



Computing Highlights, Status and Plan

Donatella Lucchesi
for the CDF Offline group

International Financial Committee
November 7th 2008



Introduction

Quoting myself one year ago:

" *CDF computing model evolved and is evolving to take advantage of GRID: **dedicated farms** --> **GRID pools** "*

Outline

- Current status for data production
- Monte Carlo generation update
- Resources usage
- Projects in progress
- Summary

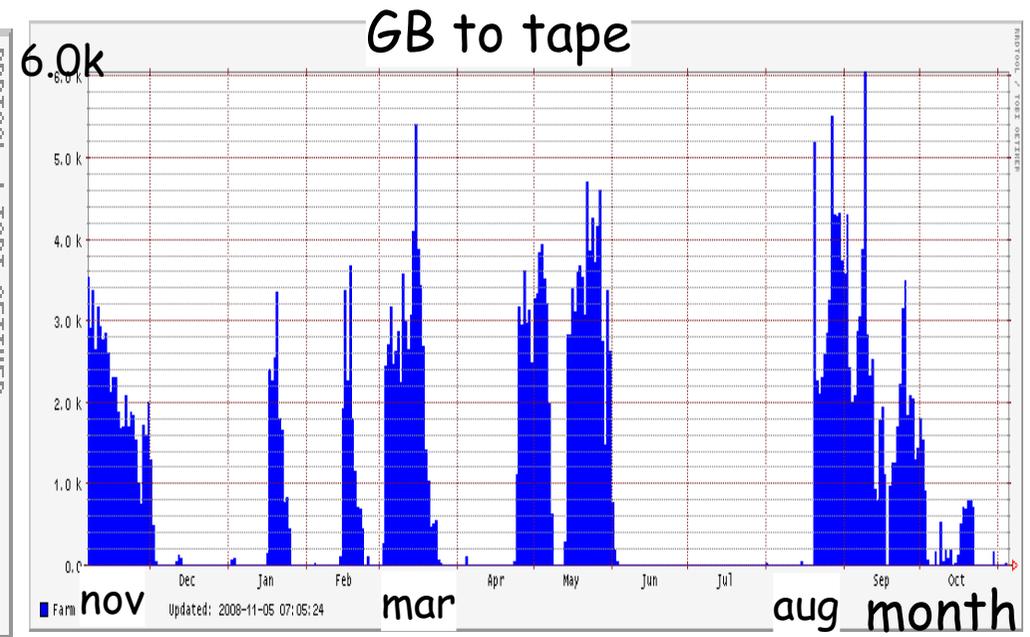
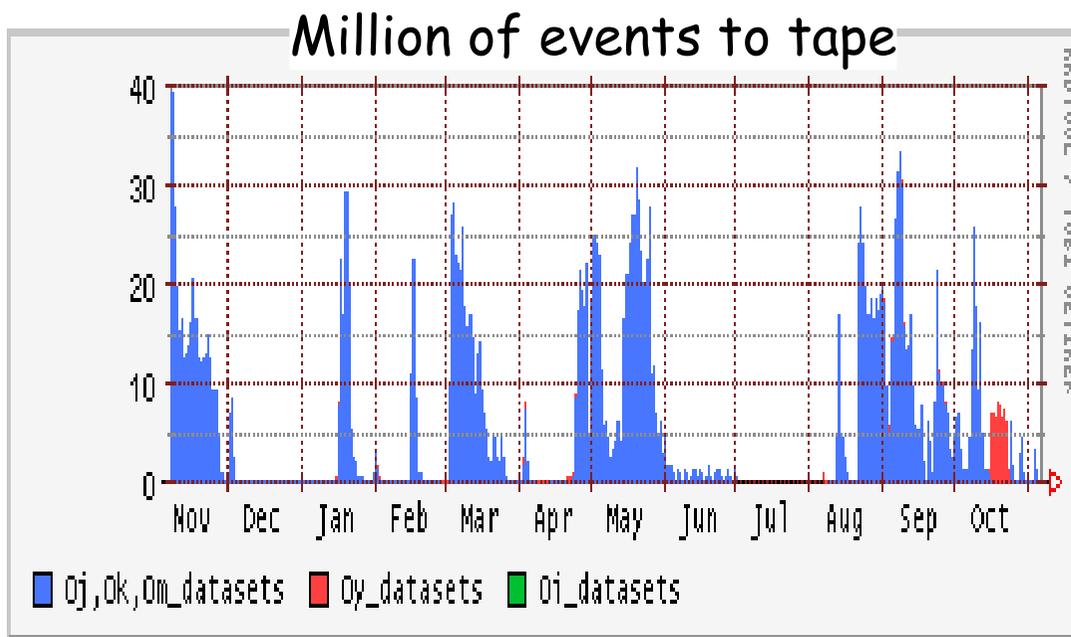


Data Processing

Data reconstruction uses the established "one-pass processing mode"

Calibration -> Production -> N-tupling

- Data split up in "periods"
- Average processing time ~2 sec/ev.





Data Processing cont'd

Data available to users in about 10-12 weeks

Working to

- reduce delays
- improve automation to reduce human operation load

Data delivery for recent run periods

Period	Start	End	Lum (pb ⁻¹)	Events (M)	N-tuples ready
13	May 13, 07	Aug 4, 07	317	545	Nov 29, 07
14	Oct 28, 07	Dec 3, 07	45	59	Feb 21, 08
15	Dec 5, 07	Jan 27, 08	159	210	Apr 7, 08
16	Jan 27, 08	Feb 27, 08	142	168	May 21, 08
17	Feb 28, 08	Apr 16, 08	188	235	Jun 6, 08
18	Apr 18, 08	Jul 1, 08	407	436	Oct 25, 08



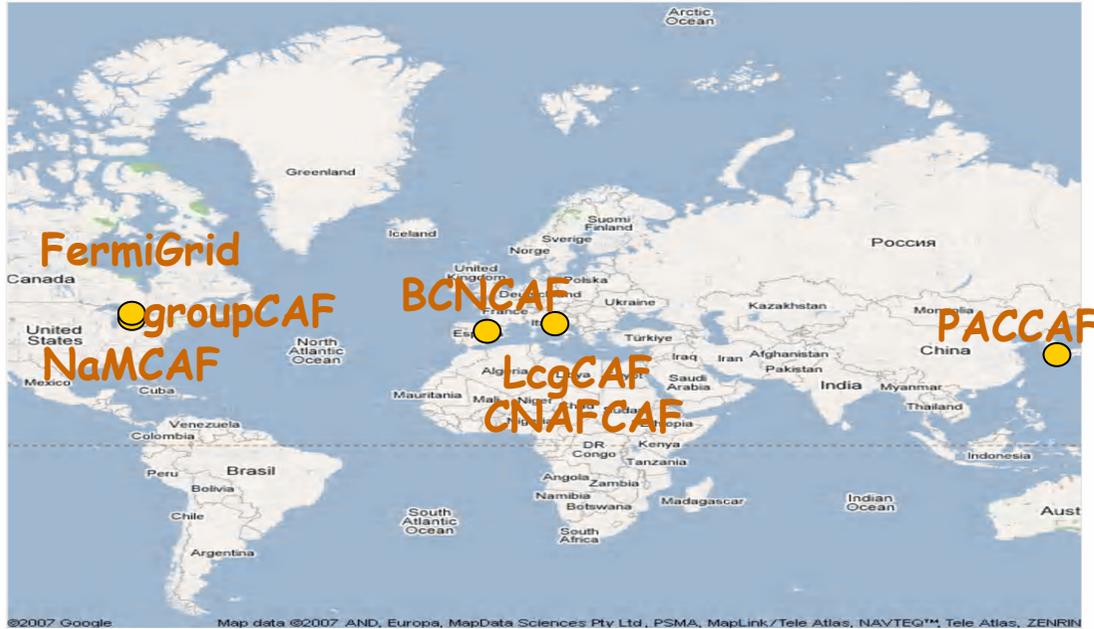
Monte Carlo Production

- The "old" MC data production model: Run-Based
 - ✗ Take into account detector configuration and luminosity
 - ✗ Require continuous production operations coordinated with data
- Changing the production model for new MC
- The new MC production model: Luminosity Profile Scaling
 - ✗ Generate MC asynchronously with data taking
 - ✗ Allow better scheduling of CPU usage
 - ✗ Significantly reduces amount of MC needed relative to run-based approach
- Possible because the detector configuration is very stable

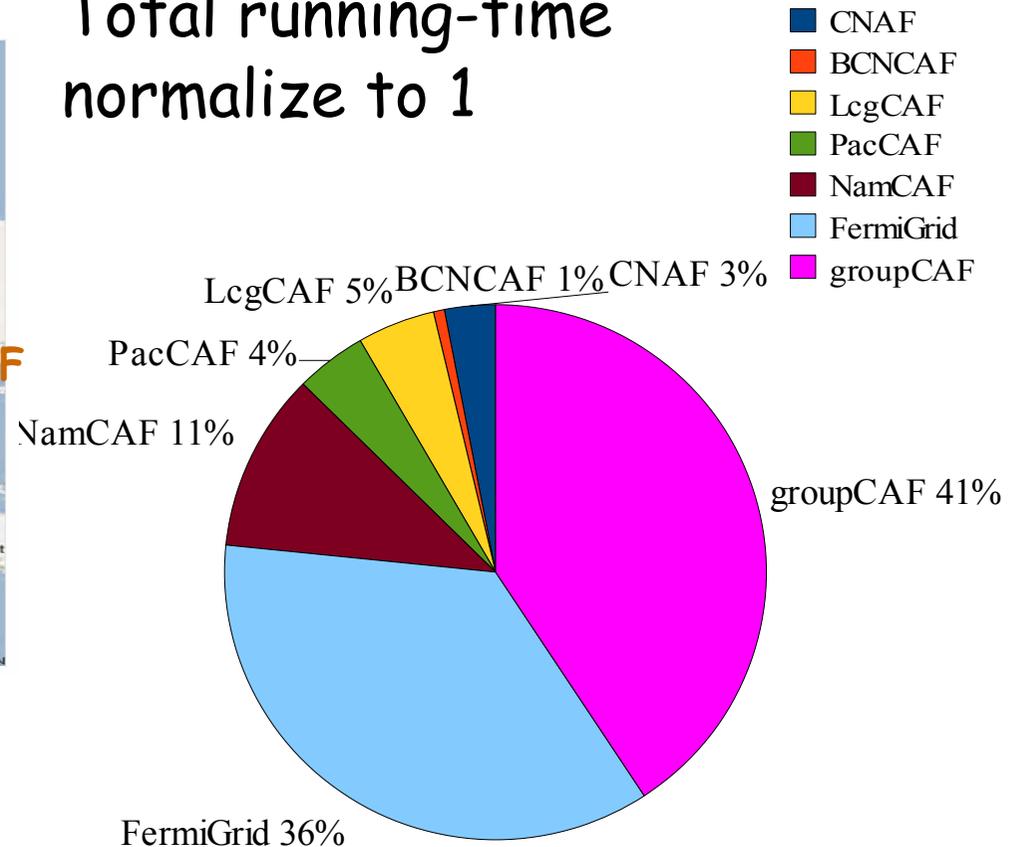


CPU's Resources Usage - Total

Farms we access for the last year



Total running-time
normalize to 1



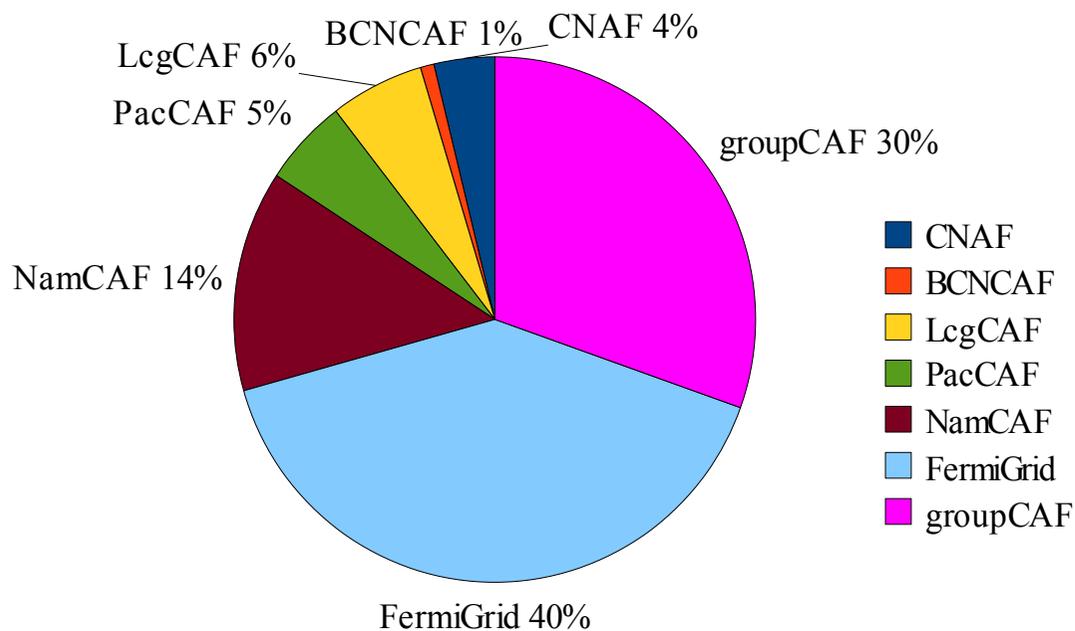
This includes:

- Data processing
- Monte Carlo generation

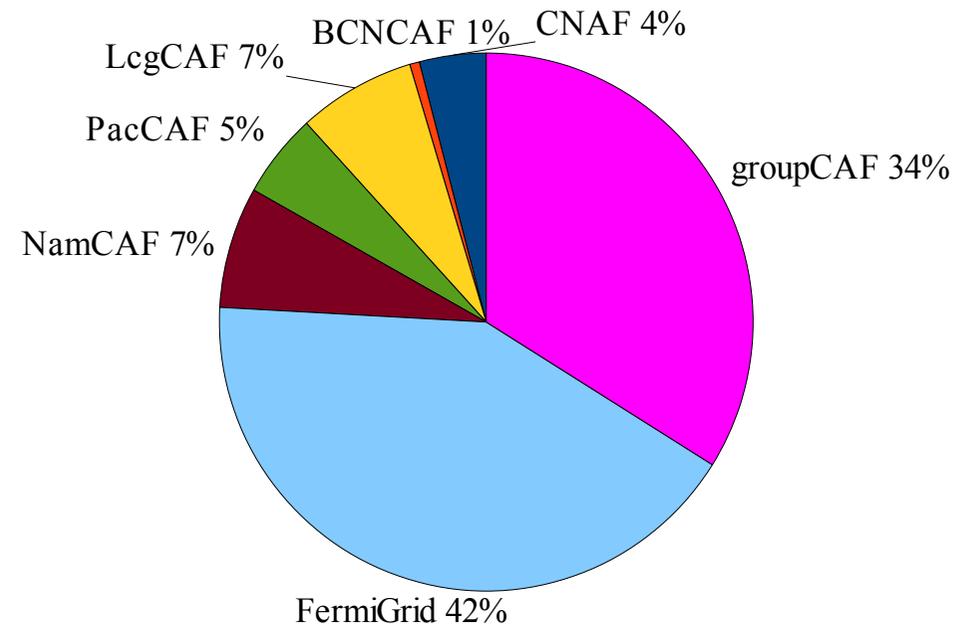


CPU Resources Usage - MC Prod & Users

Running-time for MC production and users jobs normalize to 1



Running-time for users jobs normalize to 1



About 20% of computing resources comes from off-site

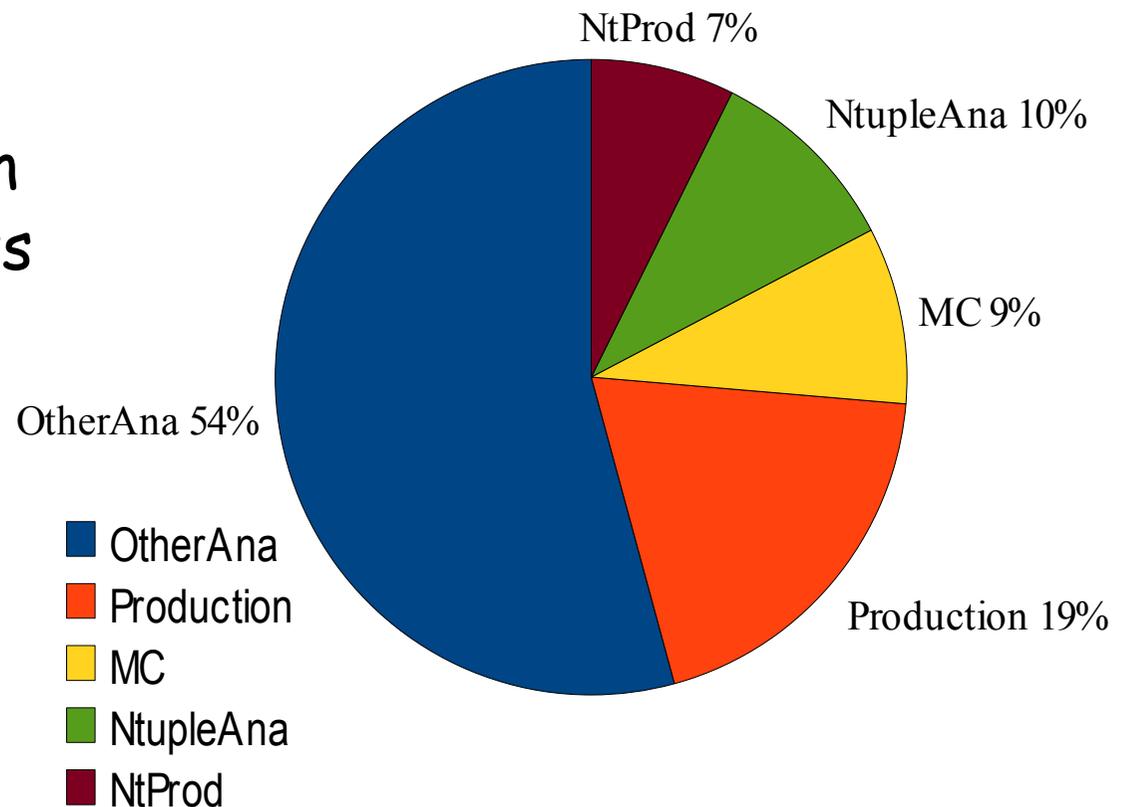


CPU's Resources Usage Analysis

Off-site

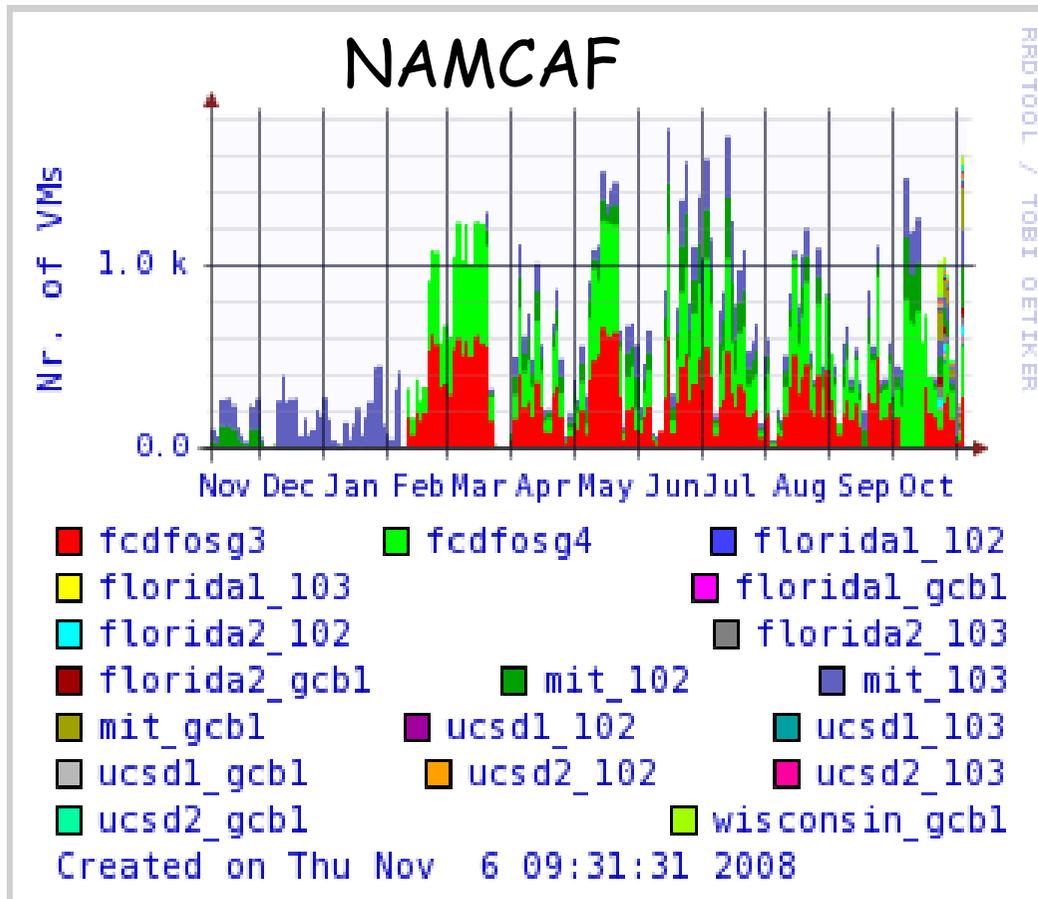
CPU's used mainly for Monte Carlo Production or pseudo-experiments generation
CNAF exception: some datasets available to users for analysis and re-processing for private detector and/or algorithms studies

On-site

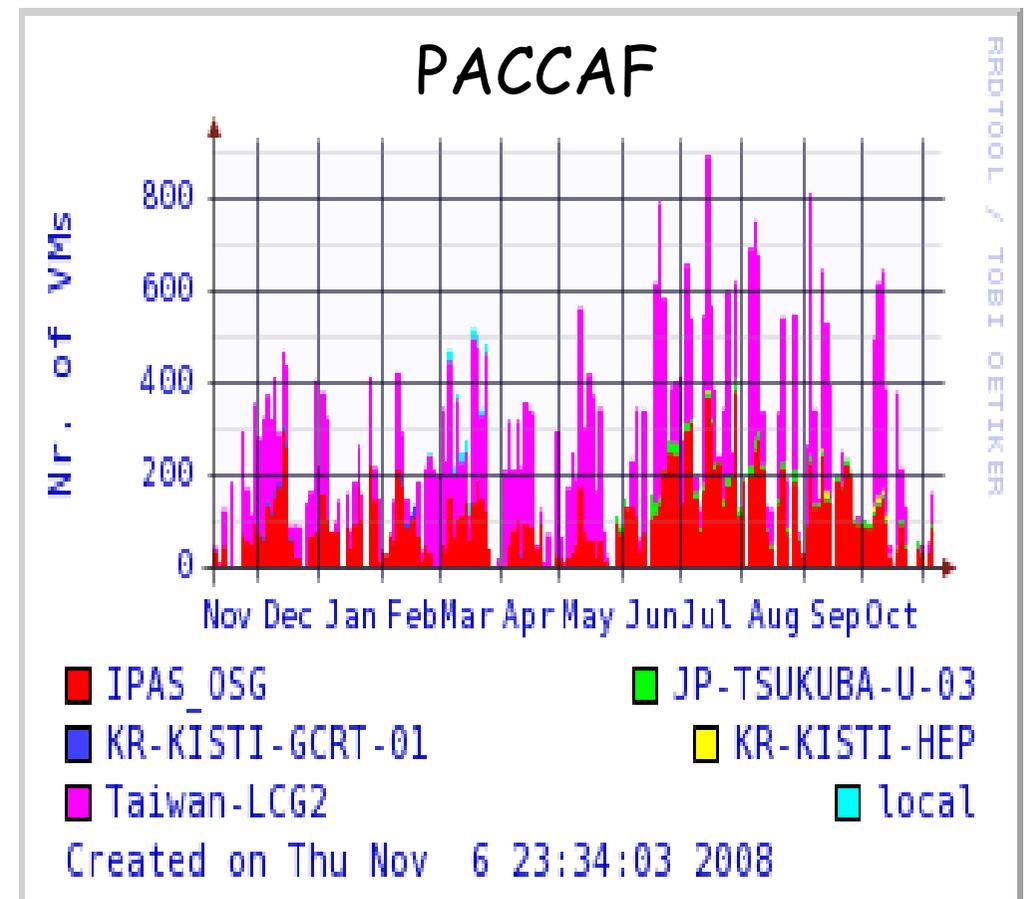




Off-site Resources: USA and Asia



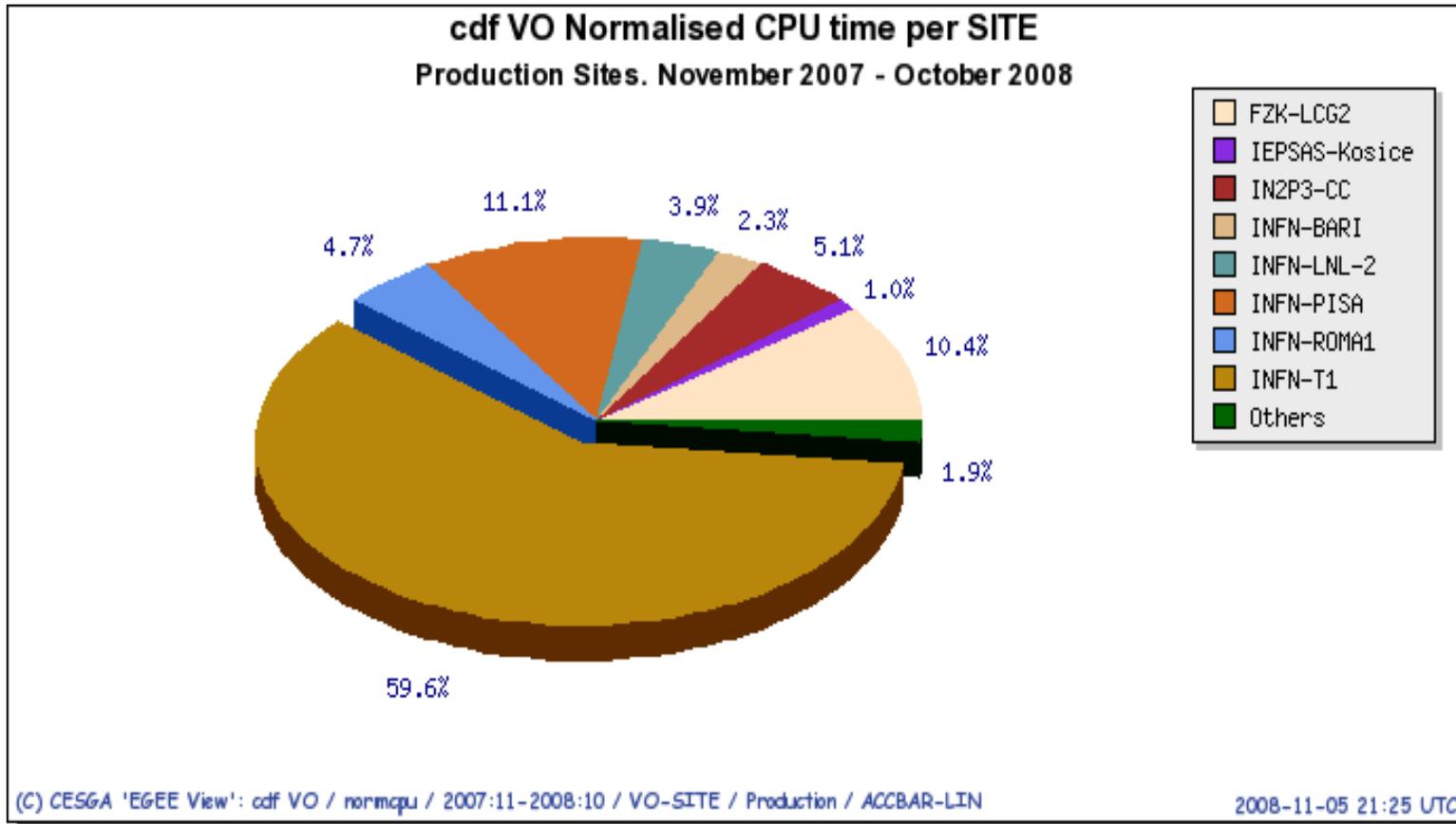
Few opportunist resources until Sept, now improving



Need to increase the usage in particular with KISTI



Off-site Resources: Europe



An important fraction comes from opportunist resources but we should or would like to improve usage of big sites



CDF moves to the Grid

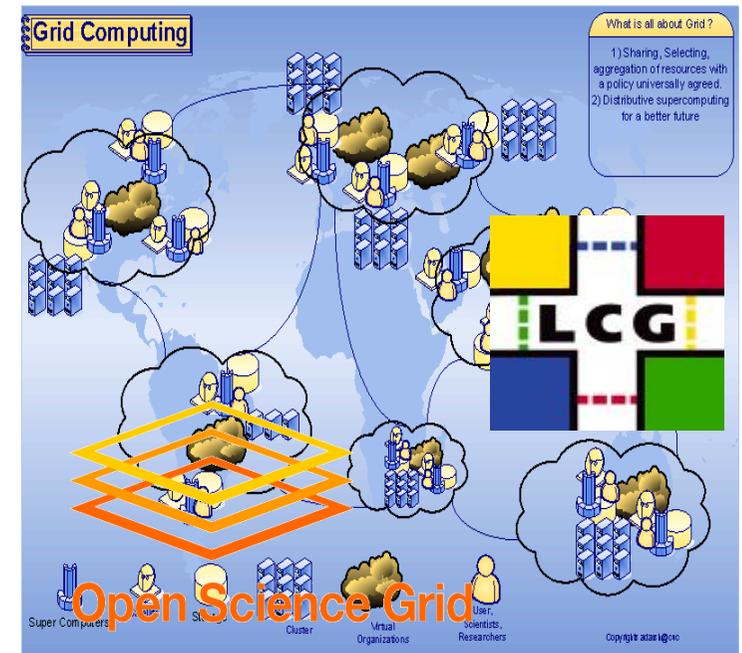
Several years ago CDF started to moving to Grid in its own way, ie. adapting the CDF structure to the Grid middleware.

Not so easy...

- ✓ Grid is distributed across many regions
- ✓ Grid has several flavors
- ✓ Sites deploy middleware following local requirements
- ✓ Middleware changes often (at begin in particular)

CDF uses two middlewares, OSG and LCG:

- cpus access
- data transfer





CPU's Resources Access

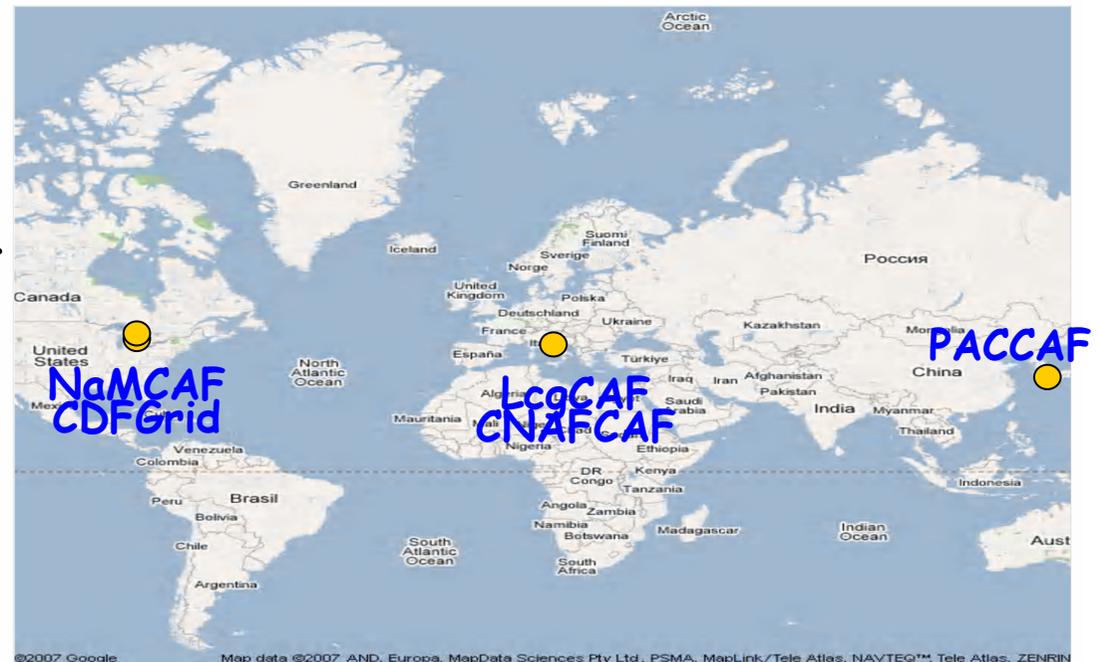
CDF went from a model with several dedicated farms to a model with few portal to access resources located in Tier centers.

USA: groupCAF and FermiGrid merged to CDFGrid to access FNAL T1
NAMCAF portal to OSG resources

EU: BCNCAF, LyonCAF
merged to LcgCAF
CNAFCAF to access CNAF T1

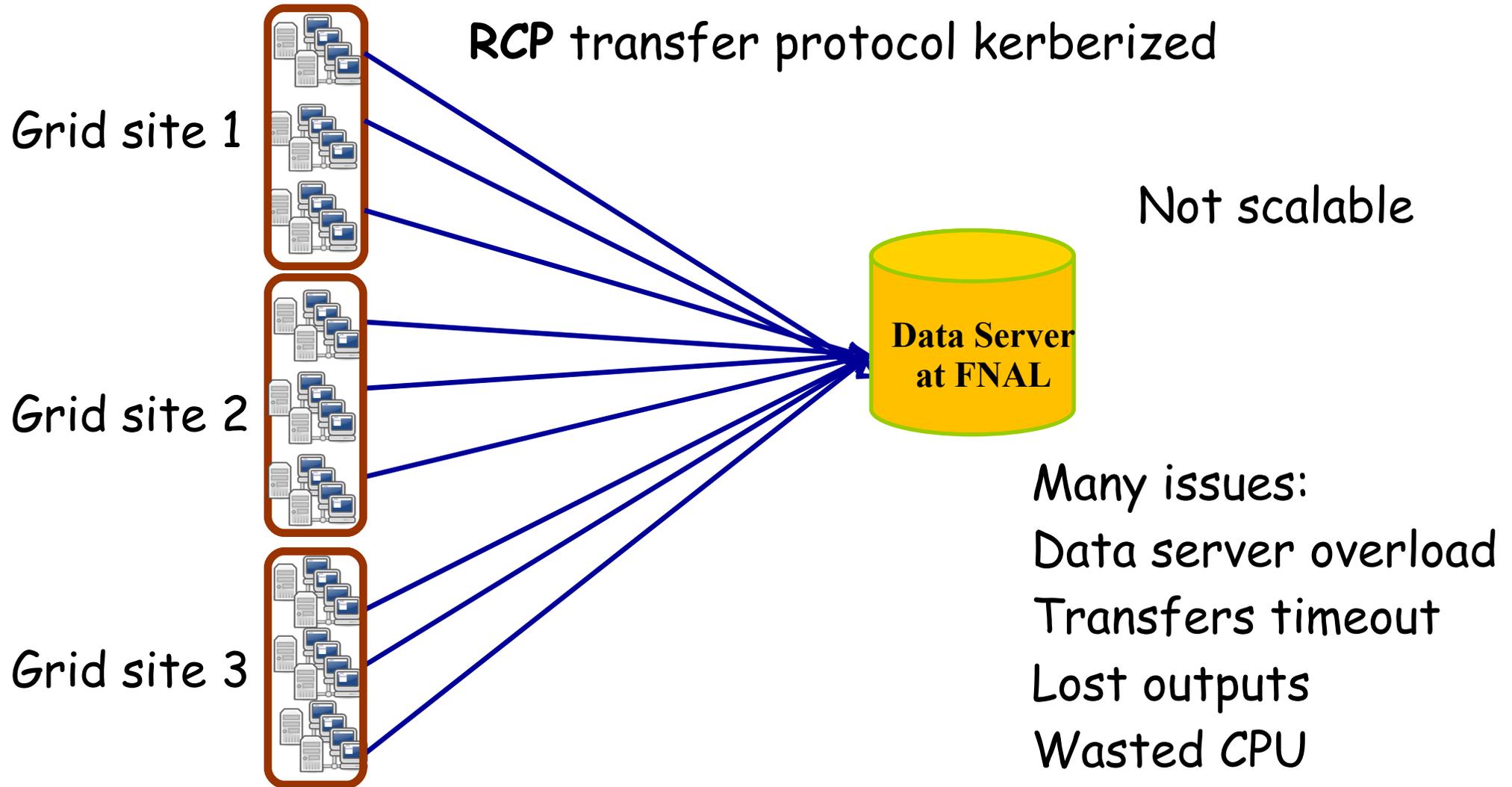
All this accomplished or close to
this year

Asia: KorCAF, JPCAF merged to
PACCAF



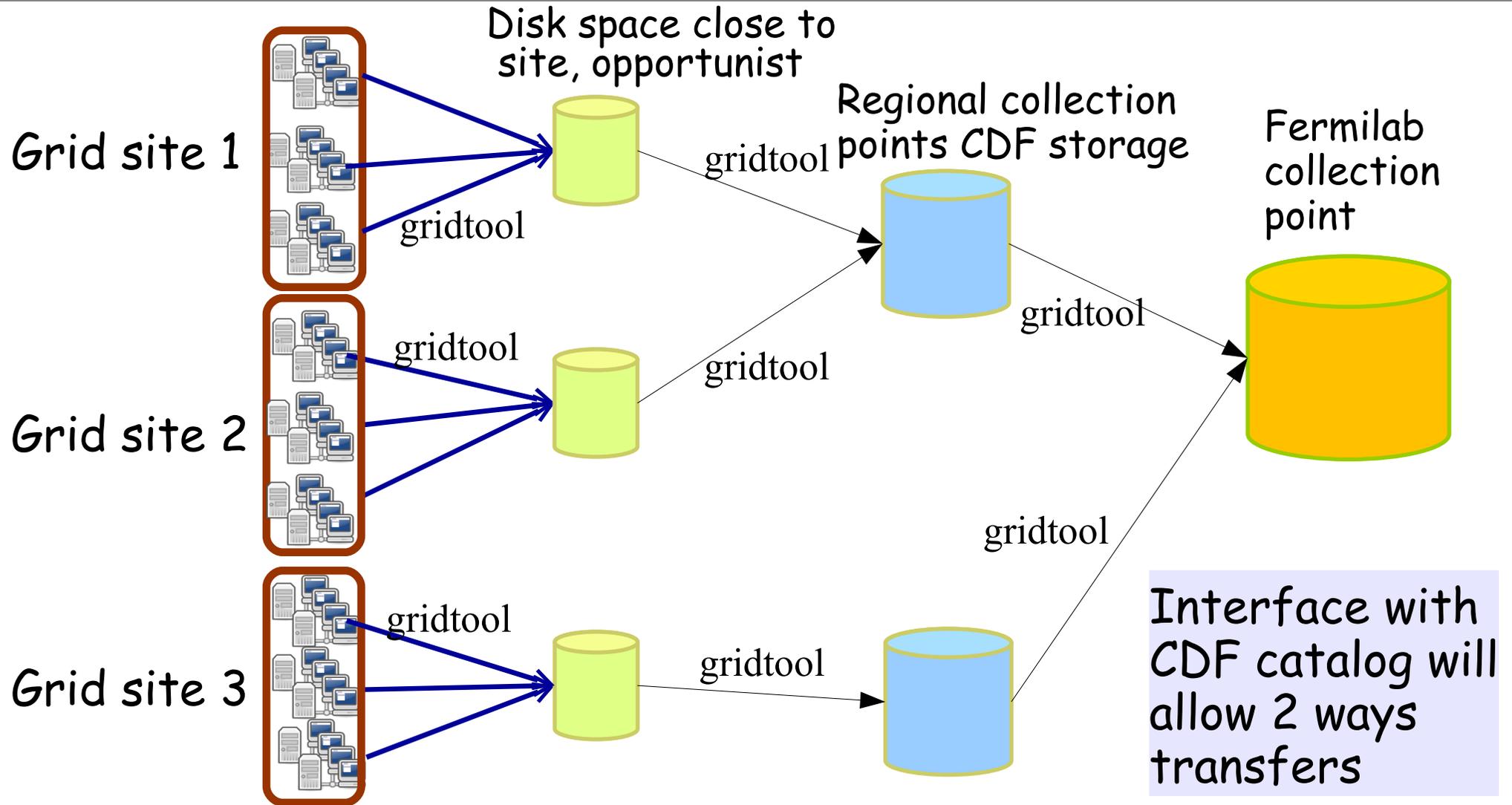


Current Data Transfer Model





New Data Transfer Model





Summary

- Data process provide users with ntuple for analysis in 12 week work to improve automation and reduce human operation load
- Transition to Grid close to completion:
 - ✓ cpus resources access complete by end of 2008
 - ➔ reduce human operational load
 - ➔ provide easy access to large number of OSG/LCG sites
 - ✓ disk space access via Gridtools in progress
 - ➔ grant access to opportunistic storage
 - ➔ speed up data transfer
- Monte Carlo production off-site

Grid resources are important to have physics results out in time for conferences we want to increase the CDF quota