

Computing Highlights, Status and Plan

Donatella Lucchesi
for the CDF Offline Group

International Financial Committee
October 30^{en} 2007

Introduction

CDF computing model evolved and is evolving to cope with GRID: **dedicated farms** --> **GRID pools**

I will show a snapshot of the current CDF configuration and the works in progress for:

- Software and algorithms improvements
- Data Production
- Current on-site & off-site farms and GRID resources use
- On going developments

Software and Algorithms Improvements

- "Gen 6" big success for Data Production, Data Analysis and Monte Carlo generation
- Can we do better? Difficult, but yes

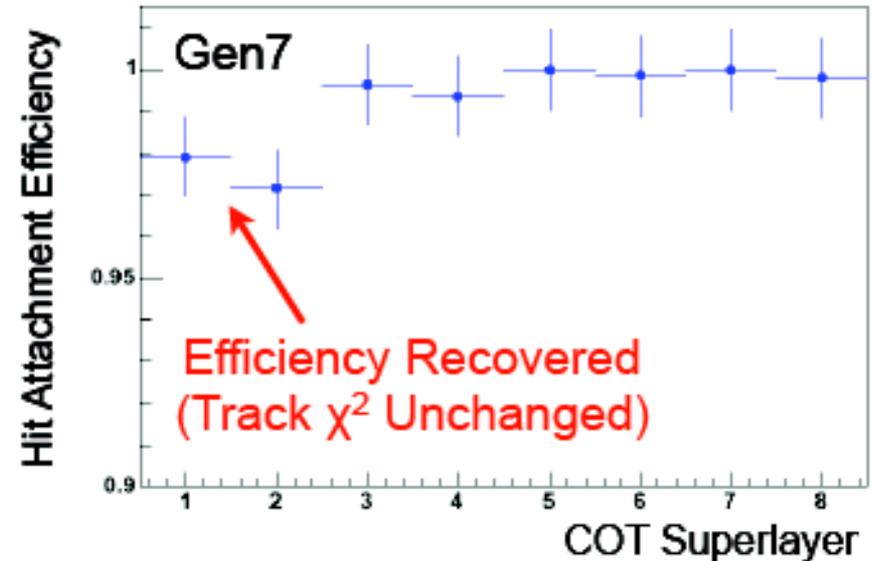
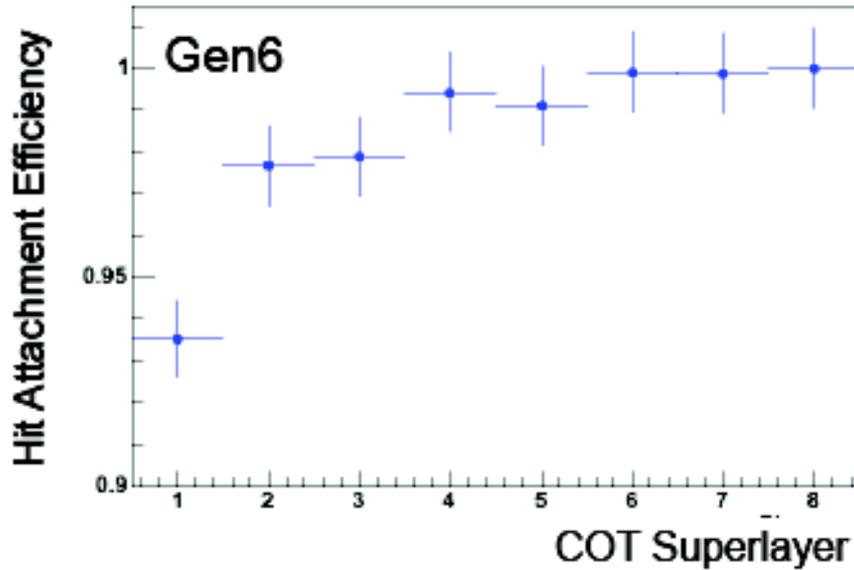
"Gen 7" has :

- ✓ Improved Tracking
- ✓ Better Calorimeter Simulation
- ✓ New Trigger Simulation to reflect trigger upgrades
- ✓ SL4 compatibility

Tracking Improvements

U. Husemann,
Tracking Group

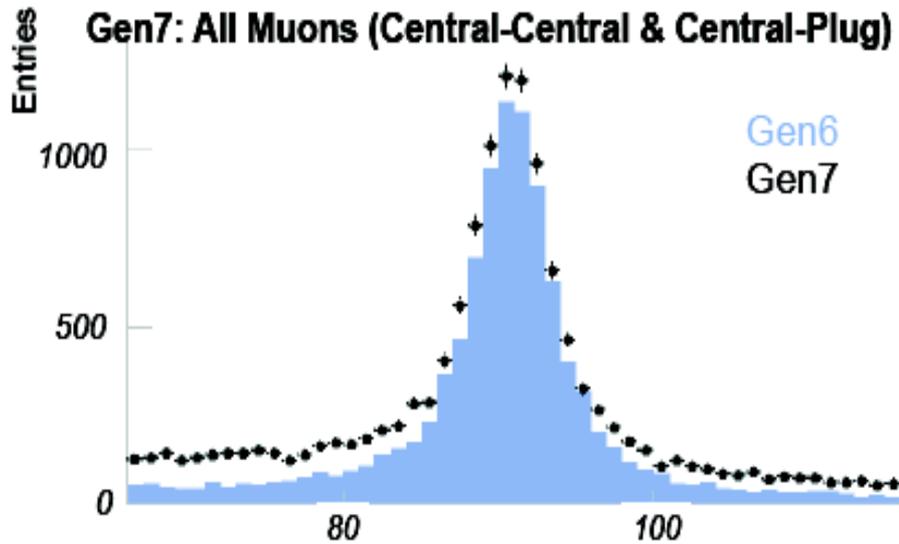
Increased COT performances at high luminosity



Tracking Improvements cont'd

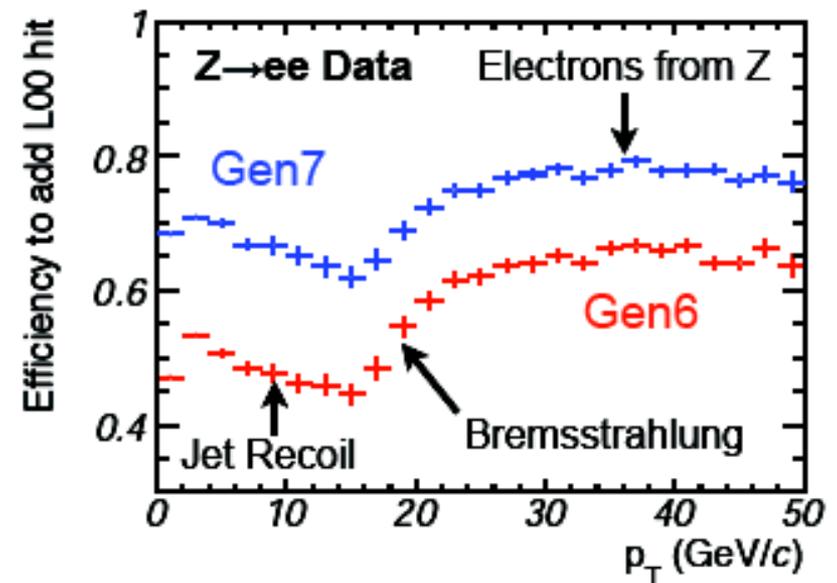
U. Husemann,
Tracking Group

Forward tracking capabilities, up to $|\eta| \sim 2.8$



35% more events
in 1σ around Z peak

Efficiency to add a L00 hit
improved up to 40%



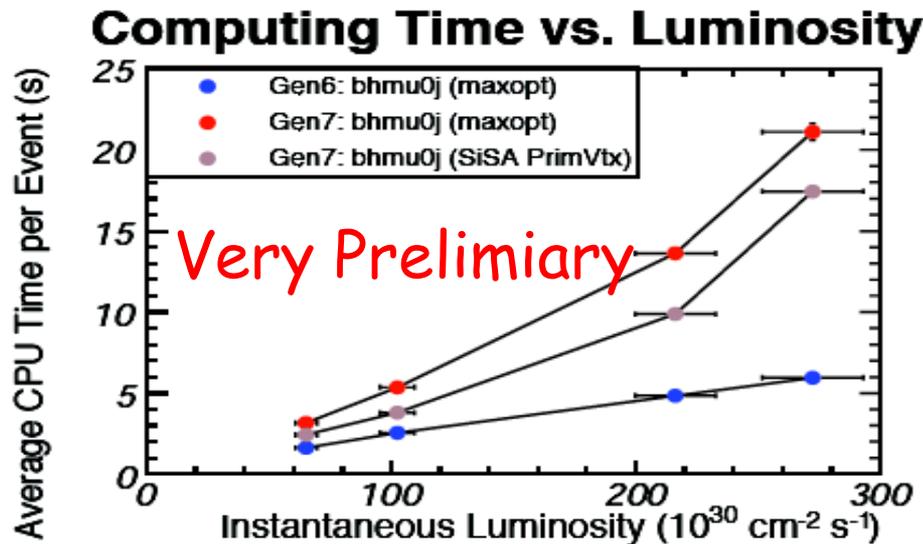
Tracking Improvements cont'd

U. Husemann,
Tracking Group

Gen7 b-tag:

- ✓ Increased b-tagging efficiency of 4% for central jets
- ✓ Under optimization forward b-tagging

Timing

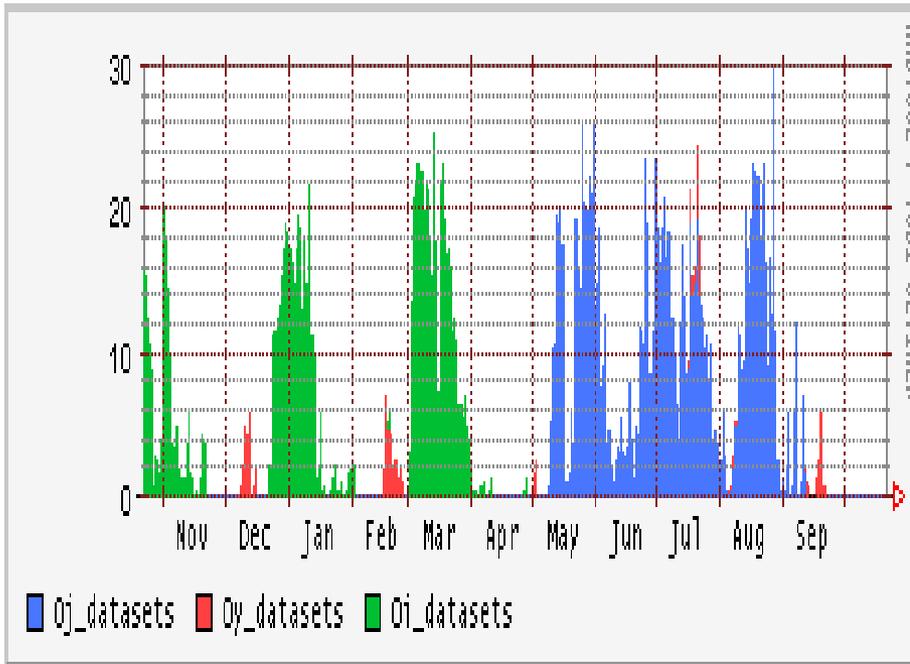


Not official yet, but we know we have CPU consumption of 1.3xGen 6

Data Processing

T. Miao

Well established "one-pass processing mode":
Calibration -> Production -> N-tupling



Data split up in "periods" of
 $\sim 260\text{pb}^{-1}$ ~ 400 million events

A good daily production rate:

15-20 million events

Peak rate with smooth

DH/CAF:

25-28 million events

Data available to users in about 6-8 weeks

Working to reduce delays

October 30th

Computing Infrastructure

Goal is to create a uniform, GRID-enabled computing platform for all computing:

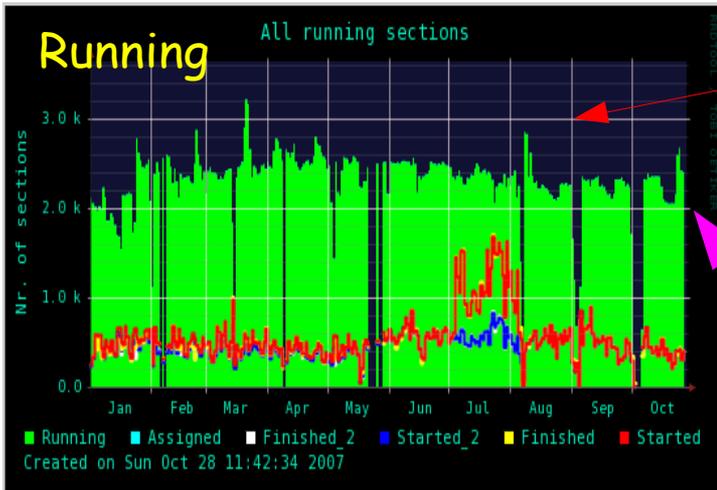
- ✓ Data Processing and N-tuples production
- ✓ Monte Carlo Production
- ✓ User Analysis

Accomplishments in 2007

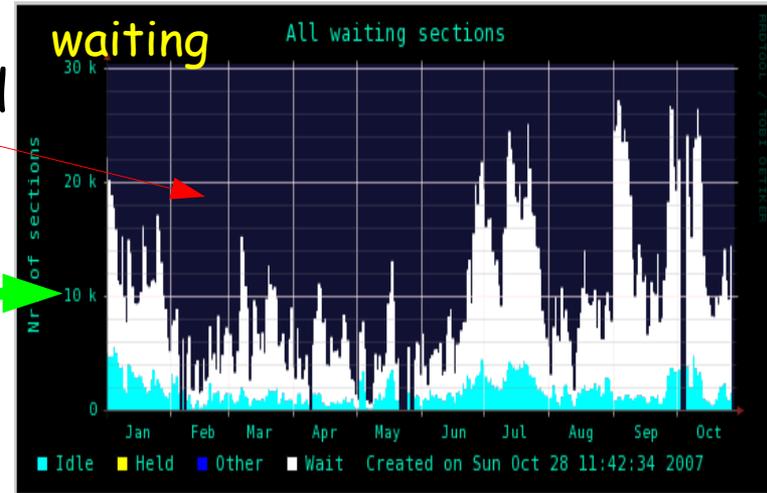
- ✓ Production Farm merged into the CAF (dedicated farm)
- ✓ Improvements in FermiGrid resources exploiting
- ✓ In progress the transition of CAF to FermiGrid pool
All new resources go in FermiGrid pool
- ✓ Reduced number of dCAF and increased usage of GRID resources

On-site overview

CAF Farm, submission via CAF portal



Large use and demand

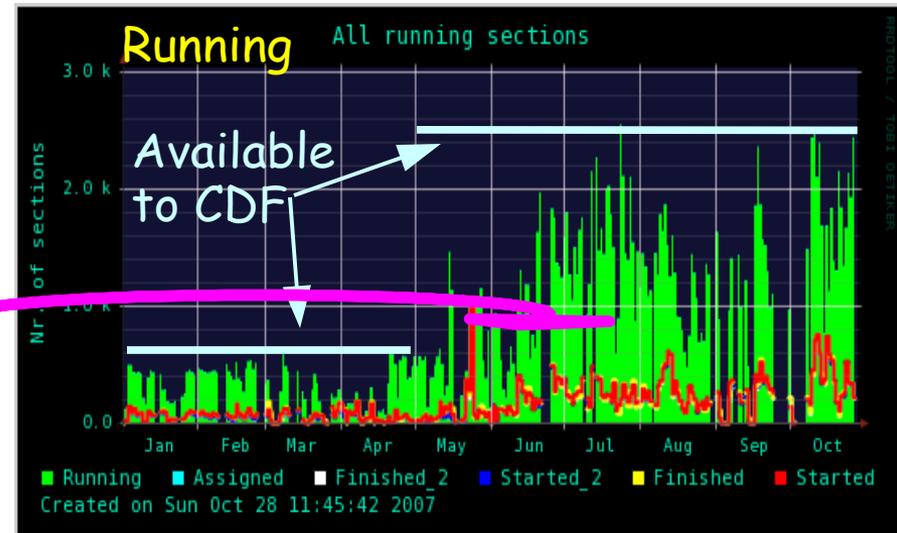


2K 10K

FermiGrid pool

Submission via "glidein" portal

Start improve FermiGrid use

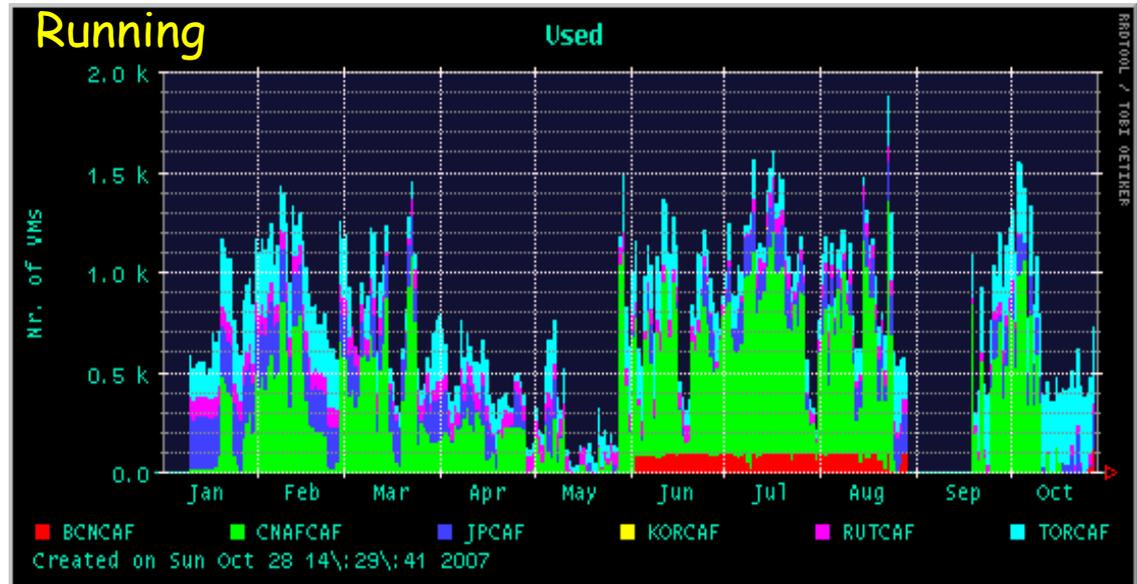


October 30th

Off-site overview

Off-site dedicated farm use

These farms are CAF and GlideCAF maintained by local CDF personnel with CAFTeam support

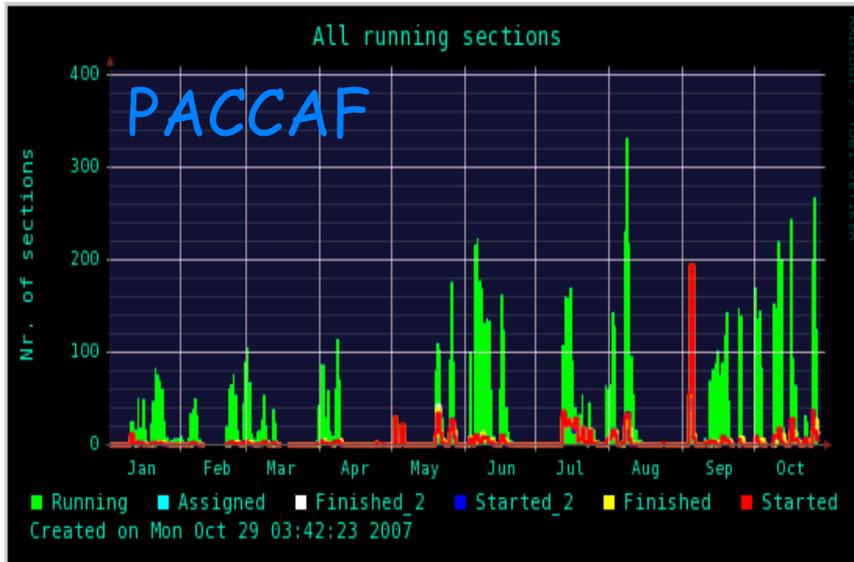


Plan: access them via GRID portals -> reduce maintenance

Possible exception:

CNAF because it hosts several data for analysis

GRID Portals overview: "glidein" based (OSG)

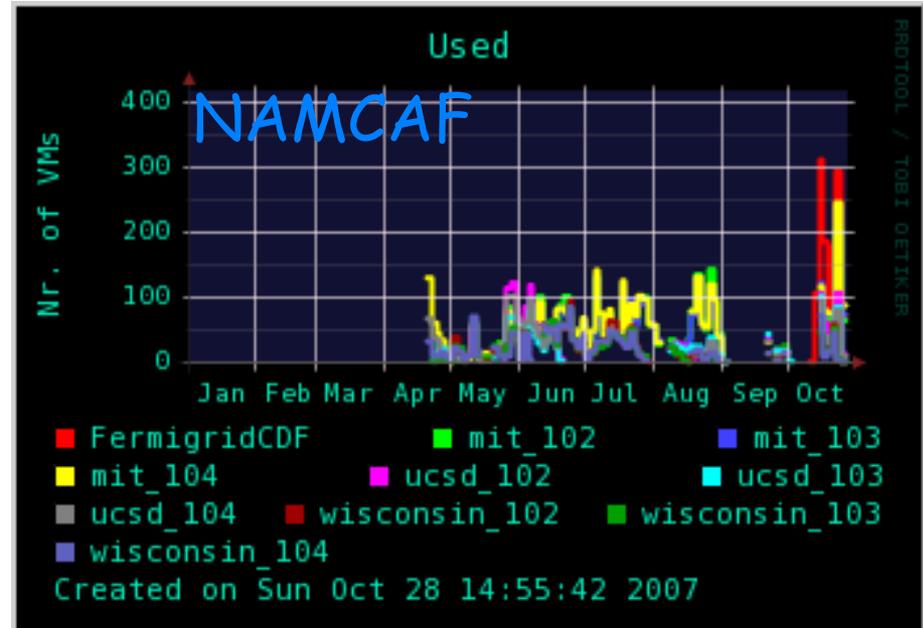


Sites:

IPAS_OSG, Taiwan-LCG2
KR-KISTI-GCRT-01

Plan:

- include JPCAF, KORCAF
- expand KISTI



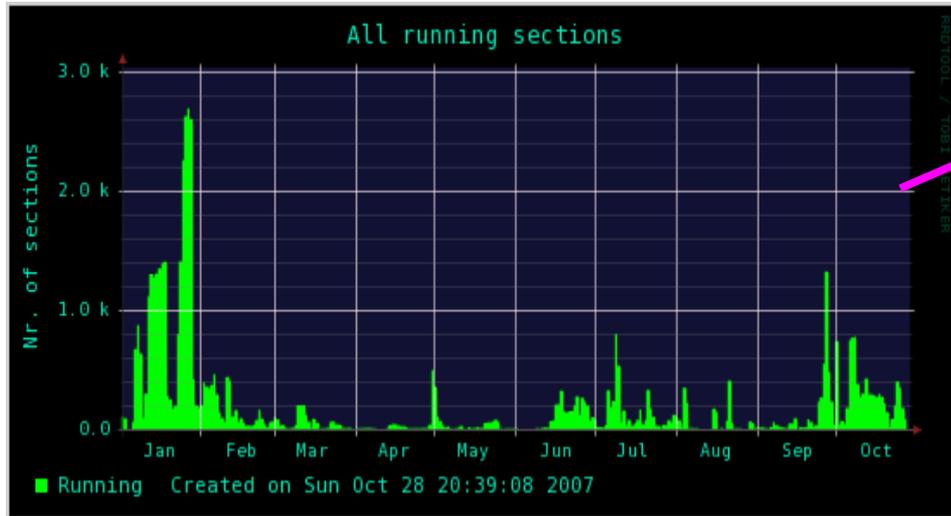
Sites:

UCSD, Wisconsin, MIT
FermiGrid

Soon also:

Florida and McGill

GRID Portals overview: "WMS" based (LCG)



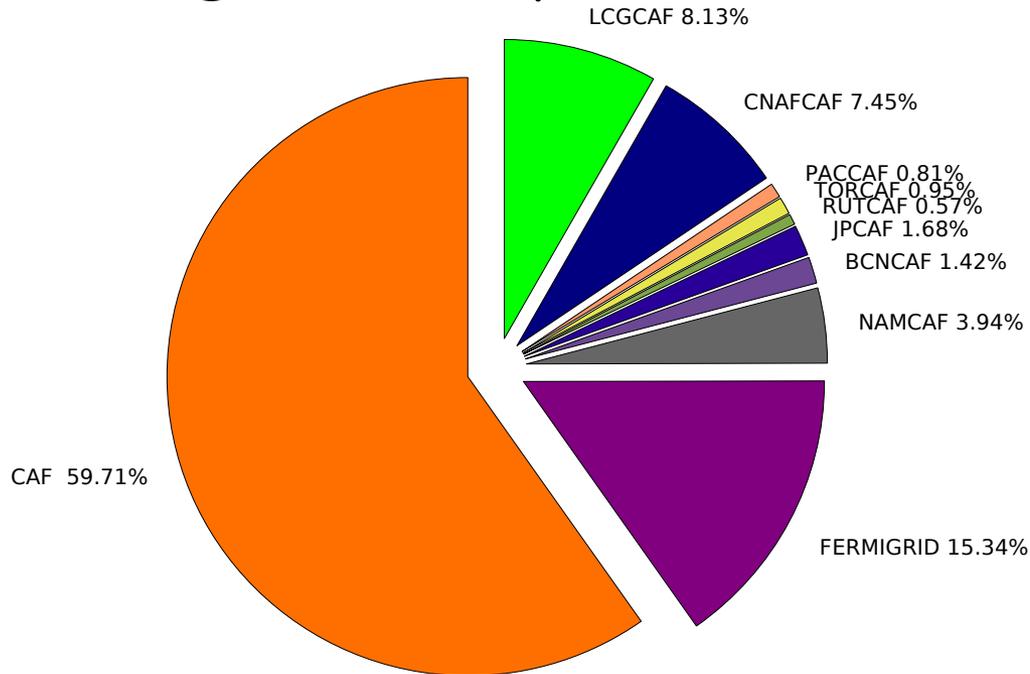
GRID resources availability not predictable -> spikes in running jobs
Some sites asked us "CDF" to use their resources because a running experiments with "real users" helps a lot in debugging site configuration

October 30th

Site	Country
CNAF-T1	Italy
INFN-Padova	Italy
INFN-Catania	Italy
INFN-Bari	Italy
INFN-Legnaro	Italy
INFN-Roma1	Italy
INFN-Roma2	Italy
INFN-Pisa	Italy
FZK-LCG2	Germany
IN2P3-CC	France
IEPSAS	Slovakia
IFAE	Spain
PIC	Spain
UKI-LT2-UCL-HE	UK
Liverpool	UK

Resources Usage: All Farms

Averaged over a year: Jan07-Oct07

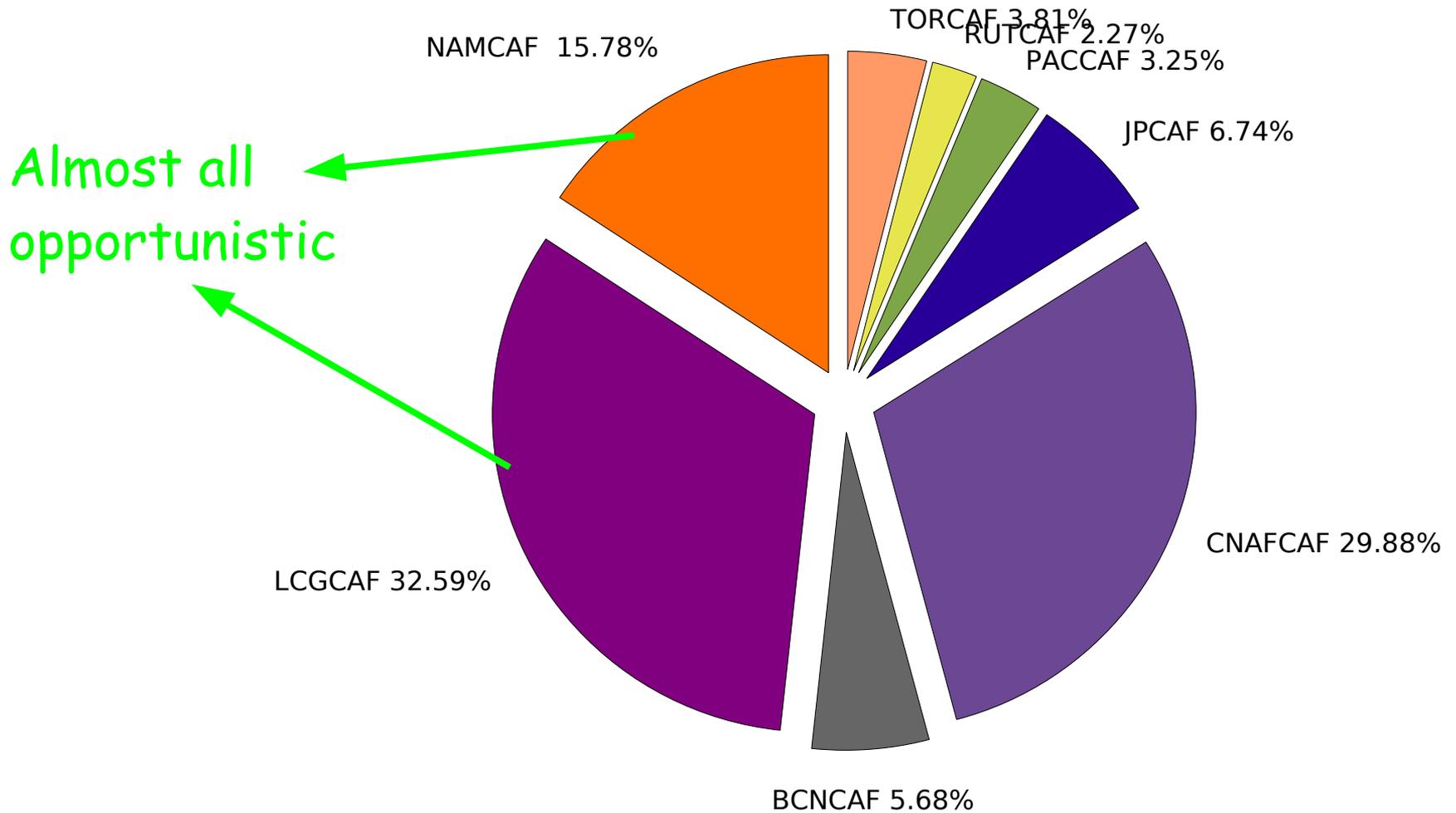


This is an "evolving pie":
thanks to FermiGrid use
Aug-Oct onsite share
has been 50%:50%

Quite soon:

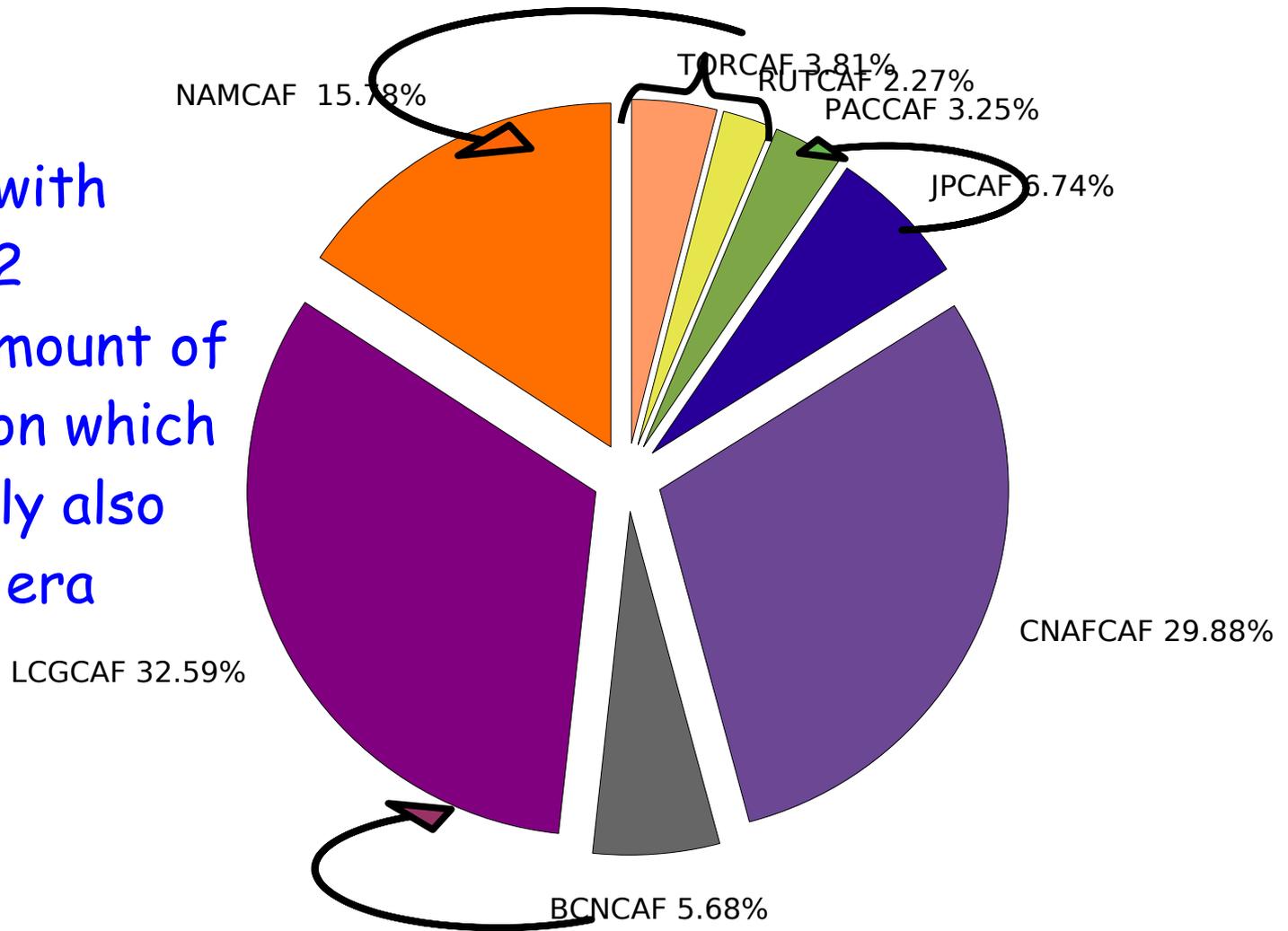
- dcafs will disappear
- caf will be merged in FermiGrid

Offsite Resources Use



Offsite Resources Development

Negotiate with Tier1/Tier2 a certain amount of resources on which CDF can rely also in the LHC era

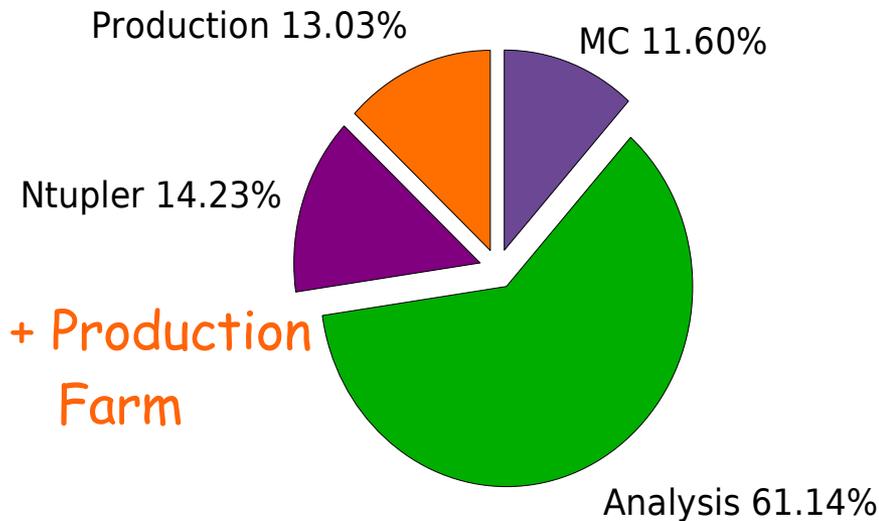


On-site Resources Use

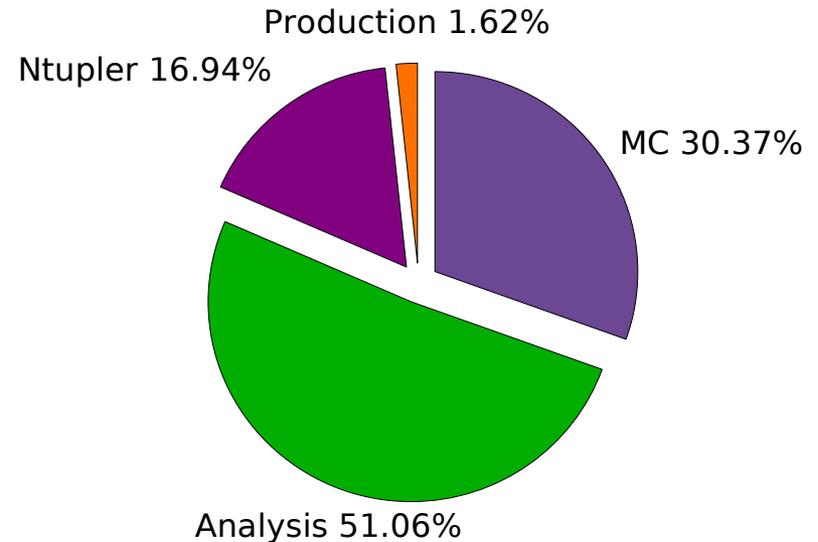
Production: Productionexe
MC: "CDFSim"

Ntupler: STN+top+Bs ntuple
Analysis: all the rest

CAF



FermiGrid



MC has to moved off-site

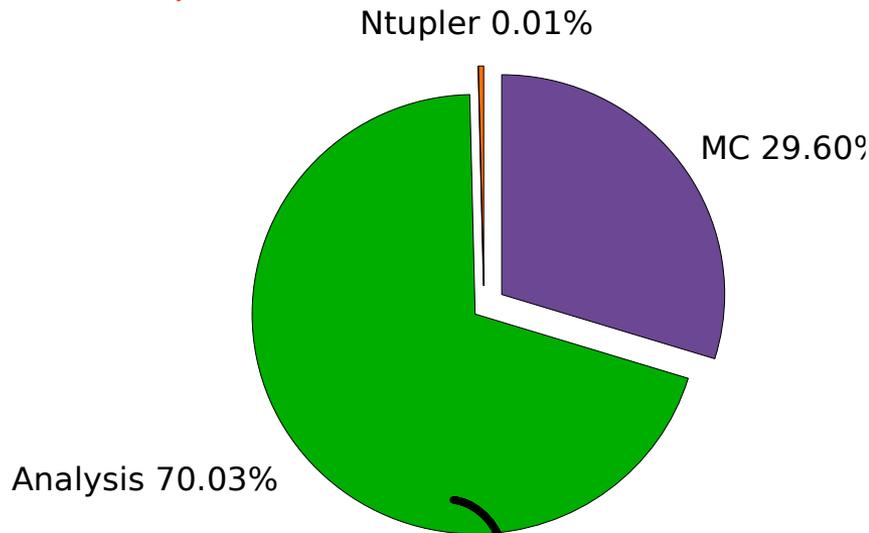
On-site resources mainly dedicated to data production & analysis

Off-site Resources Use

Production: Production.exe
MC: "CDFSim"

Ntupler: STN+top+Bs ntuple
Analysis: all the rest

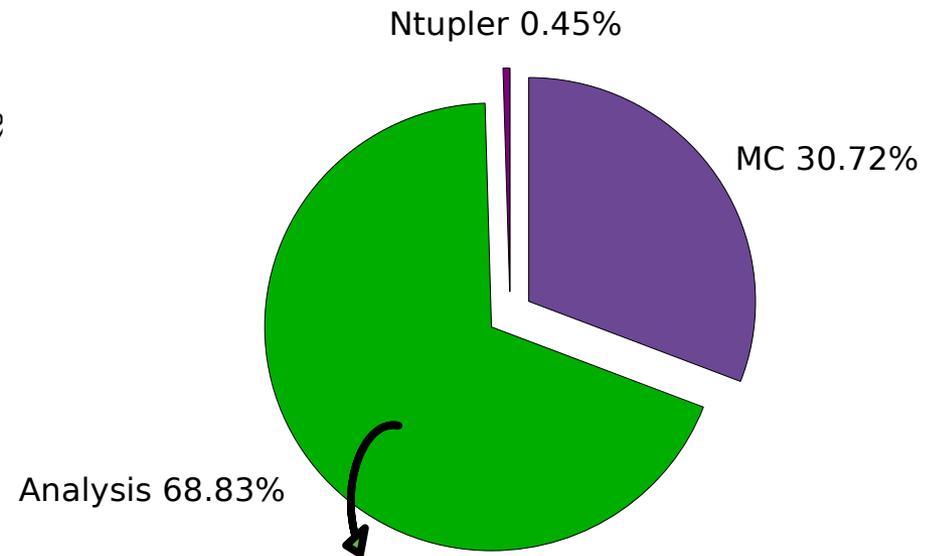
dCAF



Mostly
Pseudo-experiments

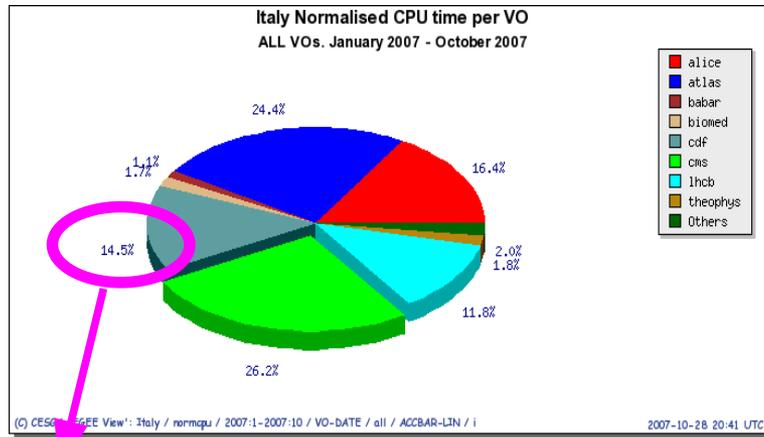
October 30th

CNAFCAF

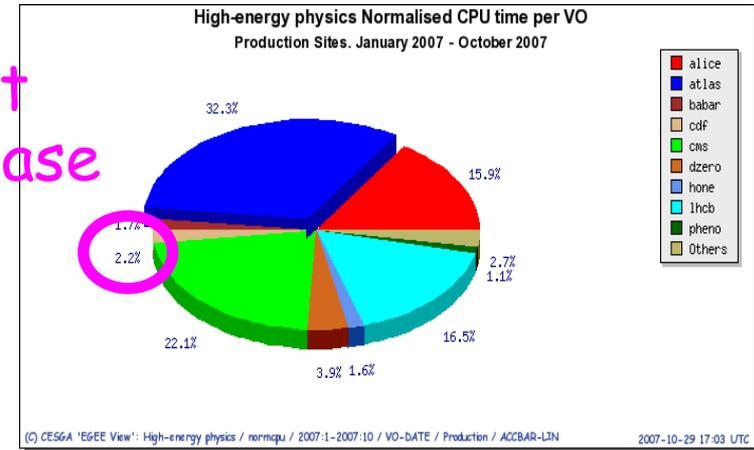


Not negligible contribution
of user analysis

VO Use

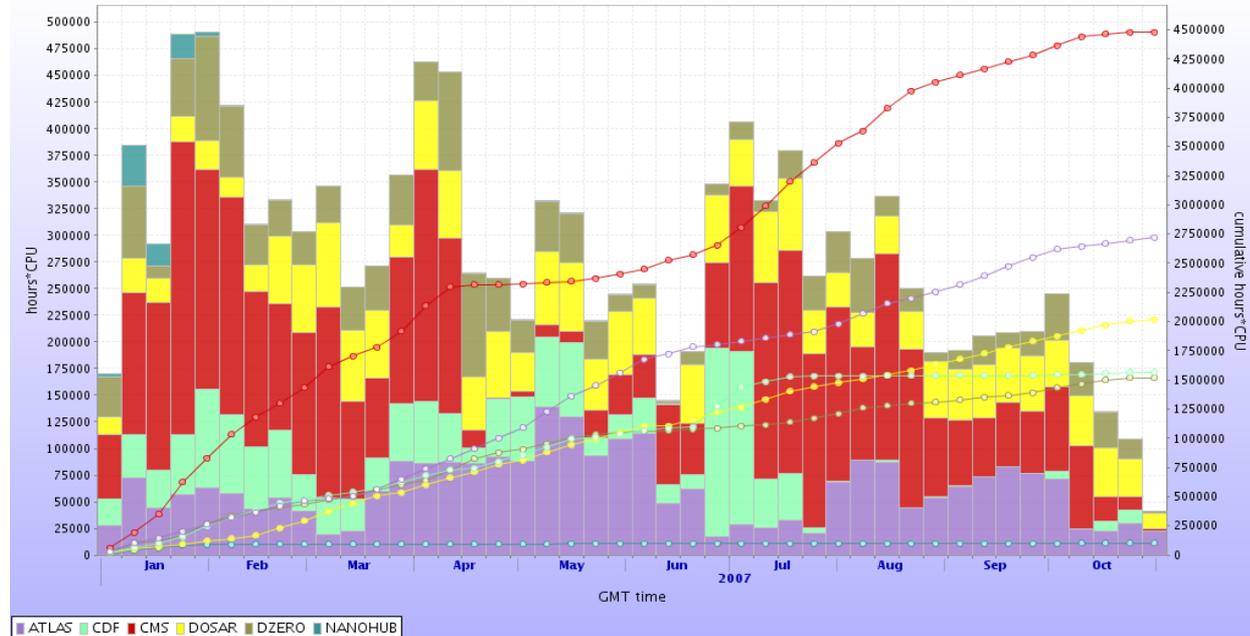


We want to increase



As big as one LHC experiment

Integrated CPU time consumed per VO



October 30th

Summary

CDF computing model evolved and is evolving to cope with GRID. Many progress:

- Reconstruction and Simulation software improved to fully exploit CDF detector and upgrades
- Data production and N-tupling procedure mature
- Moving on-site resources to FermiGrid pool
- GRID resources use a reality, need to increase it

GRID resources are critical for our Physics Program and we would like to secure a CDF quota