CDF Computing Overview

Igor Sfiligoi
for the CDF Offline Group

International Finance Committee
April 29th, 2006

Outline

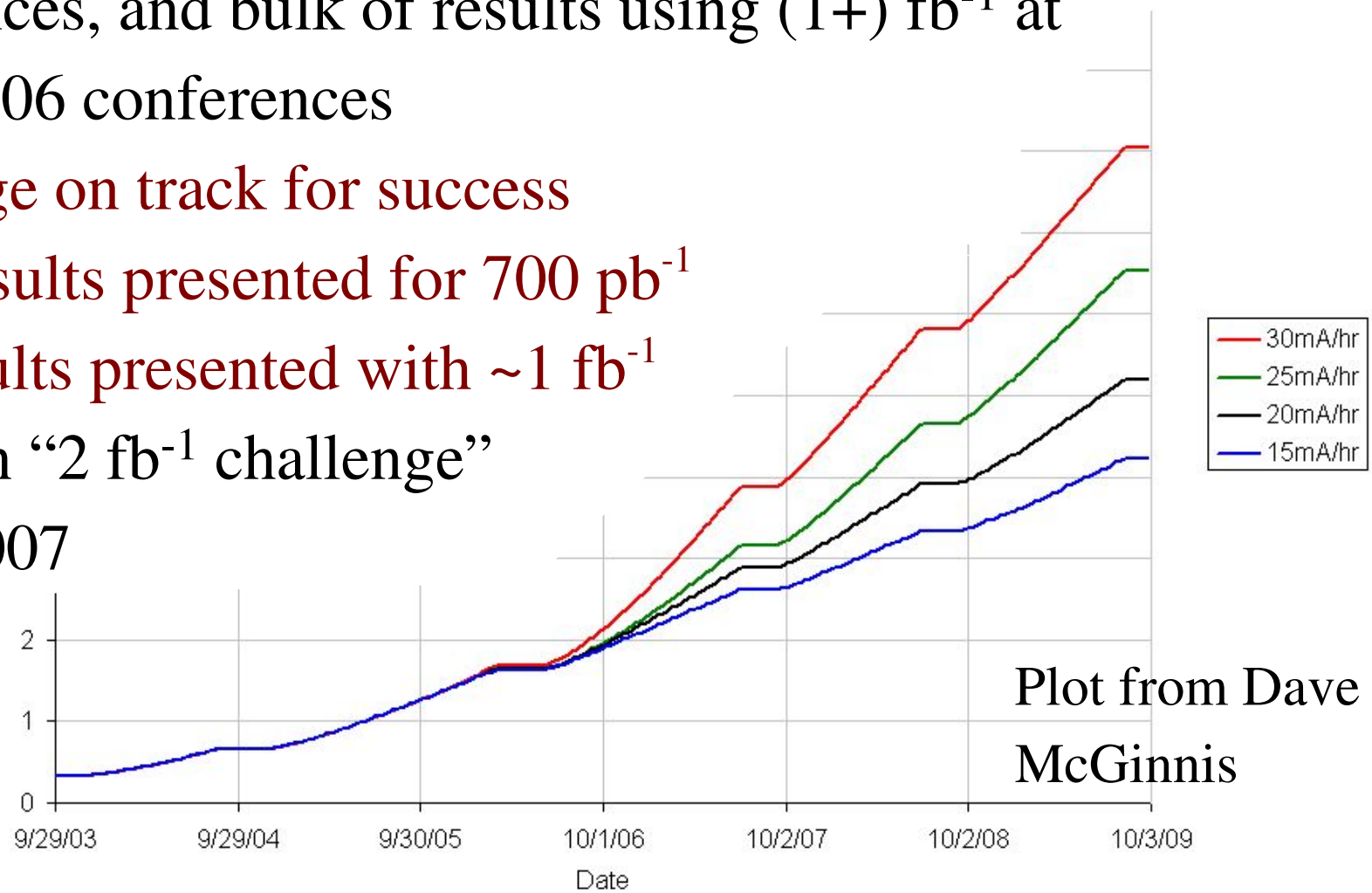
- Mission
- Software and Algorithms
- Operations
- Infrastructure Development
- dCAF status report
- Summary

Mission Summary

- Rapid turnaround from taking collider data to physics results and publications
- Use distributed, shared GRID computing resources to meet growing needs for MC and analysis
  - Use standard, supported tools for computing and data-handling
  - Minimize disruption in user interfaces to maximize physics productivity
- Finalize reconstruction and simulation software to allow collaboration to focus on data analysis
- Streamline operations to reduce personnel needs

CDF Physics Goals

- "1 fb^{-1} challenge for 2005-2006"
  - Present results using 1 fb^{-1} of analyzed data at winter conferences, and bulk of results using (1+) fb^{-1} at summer 06 conferences
  - Challenge on track for success
    - 23 results presented for 700 pb^{-1}
    - 4 results presented with ~1 fb^{-1}
- Follow with "2 fb^{-1} challenge" for 2006-2007



Plot from Dave McGinnis

Software Status

- Reconstruction code has achieved high level of physics performance and operational stability
  - Planning to make Final Production Release in 2006, to include improvements in forward tracking
- Simulation code reached high level of maturity
  - Good agreement with data
  - Simulation machinery incorporates run-dependent
    - detector configurations
    - Multiple interactions as a function of instantaneous luminosity
  - Allows straight-forward extension of Monte Carlo datasets to match collider data

Data Processing

- Achieved "one-pass processing mode" in 2005
  - Calibrations delivered within 4 weeks of data-taking
  - Delivered reconstructed data for physics analysis within 2 months.

	Period Lumi	Total Lumi	Available for Physics
Dec. 04 - Mar 19, 2005	130	680	July 2005
Mar. 19 - May 20, 2005	130	810	Aug 2005
May 20 - Jul 20, 2005	100	910	Sept 2005
Jul 20 - Sep 04, 2005	95	1005	Oct 2005
Sep.05 - Nov.09, 2005	135	1140	Jan 2006
Nov.10 - Jan.14, 2006	110	1250	Mar 2006
Jan.15 - Feb.22, 2006	50	1300	Apr 2006

~6 weeks!

Data Processing (2)

- We have generalized and streamlined the production farm infrastructure
  - Reduced FTE load needed to run production
  - Improved and Automated Error Recovery
- Extended use of farm infrastructure for making production ntuples for the whole collaboration - saves the collaboration time and effort
- Plan to lower the time between data taking and availability for physics analysis down to 4 weeks

Computing Infrastructure Development

- Goal: create a uniform, GRID-enabled computing platform for all computing
  - Data processing & N-tuple production
  - Monte Carlo production
  - User analysis
- Accomplishments in 2006
  - New reconstruction farm control infrastructure
  - Start the transition of FNAL CDF resources into Grid pools
  - Offsite computing installations on both EGEE (Europe) and OSG (US) computing pools

New Reconstruction Farm

- Reconstruction farm is now integrated into a CDF-wide common computing platform
  - *"Just another CAF"*
- Used this flexibility in the winter
  - Expanded into the analysis CAF as needed
  - 25% of all the reconstruction was performed outside the production farm

Use of Grid resources

- Successful use of Condor "glide-in" technique
  - Based on the Condor batch system (developed at U. Wisconsin)
  - Allows CDF job submission to shared computing pools
    - Ensures access to large, growing off-site computing resources
  - User interface identical to the other dCAFs
  - Five successful production installations of "GlideCAF":
    - FermiGrid onsite (both in CDF and non-CDF owned pools)
    - In Europe, Italian CNAF and Lyon
    - In the US, San Diego and MIT
  - Most other dCAFs expressed interest in migrating to GlideCAF
    - Allows them to integrate into shared pools

Merging dCAFs to better exploit Grid resources

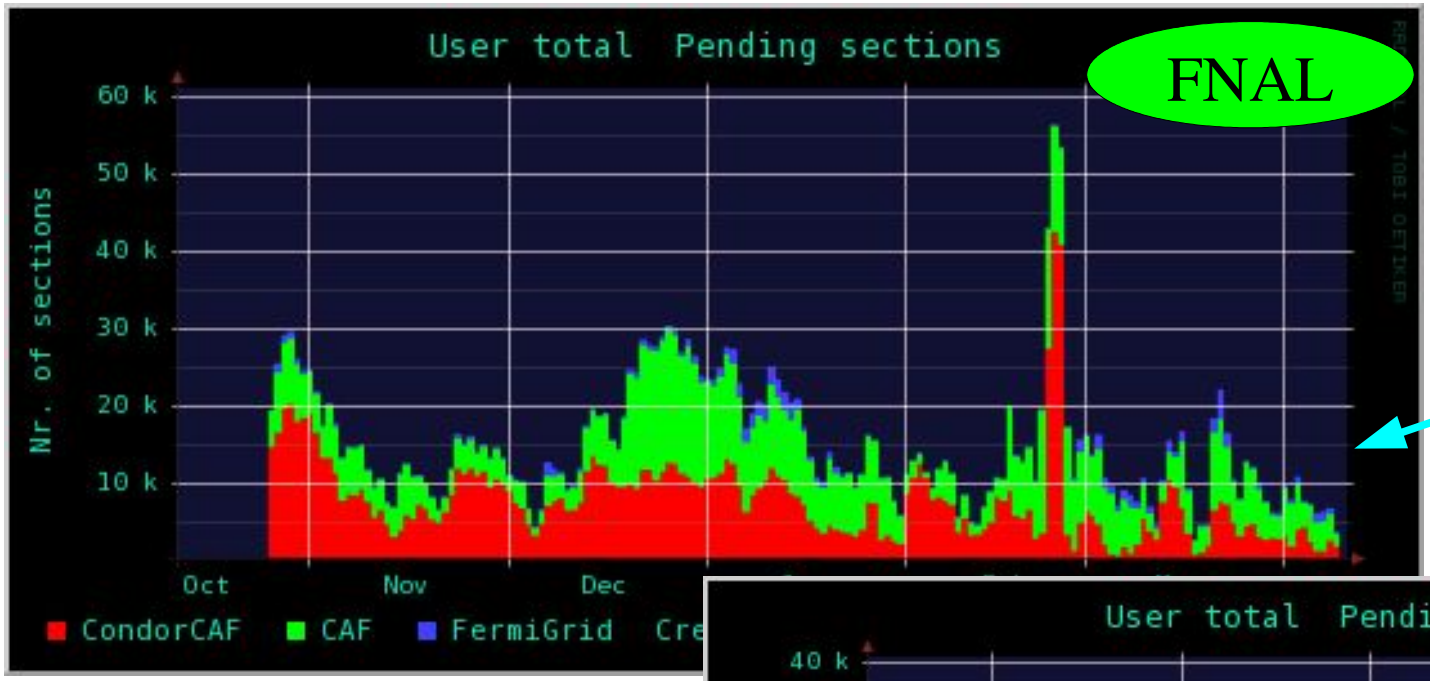
- GlideCAFs are working fine, but...
  -> Can improve utilization efficiency by reducing the number of site choices users need to make
- Merge several GlideCAFs into one
  - For firewall traversal - evaluating a Condor Proxy service
- No changes needed to the CDF-specific code
  - Works fine on small scale (few hundred CPUs)
  - Initial scalability issues resolved, more testing in progress
- In collaboration with San Diego CMS group

Exploring standard Grid interfaces

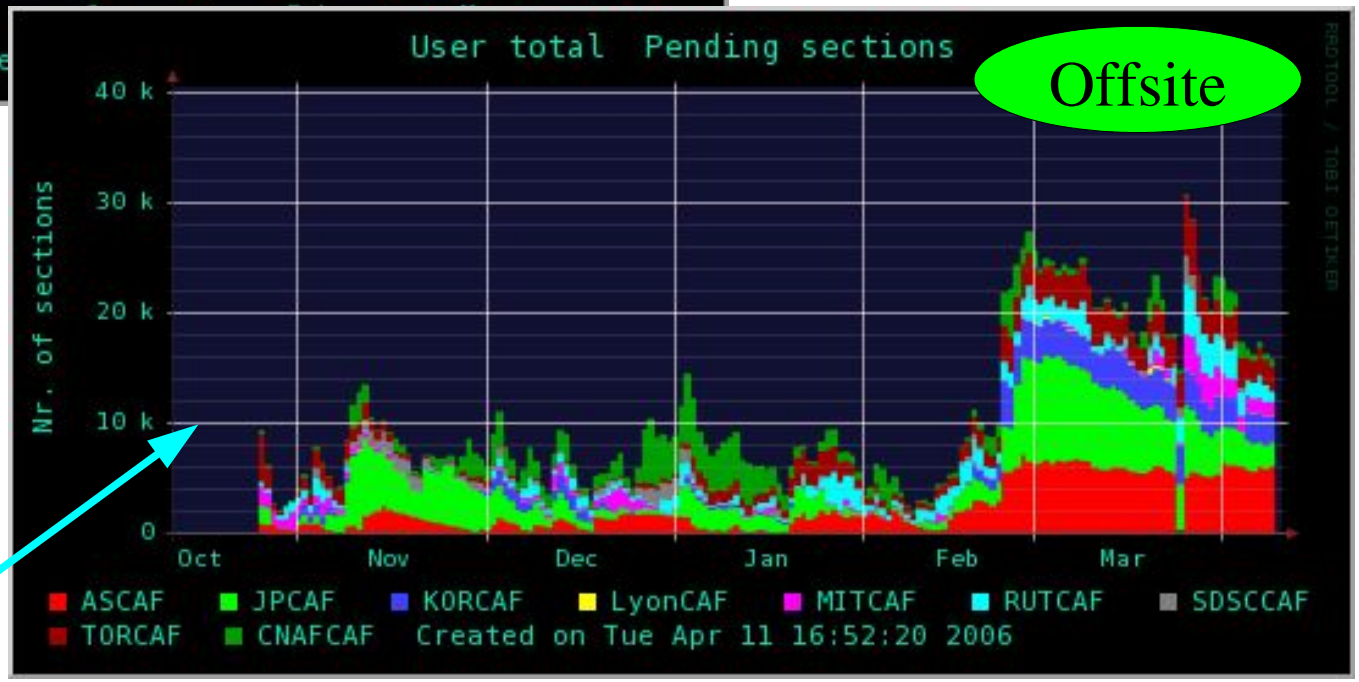
- Getting ready to use EGEE resources with LCG Grid tools
  - Trying to use the gLite Resource Broker
    -> Single point of submission for EGEE
- A new CAF interface has been written
  - Same user interface, but completely new CDF-specific code
  - Small scale MC production demonstrated by few power users
    - Needs more work to support generic users
  - Plan to start massive MC production real soon
- Led by CDF-Italy, in collaboration with INFNGrid

dCAF status report

Heavy demand for CPU



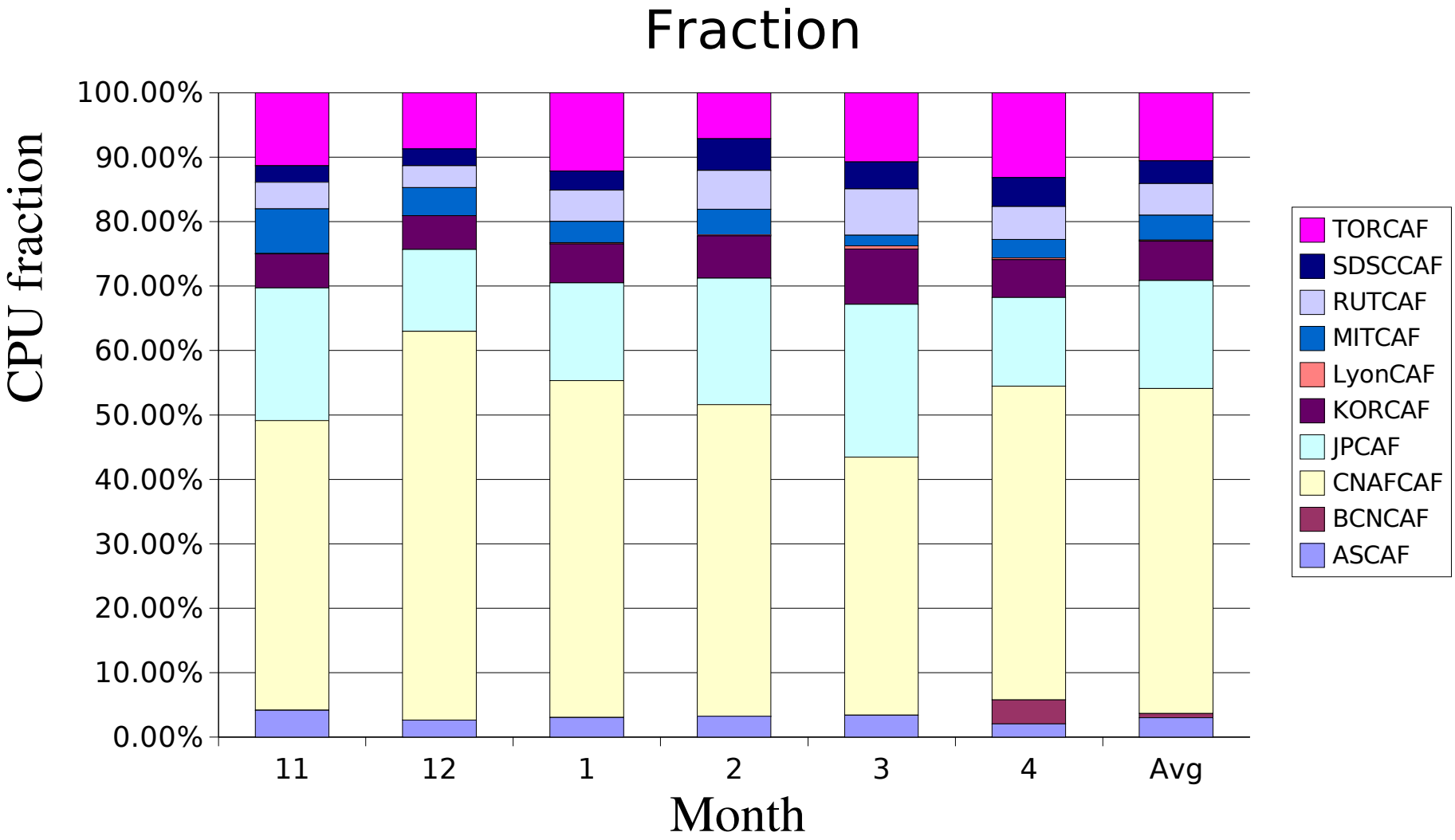
15k



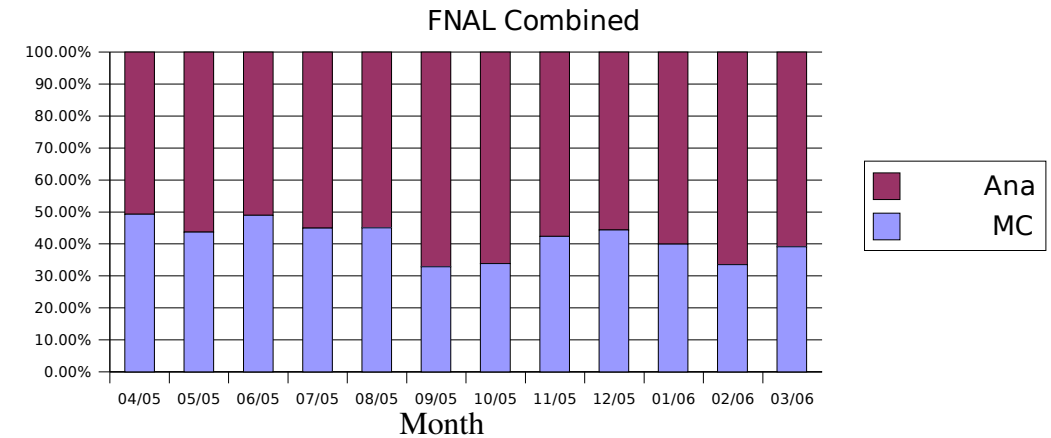
10k

Always something in some queue

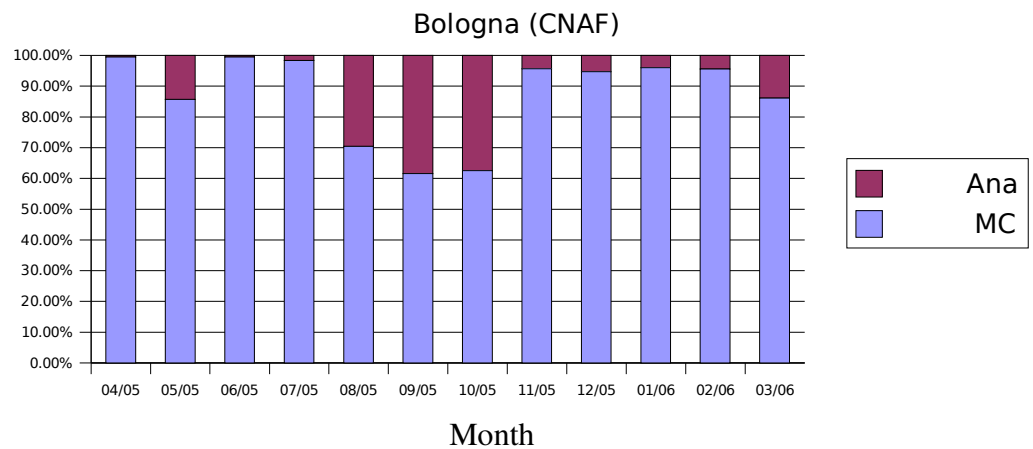
Offsite CAF resources use



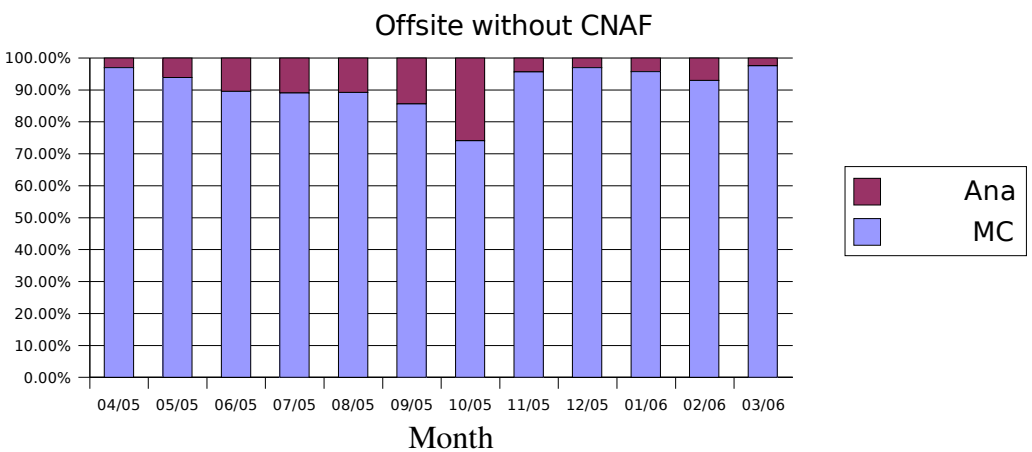
dCAF CPU usage by type



- Fermilab mostly analysis oriented, with some user MC



- CNAF had a peak of analysis, but is used for MC most of the time



- Offsite mostly used for MC

Summary

- We have achieved and sustained fast turn-around between data-taking and physics results
  - During the 2006 winter conferences:
    - 23 results presented for 700 pb^{-1}
    - 4 results presented with ~1 fb^{-1}
- Reconstruction and Simulation software mature and stable
  - Aiming for final production release this year
- CDF distributed computing working well
  - An important component in meeting our needs
- Offsite resources are critical for the CDF Physics Program
  - And will continue to be so

Backup slides