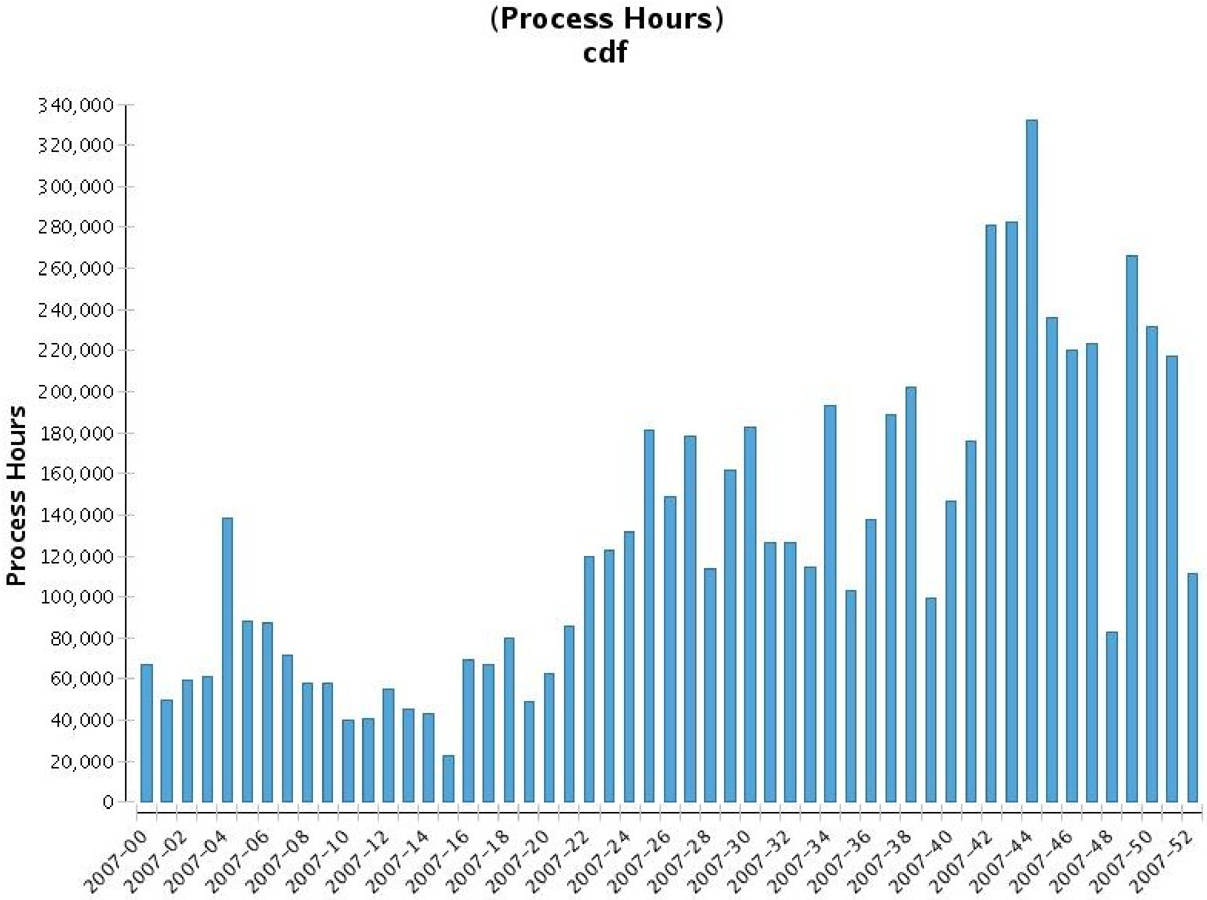
  Work in progress:

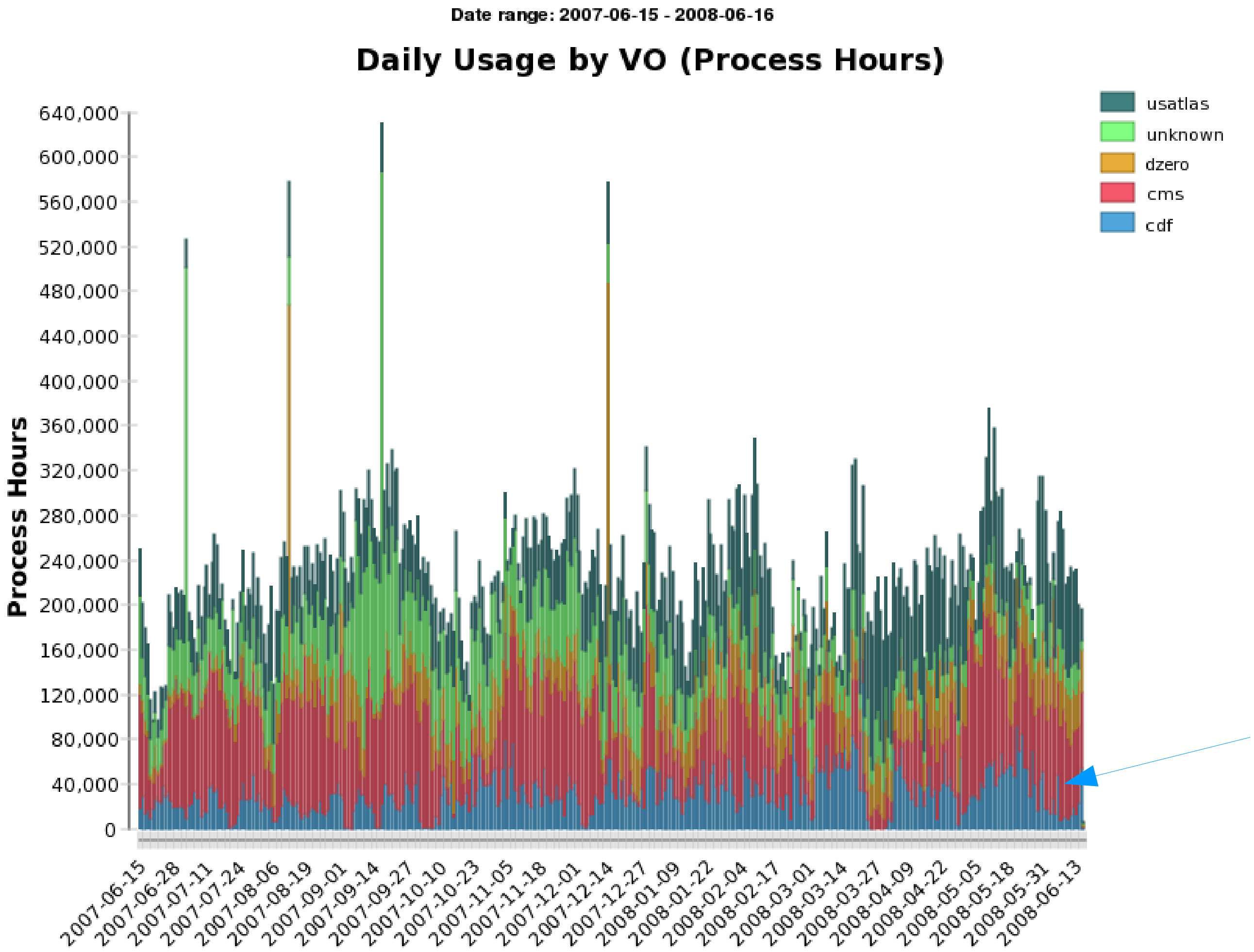
  - Prototyping SRM-based transport mechanism for MC data
    - Will use SRM-based durable storage
    - Prototype based upon existing DH system (SAM)
  - Will investigate SRM-based solution to data distribution
    - Large-scale re-processing could benefit from access to grid resources

# Basic infrastructure

- CDF software distribution
  - Locally mounted on computing owned by CDF
    - Not on CMS nodes
  - MC tarballs are self-contained (or attempt to be)
  - Investigating use of Parrot as alternative to self-contained tarballs
    - Used widely throughout LCG

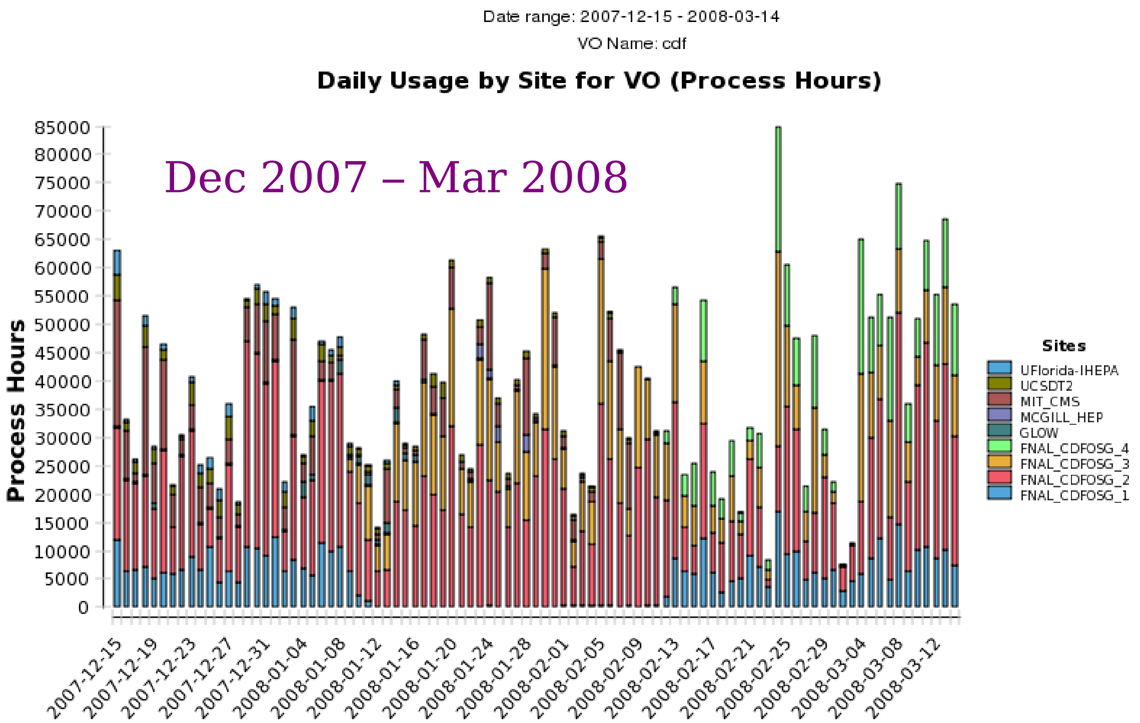
# CDF usage of OSG resources in 2007




Date range: 2007-06-15 - 2008-06-16
Daily Usage by VO (Process Hours)
Process Hours
640,000
600,000
560,000
520,000
480,000
440,000
400,000
360,000
320,000
280,000
240,000
200,000
160,000
120,000
80,000
40,000
0
usatlas
unknown
dzero
cms
cdf
2007-06-15
2007-06-28
2007-07-11
2007-07-24
2007-08-06
2007-08-19
2007-09-01
2007-09-14
2007-09-27
2007-10-10
2007-10-23
2007-11-05
2007-11-18
2007-12-01
2007-12-14
2007-12-27
2008-01-09
2008-01-22
2008-02-04
2008-02-17
2008-03-01
2008-03-14
2008-03-27
2008-04-09
2008-04-22
2008-05-05
2008-05-18
2008-05-31
2008-06-13

CDF

# CDF usage across all OSG sites

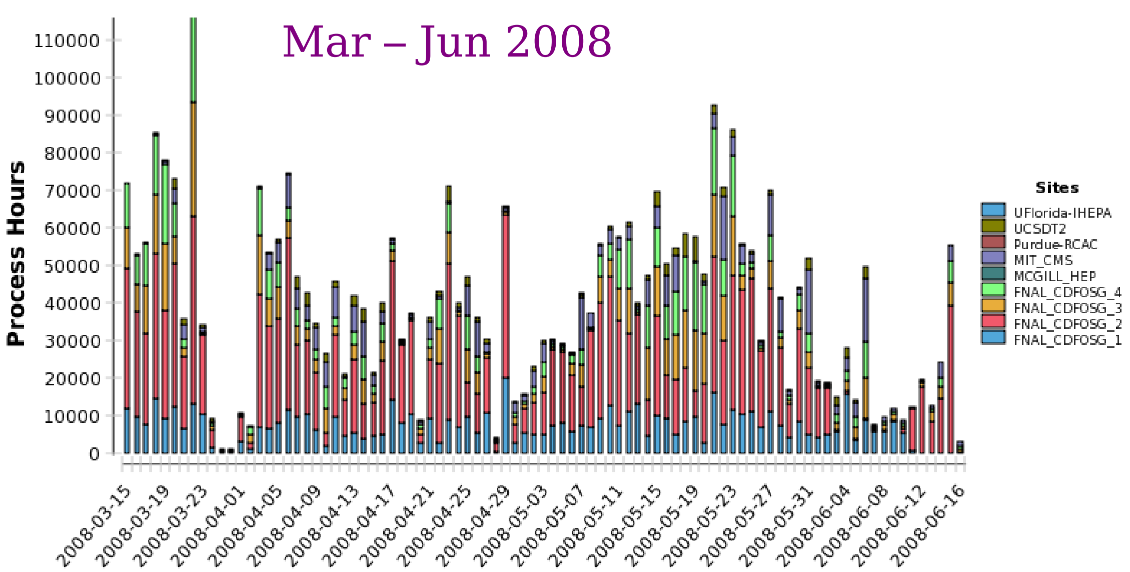


Dec 2007 – Mar 2008

Date range: 2008-03-15 - 2008-06-16

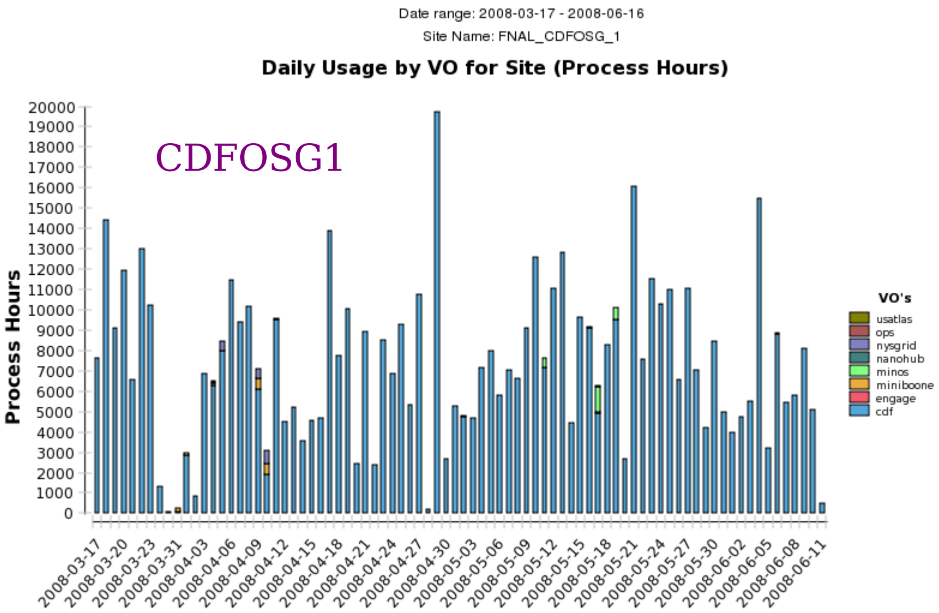
VO Name: cdf

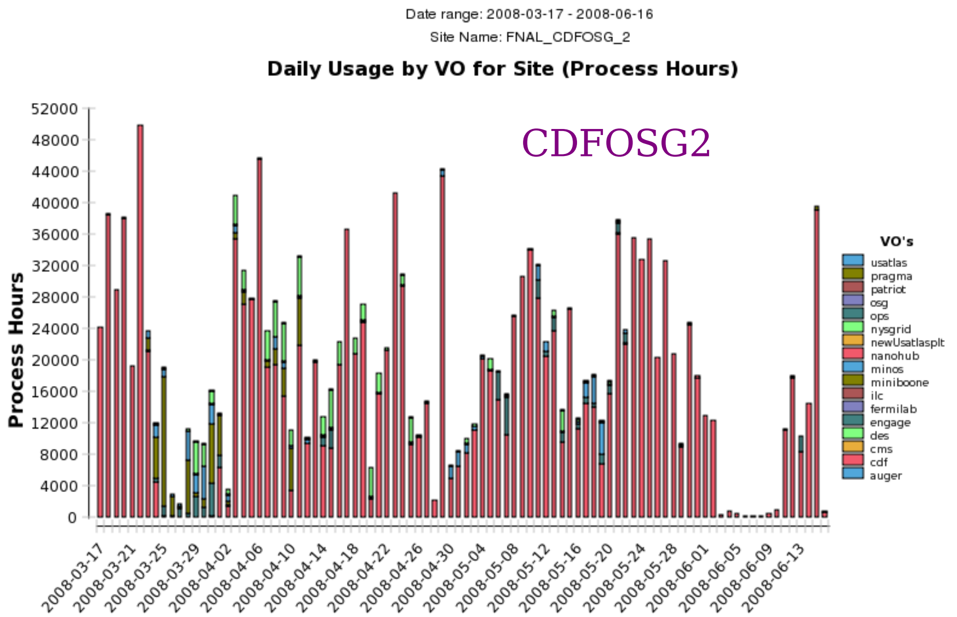
ıge by Site for VO (Process Hours)

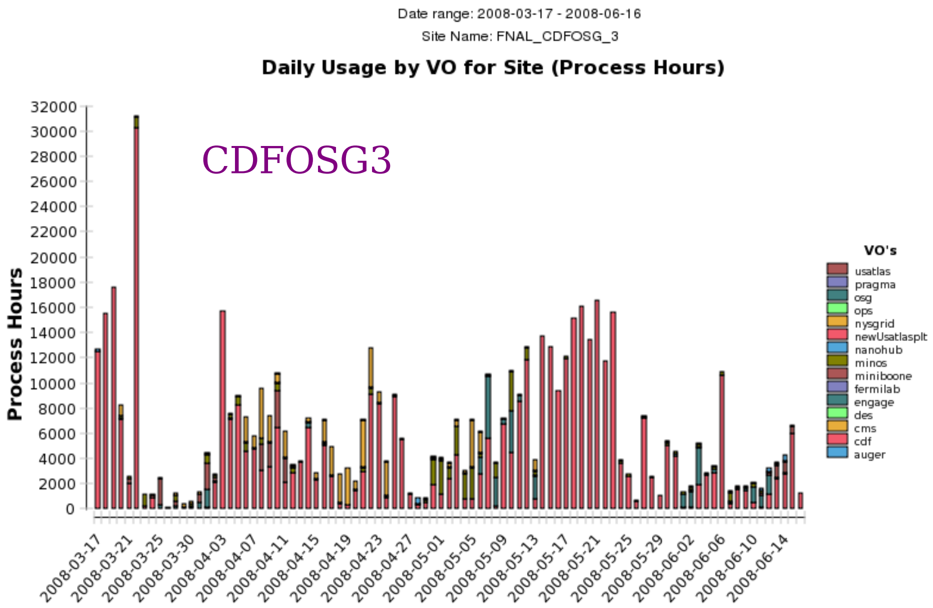


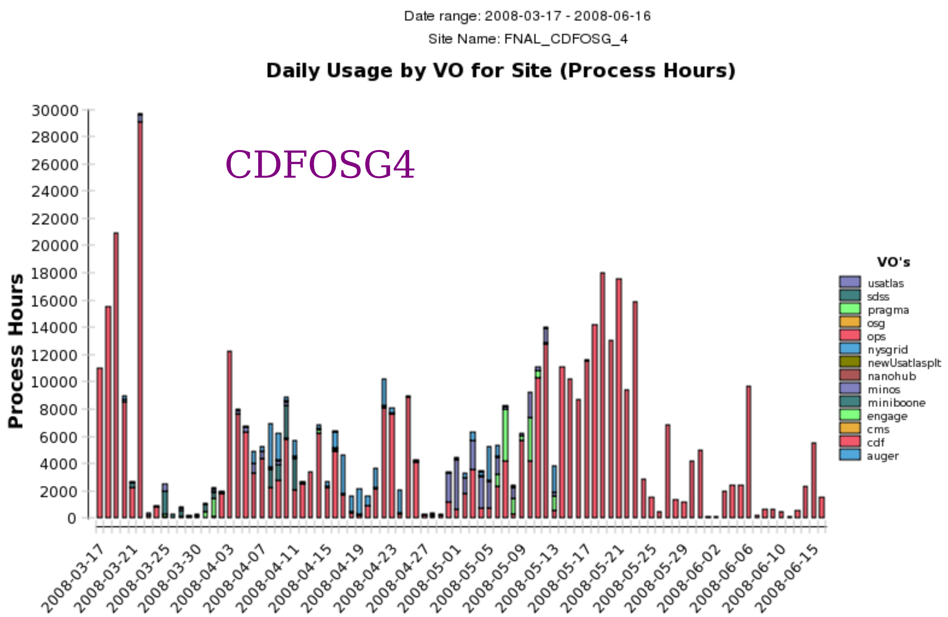
Mar – Jun 2008

# OSG usage of CDF CE's at Fermilab



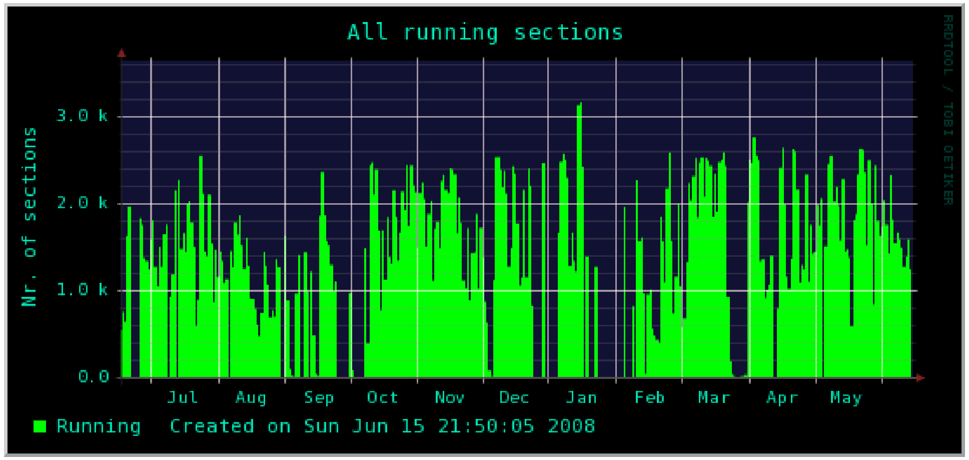






# FermigridCAF and NAMCAF

FermigridCAF

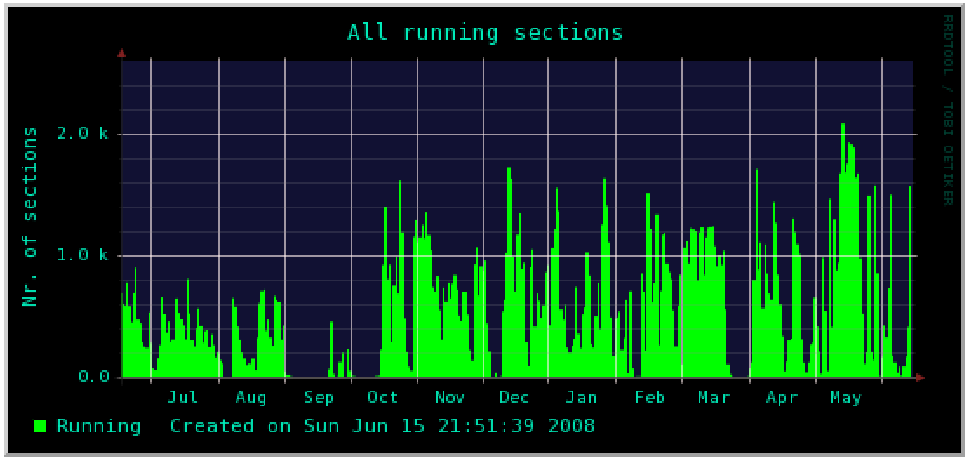


■ Running Created on Sun Jun 15 21:50:05 2008

Total capacity available for FermigridCAF is >3100 slots.

Have not been able to fill these slots, so run some CE's under NAMCAF.

NAMCAF



■ Running Created on Sun Jun 15 21:51:39 2008

# Issues

- Scaling issues with current glide-in infrastructure
  - Observe under-utilization on FermigridCAF
    - Cannot serve all existing on-site resouces
      - Have temporarily limited FermigridCAF to a subset of available CE's
      - Using NAMCAF to fill in for balance
      - Users do not or cannot exploit available resources on NAMCAF

      Not an OSG middleware problem!
    - Users do not choose effectively between FermigridCAF and legacy dedicated pool at Fermilab
  - Adopting GlideinWMS
    - Eliminates home-grown CDF-specific version
      - Improves maintainability
    - Allows glide-in functions to be run on different machines from those handling user submissions
      - Better distributes load, improves scalability

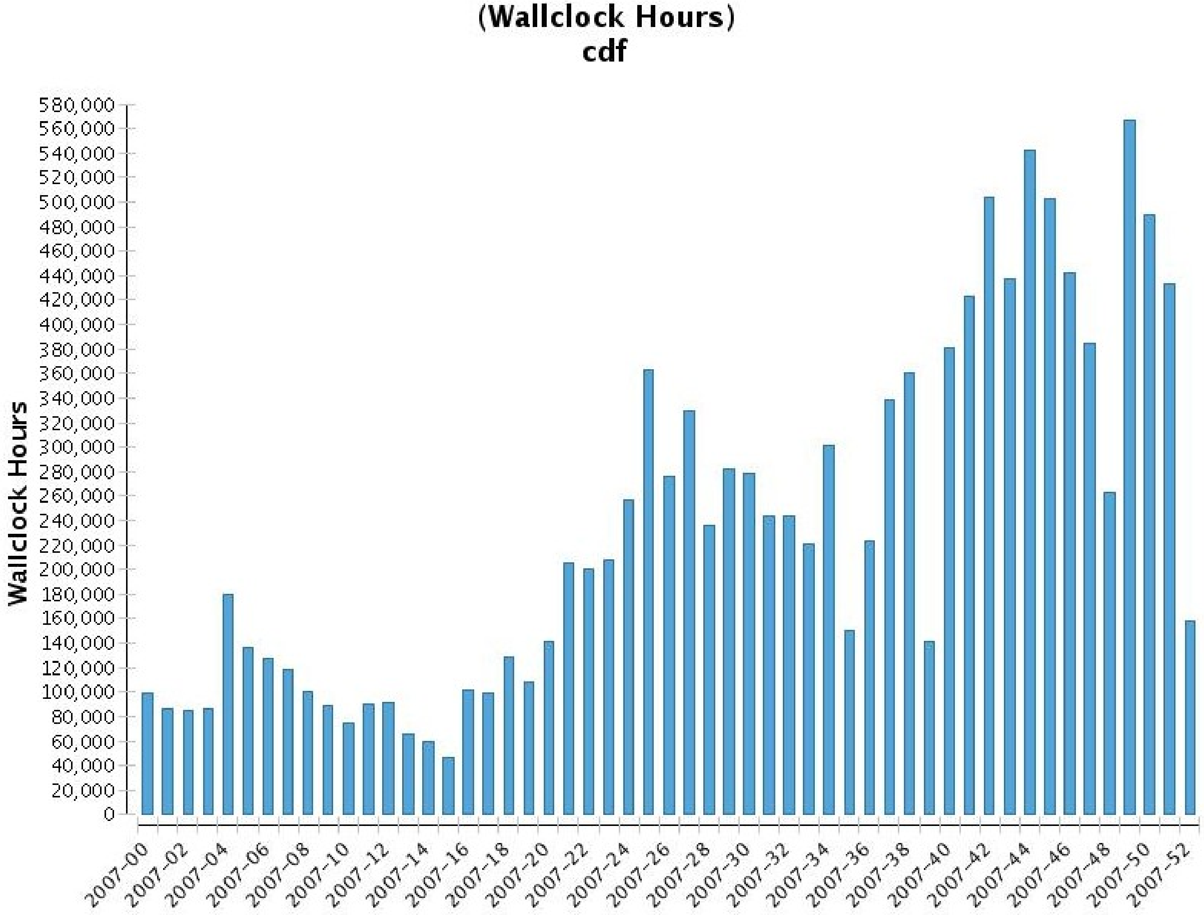
# Issues

- System space protection
  - User processes allowed to consume resources required for the OS
    - Both memory and disk
  - A rogue user process can cause a node to crash
    - Several instances at CDF of single user taking down many nodes
  - Can fix disk issues with configuration
  - Memory?

# Summary

- CDF is a large user of OSG resources, but...
  - Utilize mainly resources owned by the experiment, collaborating institutions
  - Are still in the process of migrating toward common middleware
  - Success at meeting physics goals still require dedicated pool at Fermilab
    - Have about 1200 cores in last legacy pool at Fermilab
- Have a clear roadmap for the next few months
  - Adopt GlideinWMS
  - Upgrade hardware
  - Migrate all resources into Fermigrid/OSG
  - Deploy SRM for MC transport
  - Investigate SRM for data distribution

# CDF usage of OSG resources in 2007


(Wallclock Hours)
cdf
Wallclock Hours