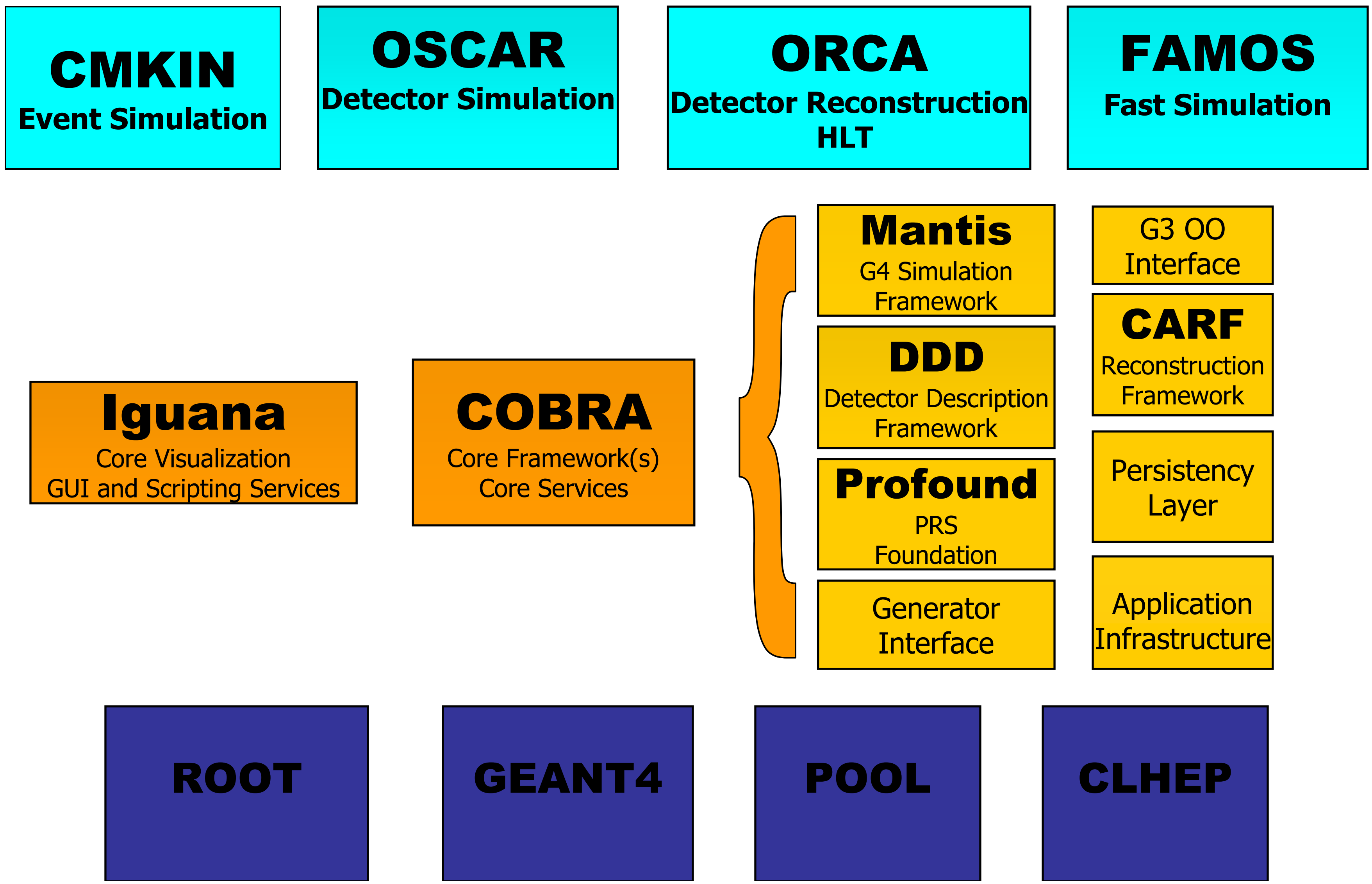
CMS Software and Computing

Claude Charlot
LLR-École Polytechnique, CNRS&IN2P3

Outline

- Software tools
  - Simulation
  - Reconstruction
  - Fast simulation and reconstruction
- Distributed computing & the Grid
  - DC04
  - Tools for distribution, publication, analysis

Simulation, Reconstruction, Analysis Software Structure

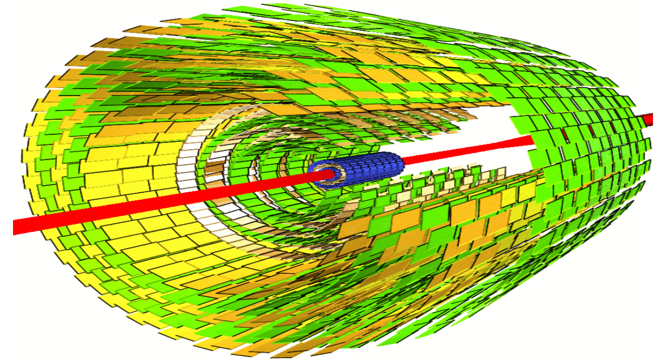


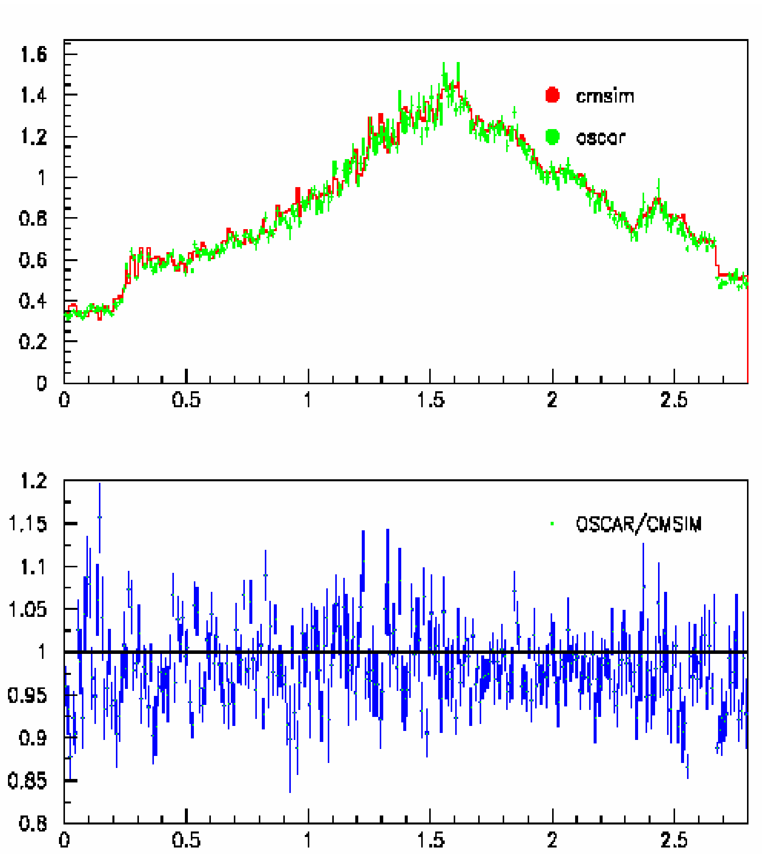
OSCAR simulation

- Full simulation based on GEANT4
- Generator interface CMS
- Detector geometry including detailed description of the magnetic field
- Physics processes for electro-magnetic and hadronic interactions
- Hits persistency
- CMS now changing from CMSIM Geant3 based simulation to OSCAR
  - A lot of comparison with cmsim and test-beam data

Tracker simulation in OSCAR

Detailed description of all active and passive components; material budget





Critical requirements for physics studies with tracker

Correct, navigable Monte Carlo truth (particle, track, vertex, history) with trace-ability of initial primary particle

Special treatment of hard brem with the assignment of new track for electron above threshold (500 MeV)

⇒ Extensive validation in terms of tracking and hit distributions

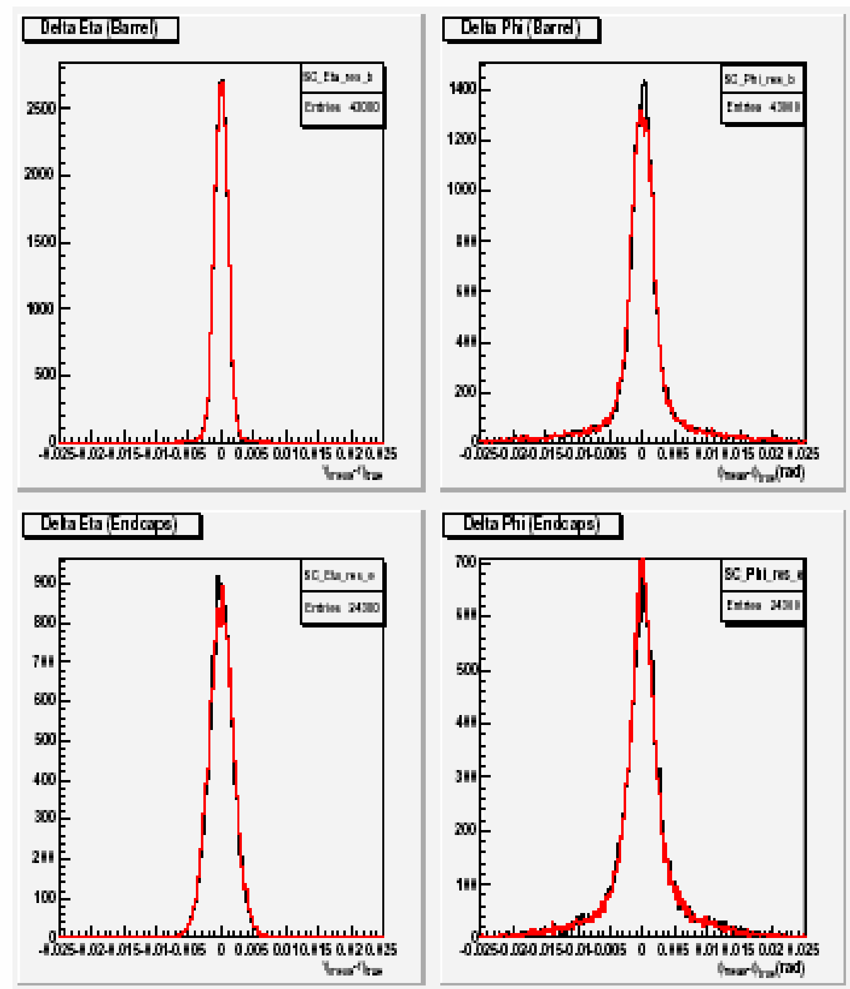
Electromagnetic Calorimeter (ECAL)

Comparisons with CMSIM/G3

Position resolution

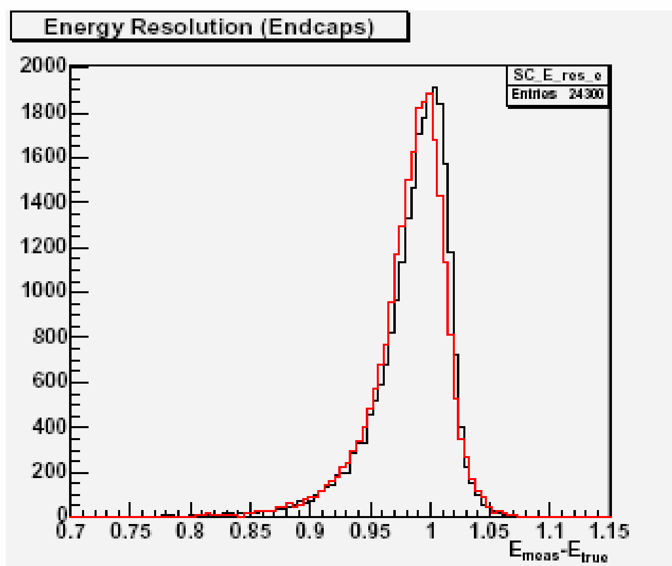
- Energy and position resolution, shower
- Hadronic showers
- Level-1 e/m trigger response
- Preshower response
- Performance studies

Red – OSCAR_2_3_0_pre5, black – CMS132



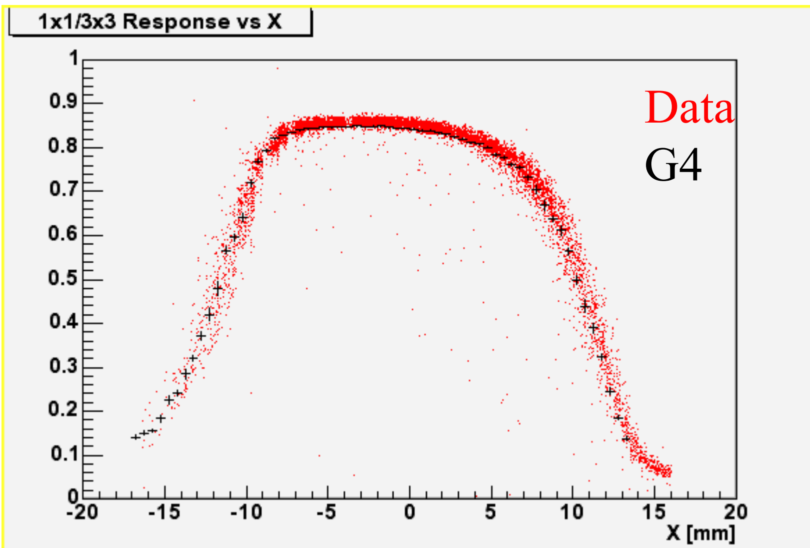
Energy resolution

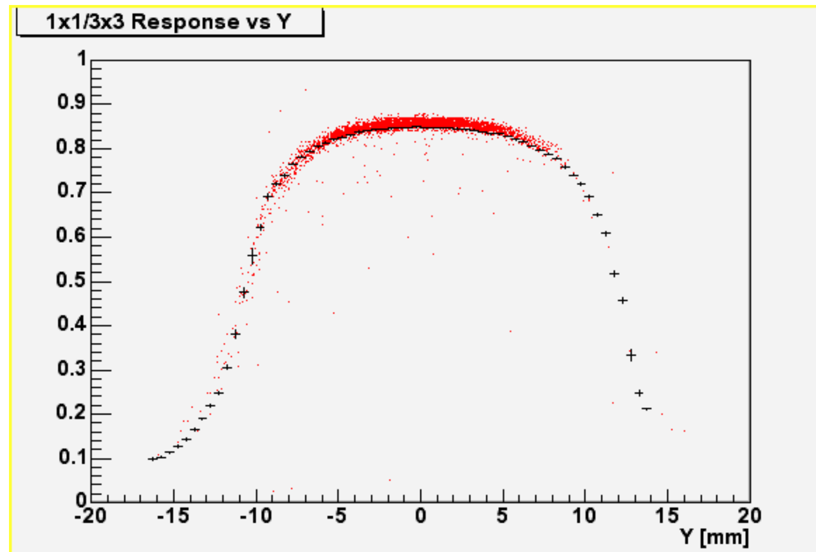
Endcap

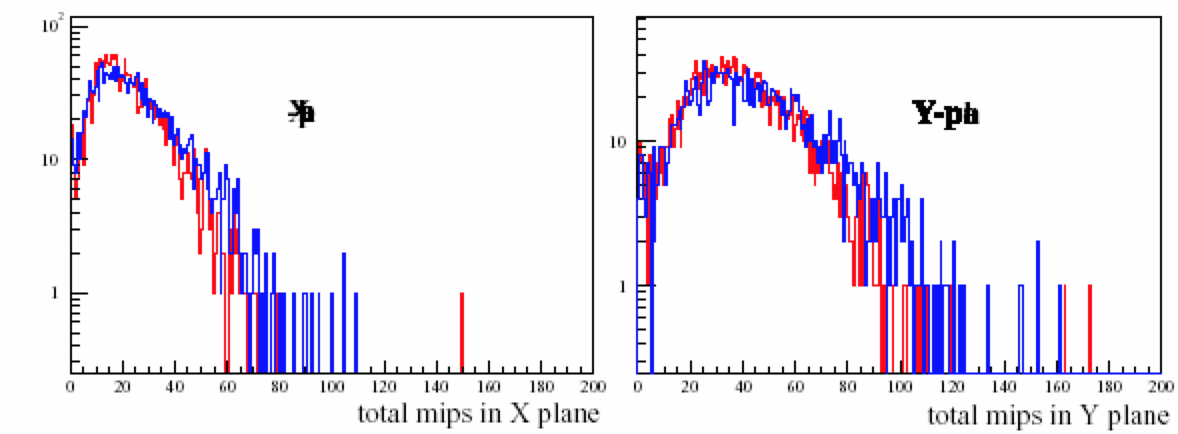


ECAL: Geant4/testbeam comparisons

Single crystal containment: E_{1x1}/E_{3x3} versus position

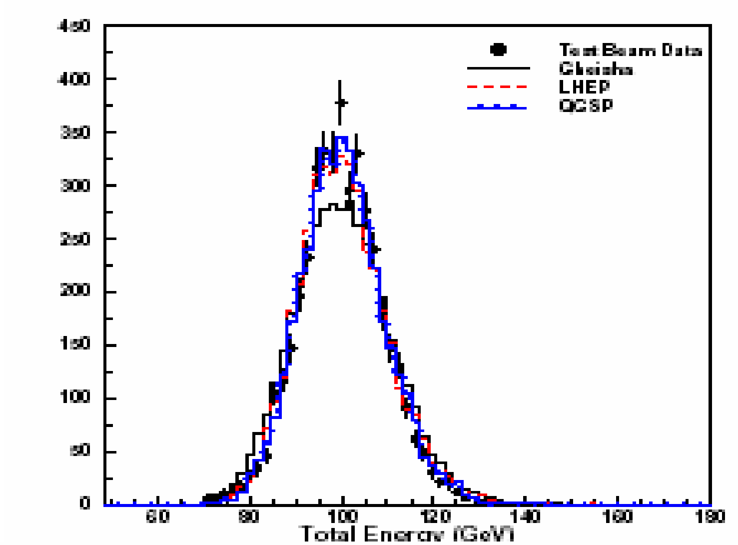


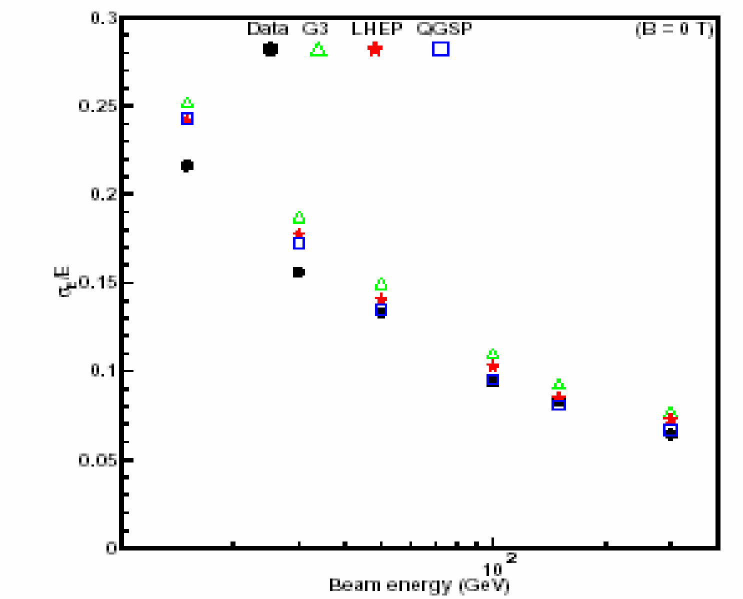




Hadronic Calorimeter (HCAL)

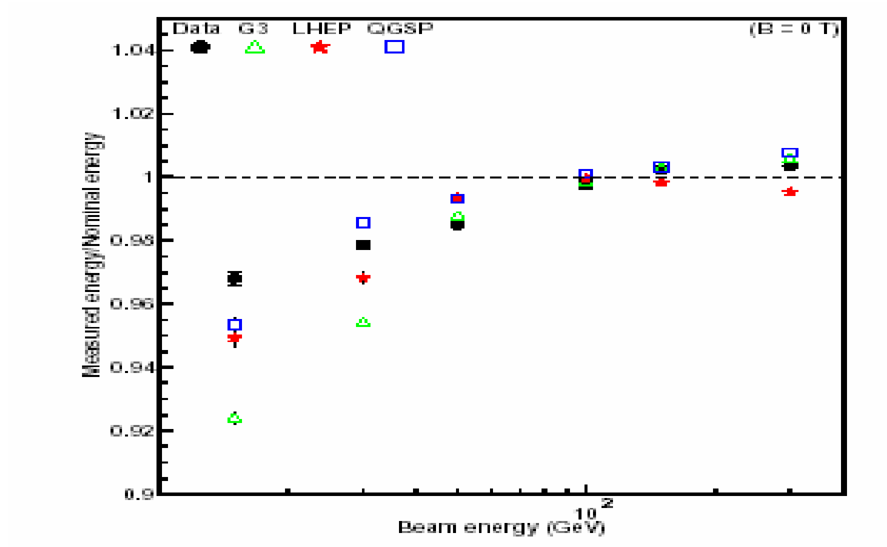
Energy resolution





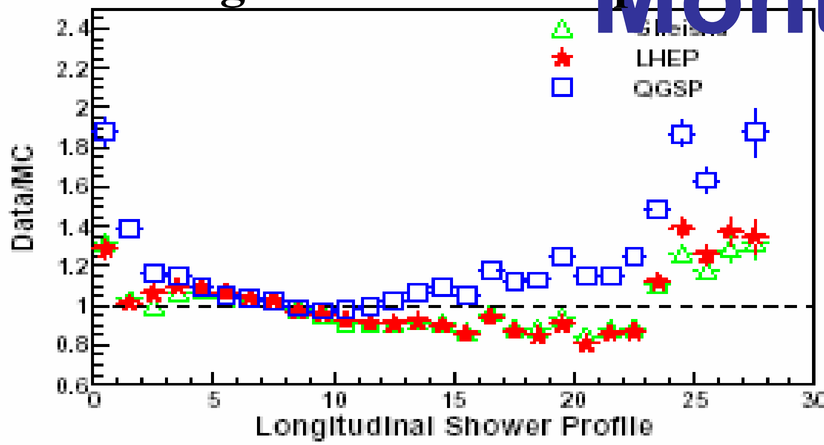
Extensive validation program with comparisons to G3 and several test beam data sets, incl. combined ECAL-HCAL runs; also in context of LCG simulation physics validation project

Non-linearity in energy response

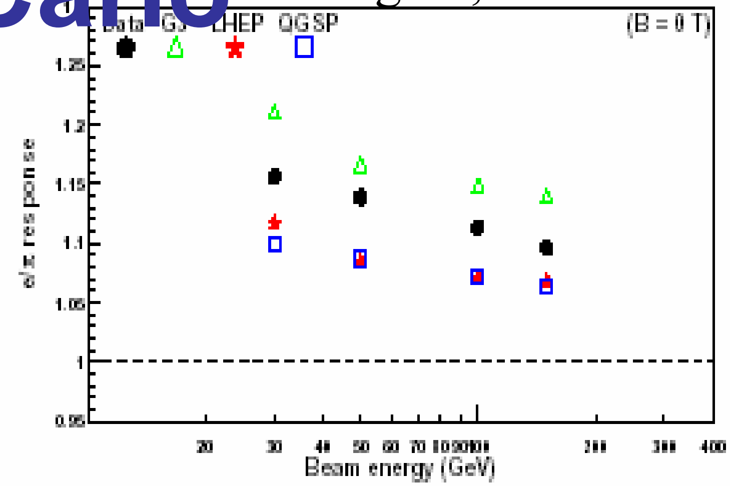


HCAL: Testbeam vs MonteCarlo

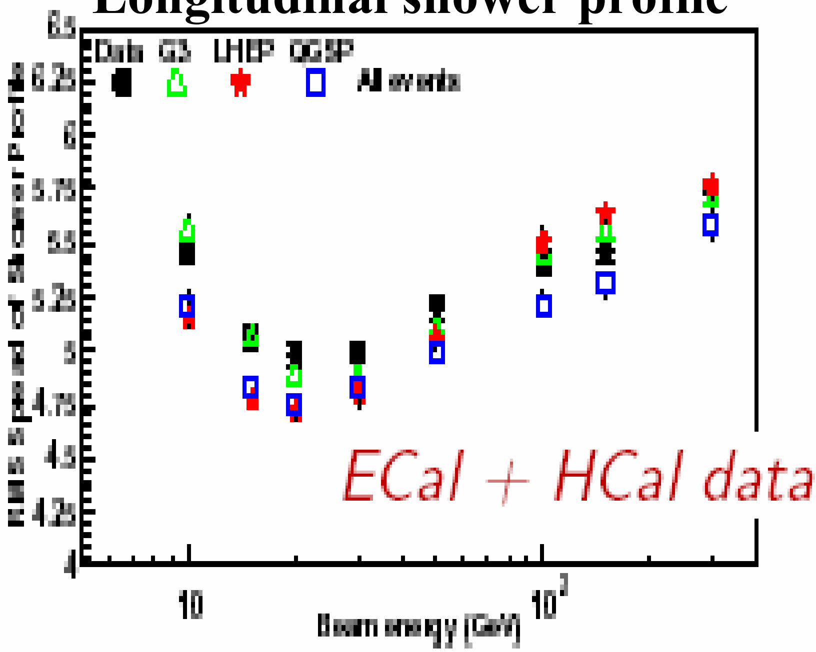
Longitudinal shower profile



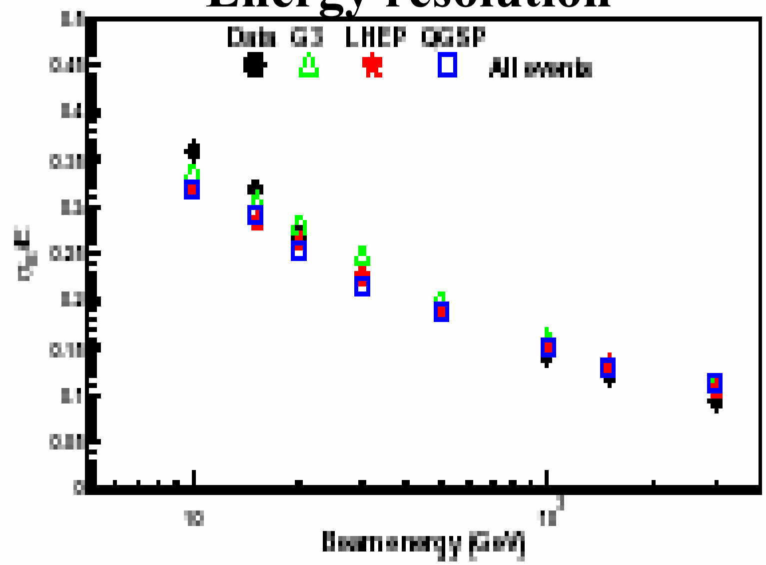
e/π: G3 ~3% higher, G4 ~4% lower



Longitudinal shower profile



Energy resolution



ORCA Reconstruction

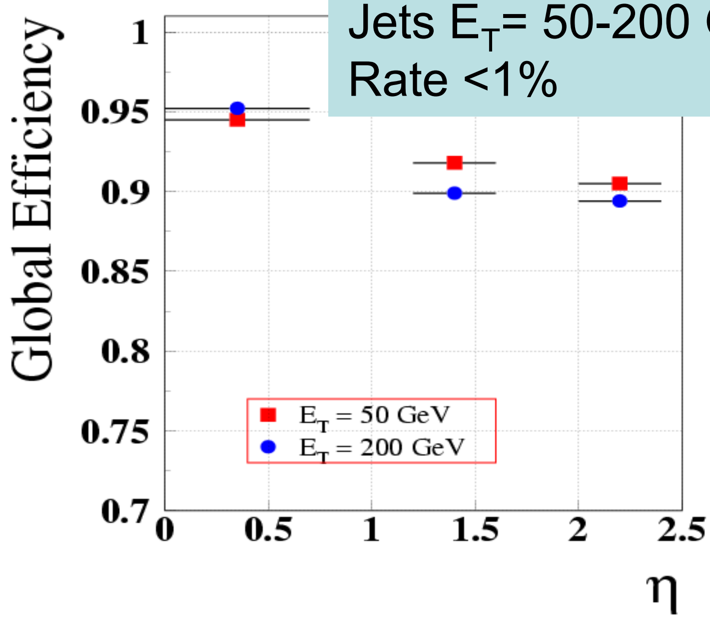
ORCA 8.1.3:
- 22 subsystems
- 373 packages
- ~6000 files

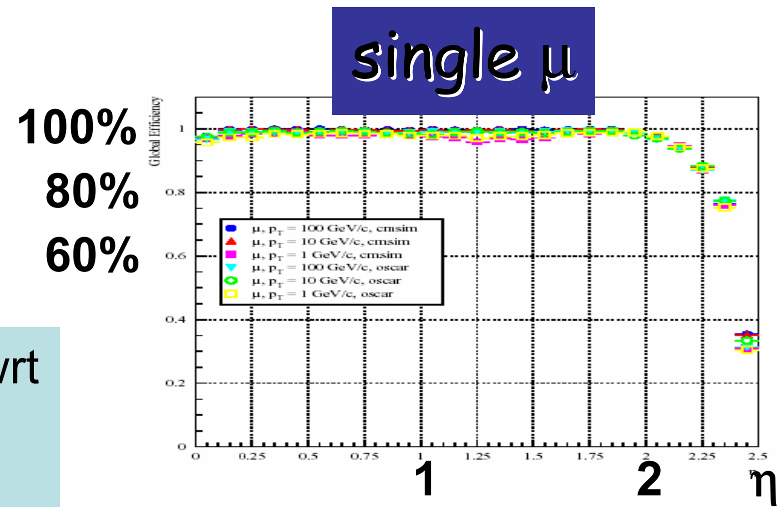
- Object Oriented Reconstruction
  - Based on COBRA
  - Using POOL and SEAL LCG software
- Detector digitization: combination of signal and pileup events followed by simulation of electronics
- Simulation and eventual validation of trigger decisions
- Reconstruction of detector objects
  - Tracks, Clusters, Vertices,..
  - The algorithms for use in High Level Trigger
- Offline reconstruction of Physics Objects
  - Jets, Electrons, Photons, Missing Et

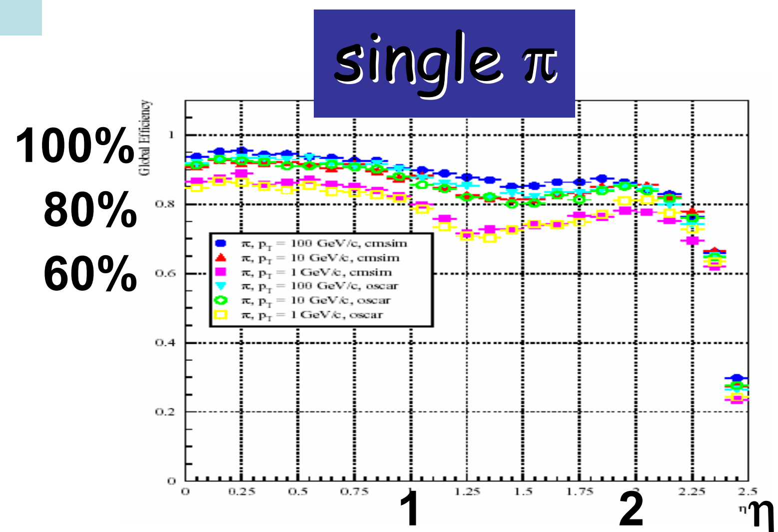
Track Reconstruction

- Multiple algorithms available
  - Kalman Filter, DAF, Gaussian Sum Filter.
- Track Object persistent for DST

No significant degradation wrt single pions.
Jets E_T= 50-200 GeV: Fake Rate <1%

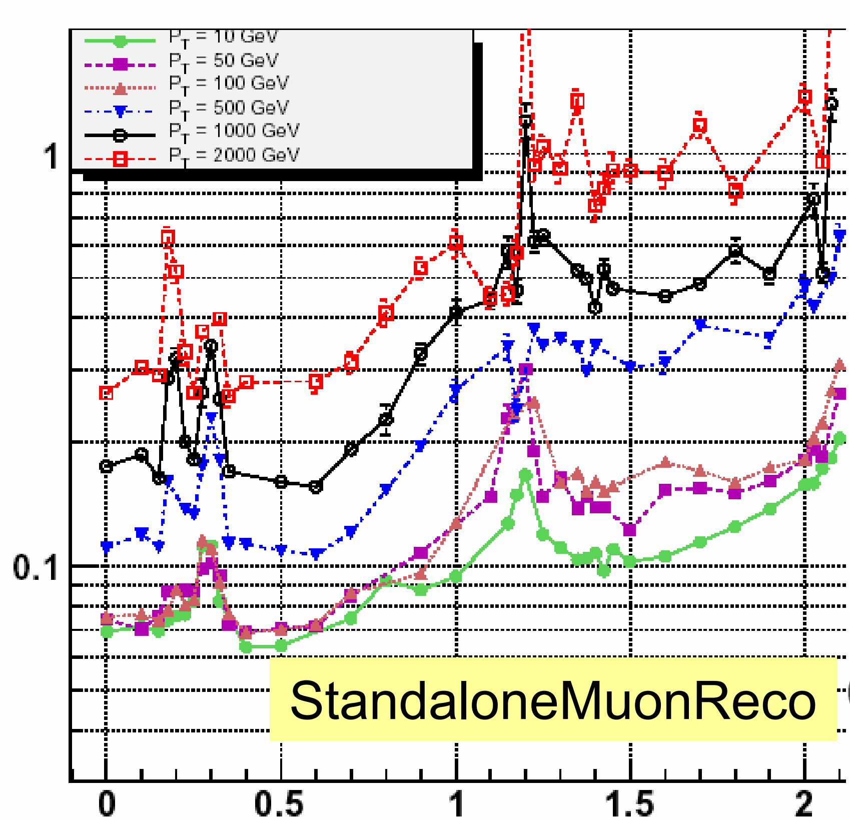


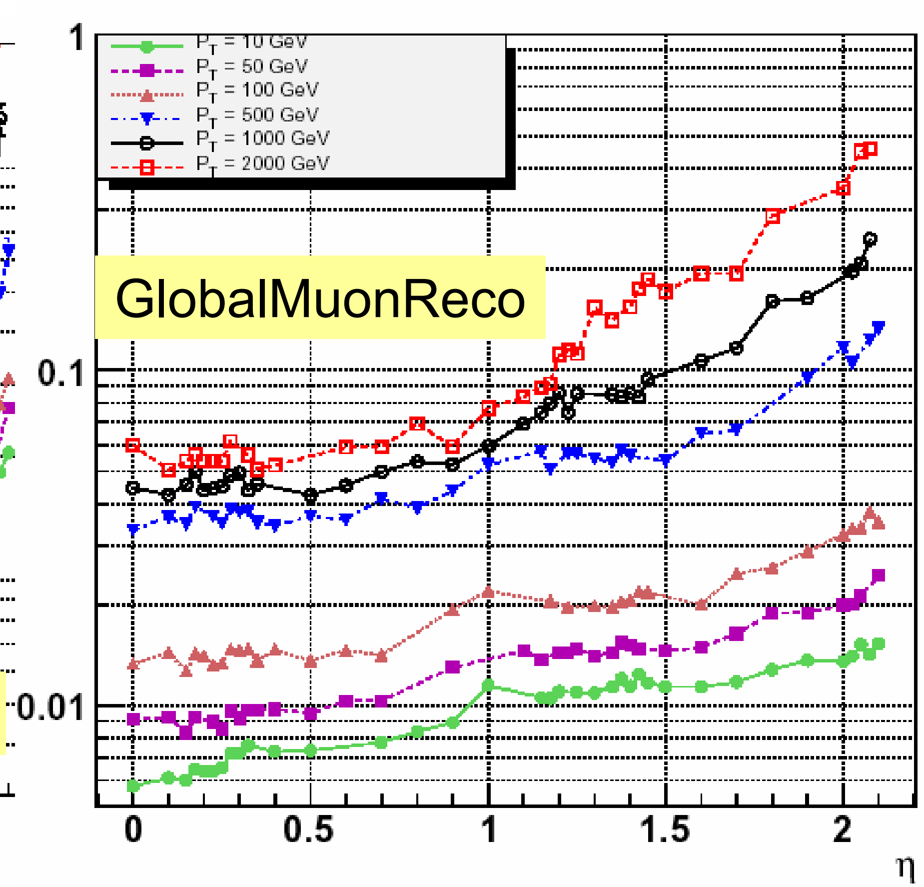




Muon reconstruction

- ORCA_8_2 muon 'off-line' reconstruction:
  - But: region eta= 2.1- 2.4 yet to be studied in detail



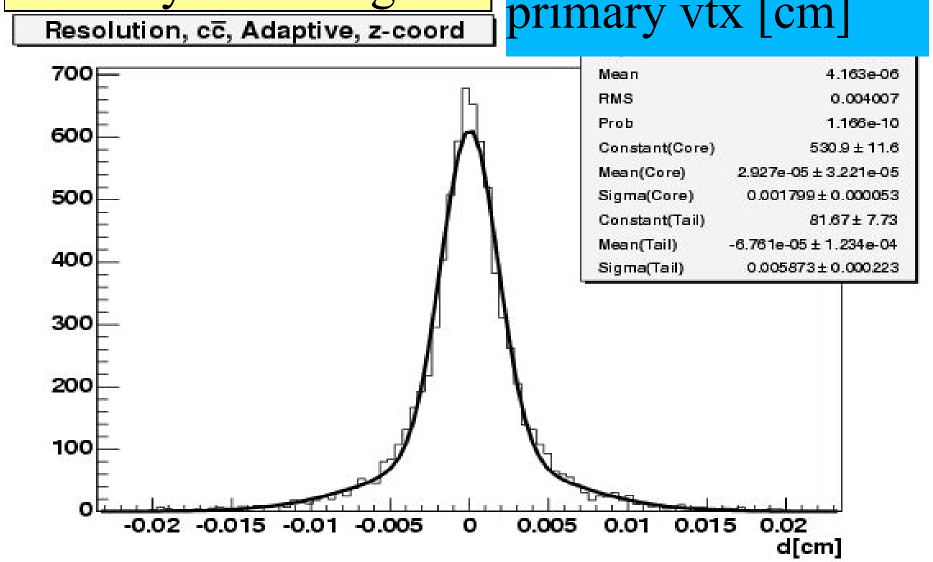


Vertex Reconstruction

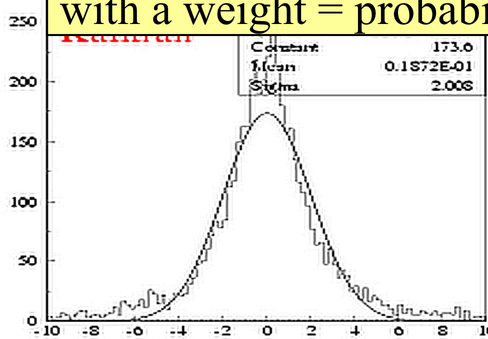
- Multiple algorithms implemented:
  - Least-squares method – KalmanVertexFitter (used in DAQ TDR)
  - Robust vertex fitting (AdaptiveVertexFitter, GaussianSumVertexFitter

AdaptiveVertexFitter
Iterative reweighting, χ^2-based
Global minimum found by annealing

c-cbar, Et = 100 GeV
$|\eta| < 1.4$
ORCA_7_6_1
Z-residuals of primary vtx [cm]

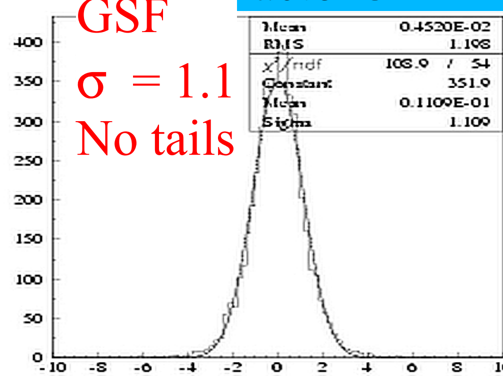


GaussianSumVertexFitter
N tracks, parameters modelled by Gaussian mixture with M components
Each component contributes to vertex with a weight = probability of compatibility



Pulls Y-coordinate
Vertex gun event
- 3 perfect tracks
- 1 track with cov.matrix wrong by factor 3

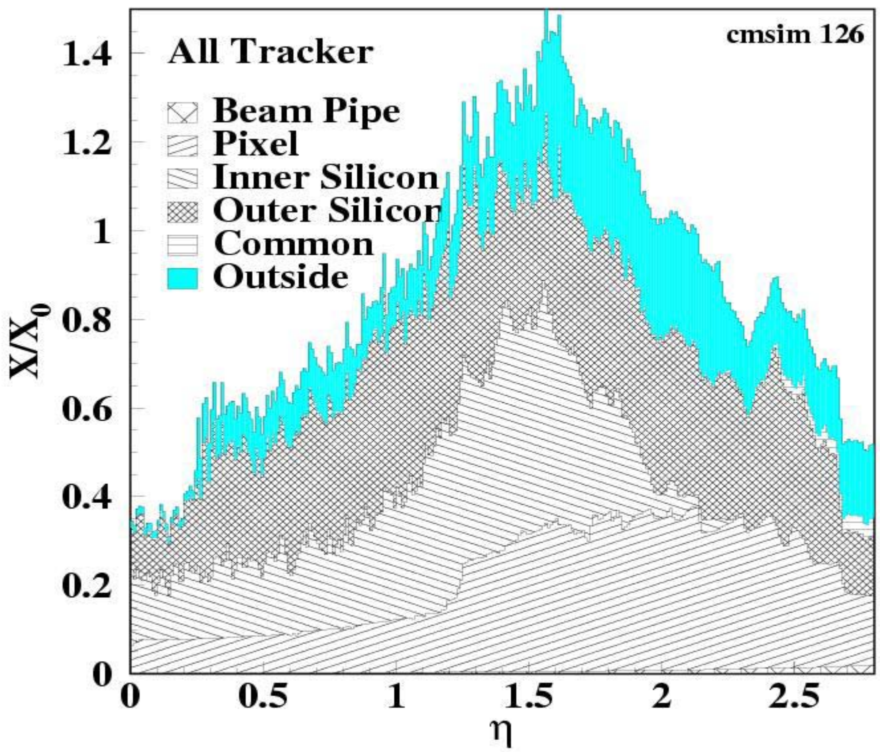
GSF
$\sigma = 1.1$
No tails



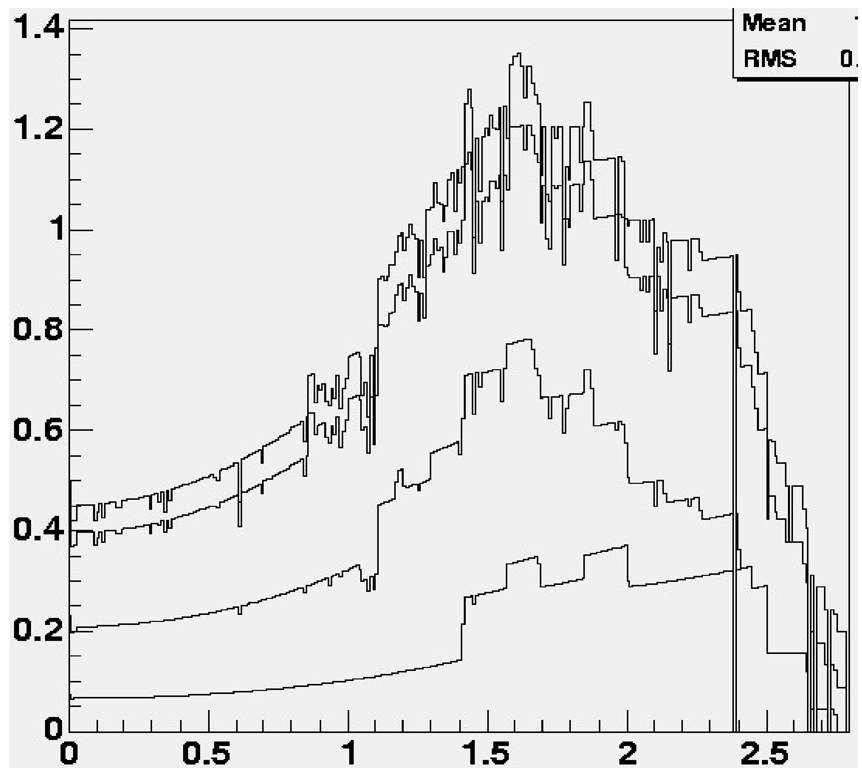
Fast simulation and reconstruction

- Fast simulation and reconstruction tool for very large simulated samples
- Fully outside Geant
  - Compatible with ORCA (same interface, persistent objects at DST level)
- Material effect description in the Tracker
- Fast rec track simulation with parametrized efficiency and smearing
- Fast em shower simulation (GFLASH)
- Parametrized responses to hadrons and jets
- Parametrized efficiency and smearing for muons

Material effect in FAMOS

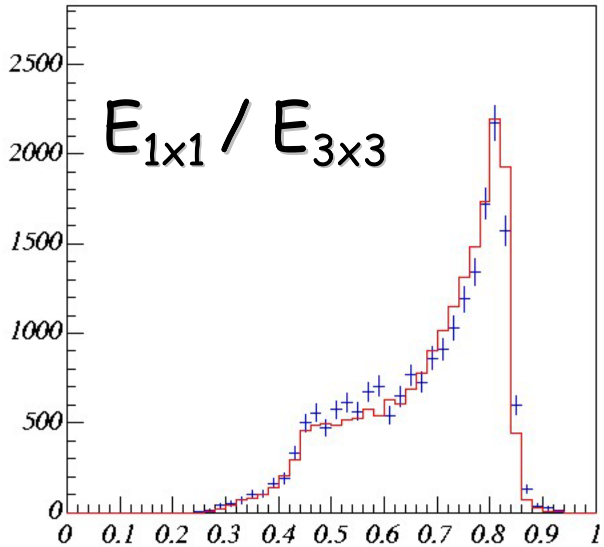


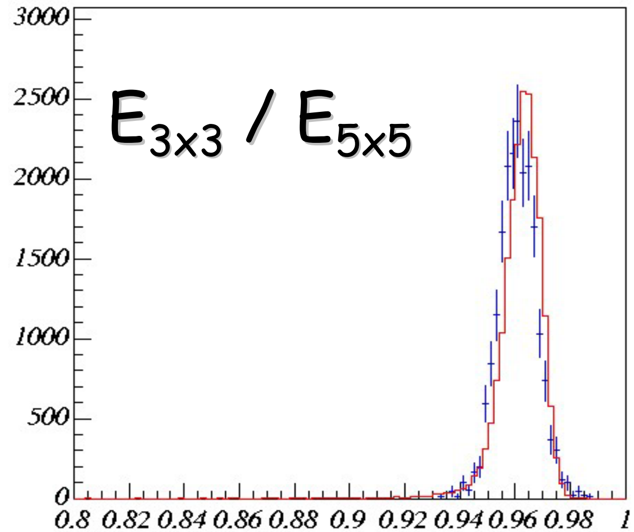
(CMSIM)



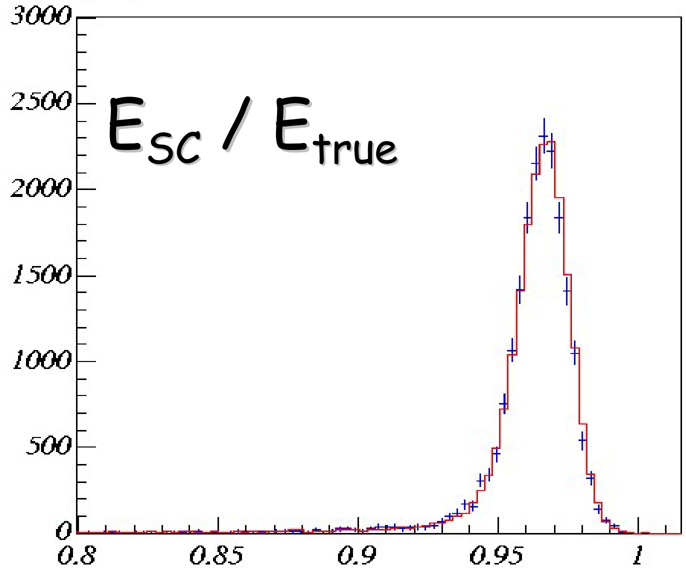
(FAMOS)

Checking/benchmarking FAMOS





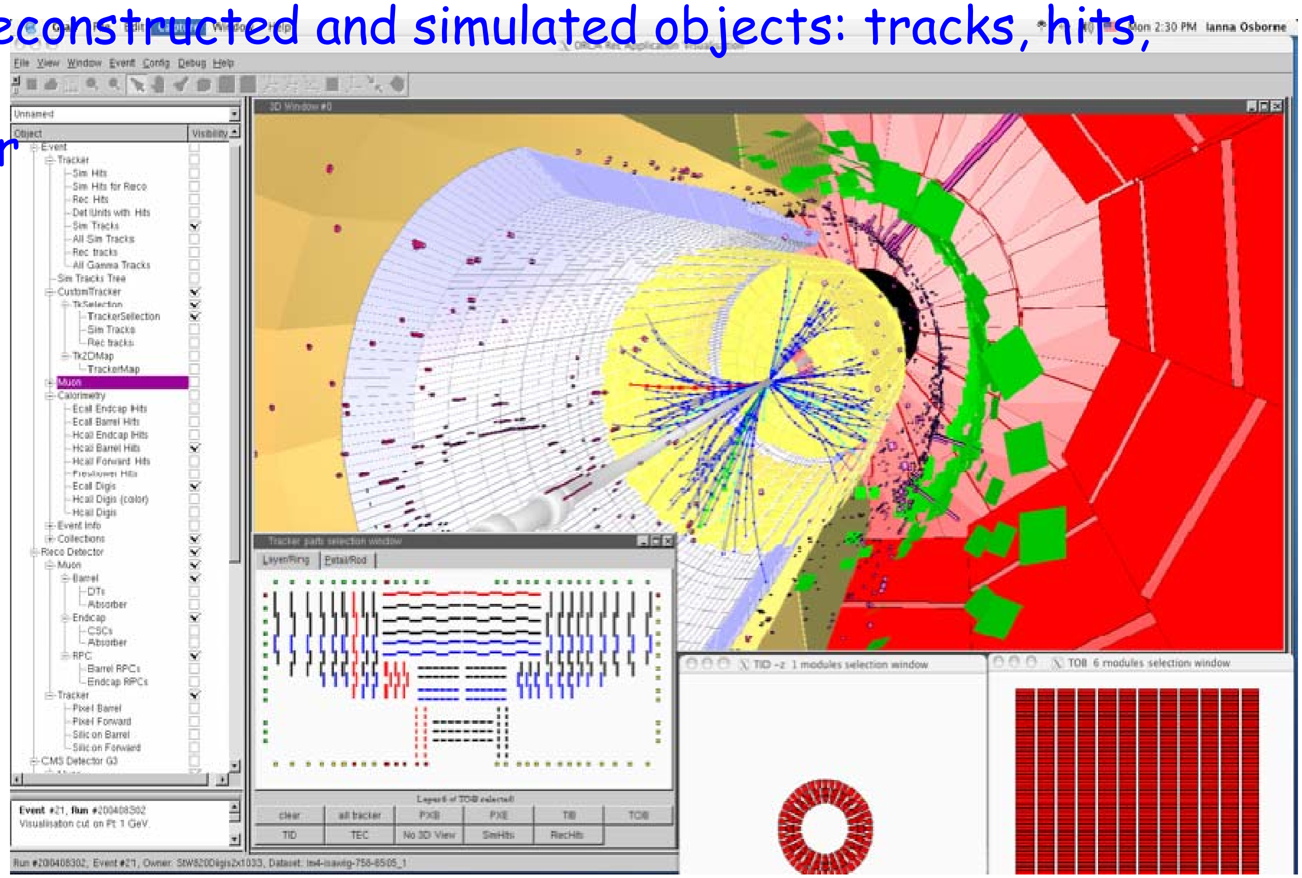
em shower containment



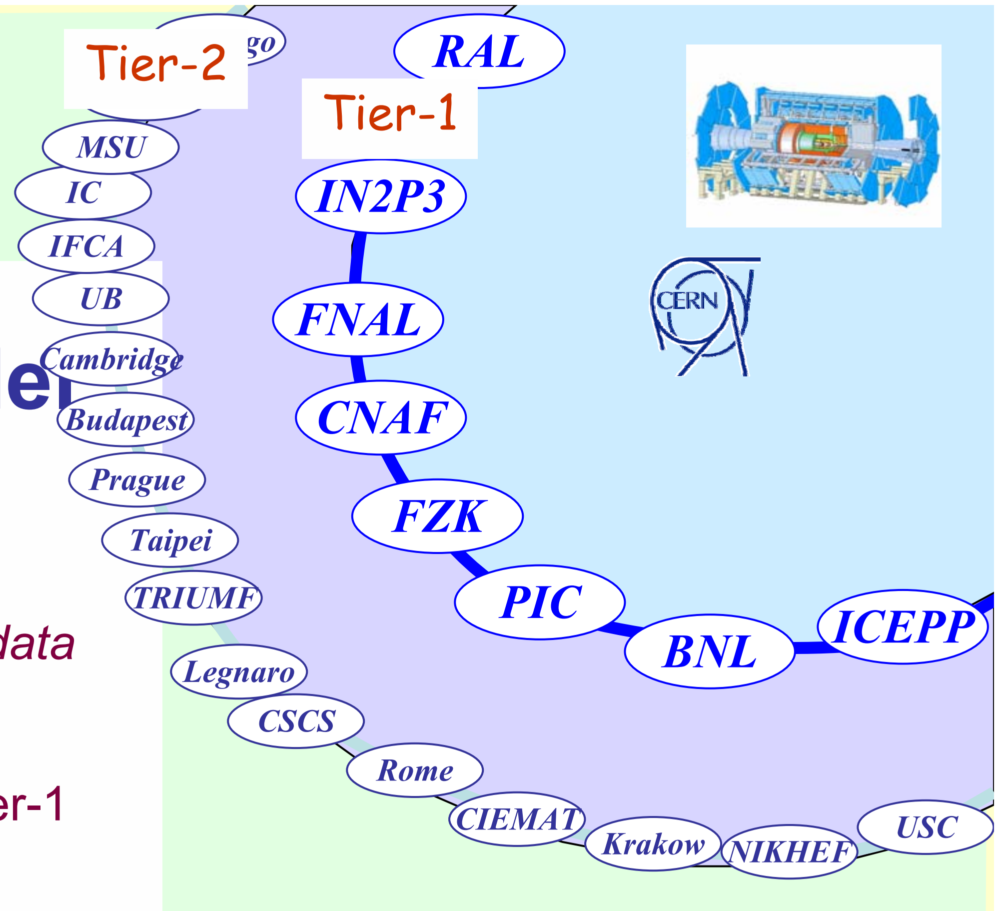
super cluster energy

IGUANACMS

- Visualisation applications for ORCA, OSCAR, test-beams (DAQ application);
- Visualisation of reconstructed and simulated objects: tracks, hits, digis, vertices, etc.;
- Full DDD detector visualisation;
- Magnetic field visualisation;
- Interactive modification of configurables at run-time;
- Custom tracker selection;
- Event browser;



LHC Computing Model

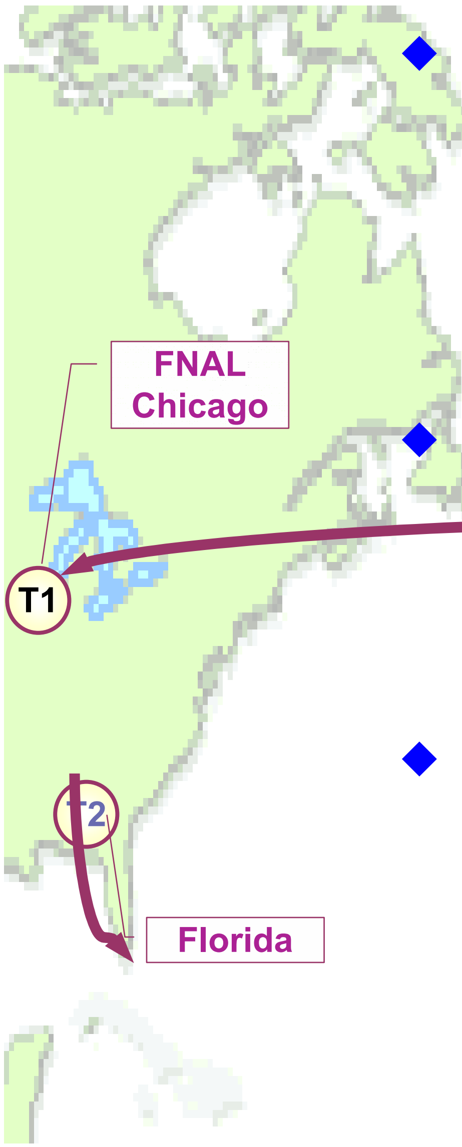


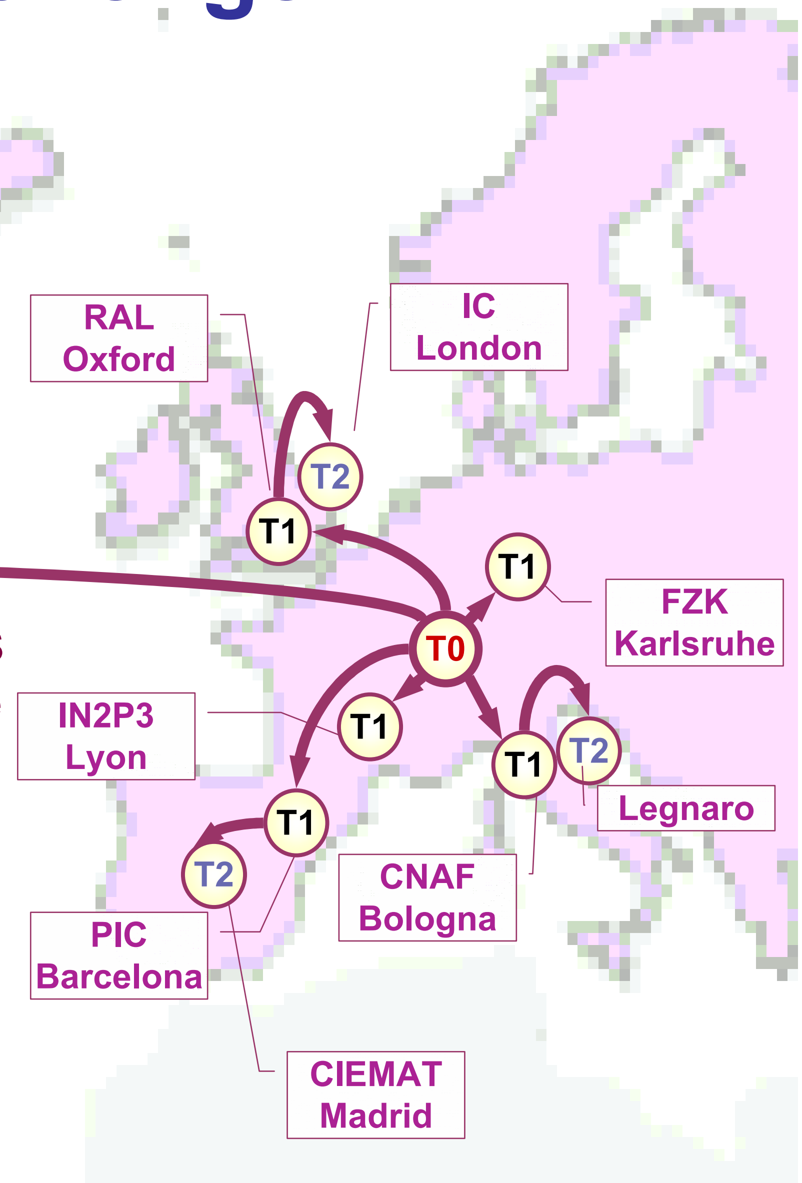
- Tier-0 – the accelerator centre
  - Filter → *raw data*
  - Reconstruction → *summary data* (ESD)
  - Record *raw data* and *ESD*
  - Distribute *raw* and *ESD* to Tier-1
- Tier-1 –
  - Permanent storage and **management** of *raw*, *ESD*, calibration data, meta-data, analysis data and databases → **grid-enabled data service**
  - Data-heavy analysis
  - Re-processing raw → ESD
  - National, regional support

- “inline” to data acquisition process
- -- high availability
- -- managed mass storage
- -- long-term commitment

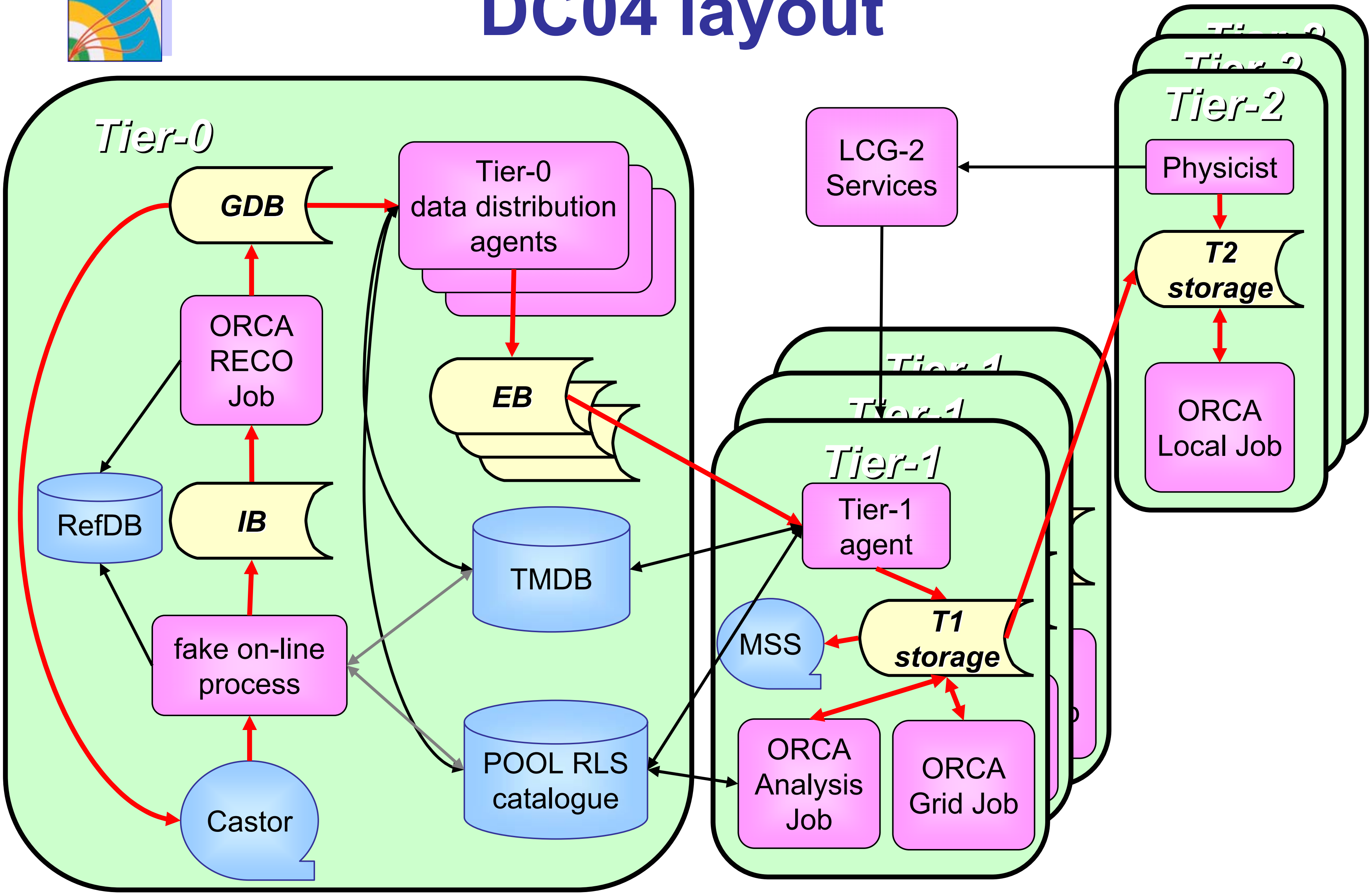
DC04 Data Challenge

- T0 at CERN in DC04
  - 25 Hz input event rate
  - Reconstruct quasi-realtime
  - Events filtered into streams
  - Record raw data and DST
  - Distribute raw data and DST to T1's
- T1 centres in DC04
  - Pull data from T0 to T1 and store
  - Make data available to PRS
  - Demonstrate quasi-realtime "fake" analysis of DST's
- T2 centres in DC04
  - Pre-challenge production at > 30 sites
  - Modest tests of DST analysis



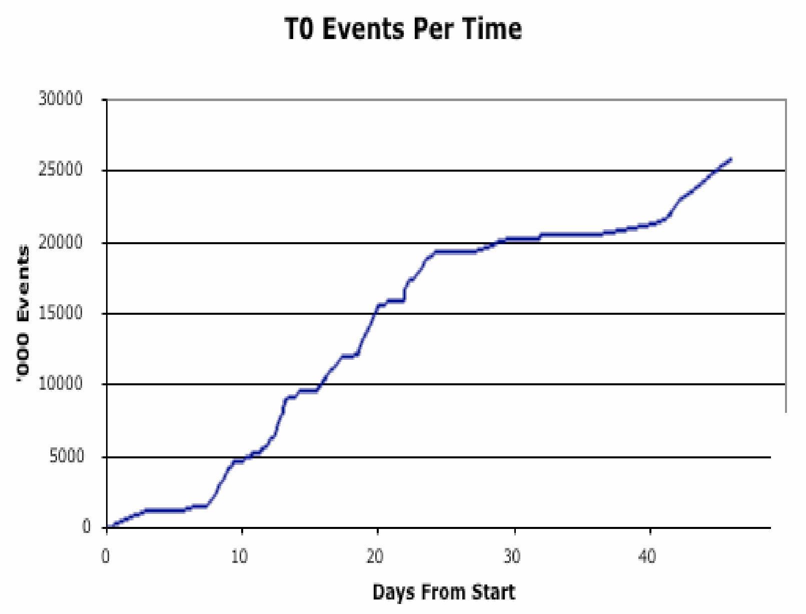


DC04 layout

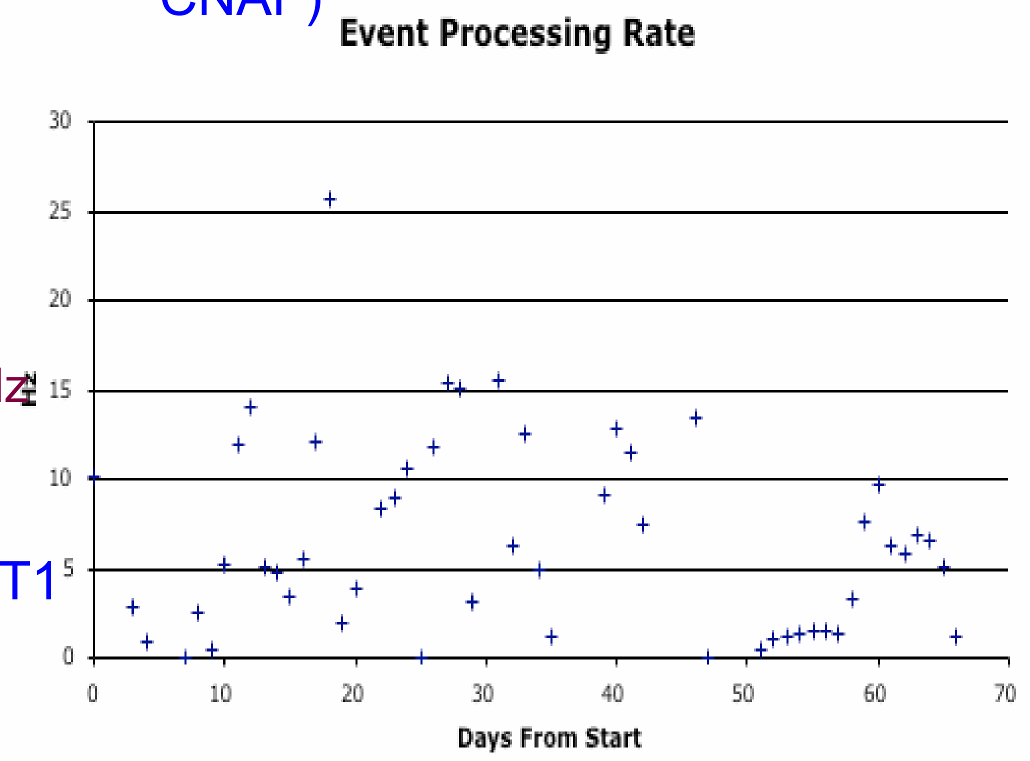


Tier-0
GDB
Tier-0 data distribution agents
ORCA RECO Job
EB
RefDB
IB
TMDB
fake on-line process
POOL RLS catalogue
Castor
LCG-2 Services
Tier-1
Tier-1 agent
MSS
T1 storage
ORCA Analysis Job
ORCA Grid Job
Tier-2
Physicist
T2 storage
ORCA Local Job

DC04 Processing Rate

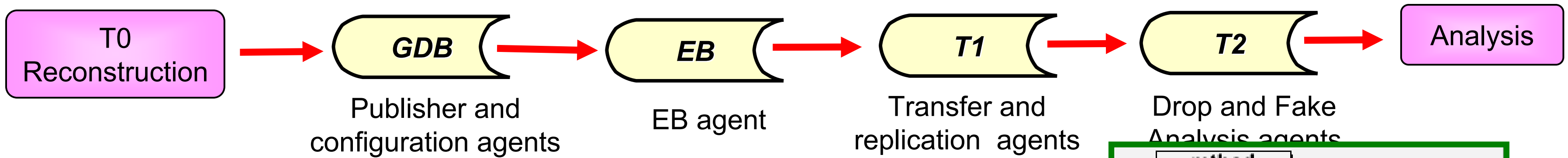


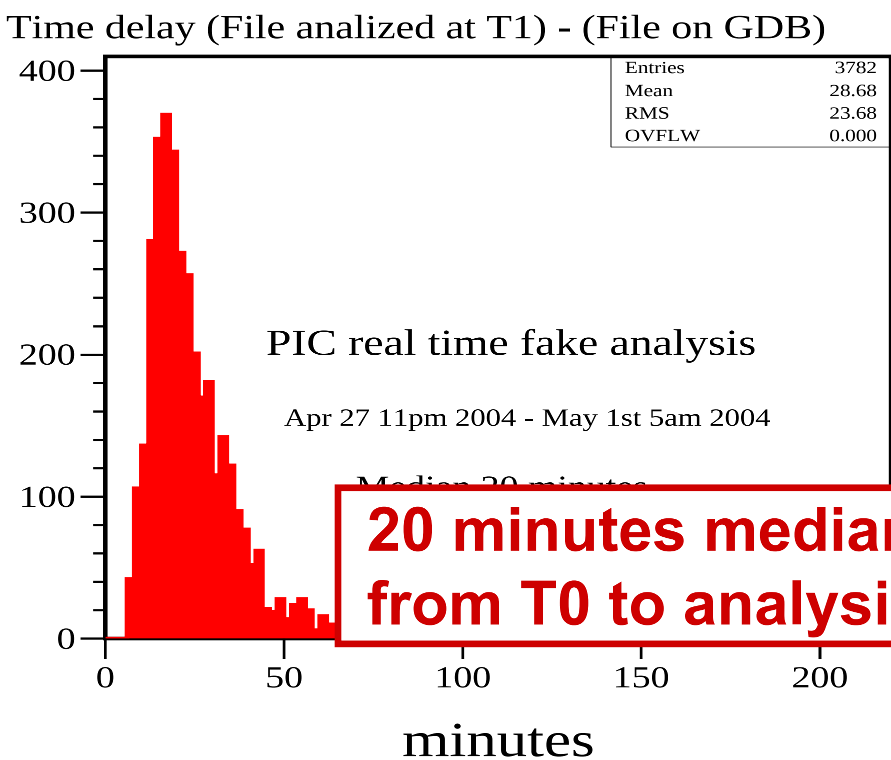
- Processed about 30M events
  - First version of DST code was not fully useful for Physicists
  - Post-DC04 3rd version ready for production in next weeks
- Generally kept up with SRM based transfers. (FNAL, PIC, CNAF)



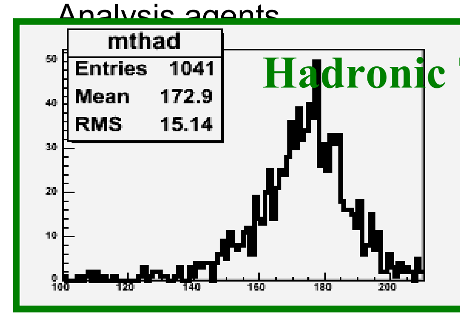
- Got above 25Hz on many short occasions
  - But only one full day above 25Hz with full system
- RLS, Castor, overloaded control systems, T1 Storage Elements, T1 MSS, …

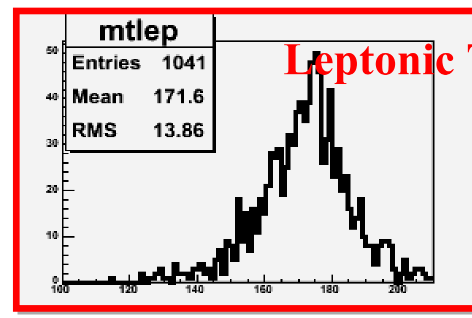
From Tier-0 Reconstruction to Analysis at Tier-1

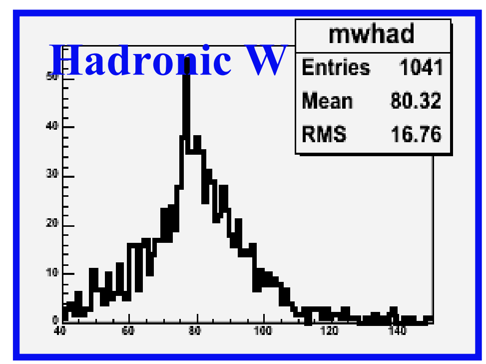




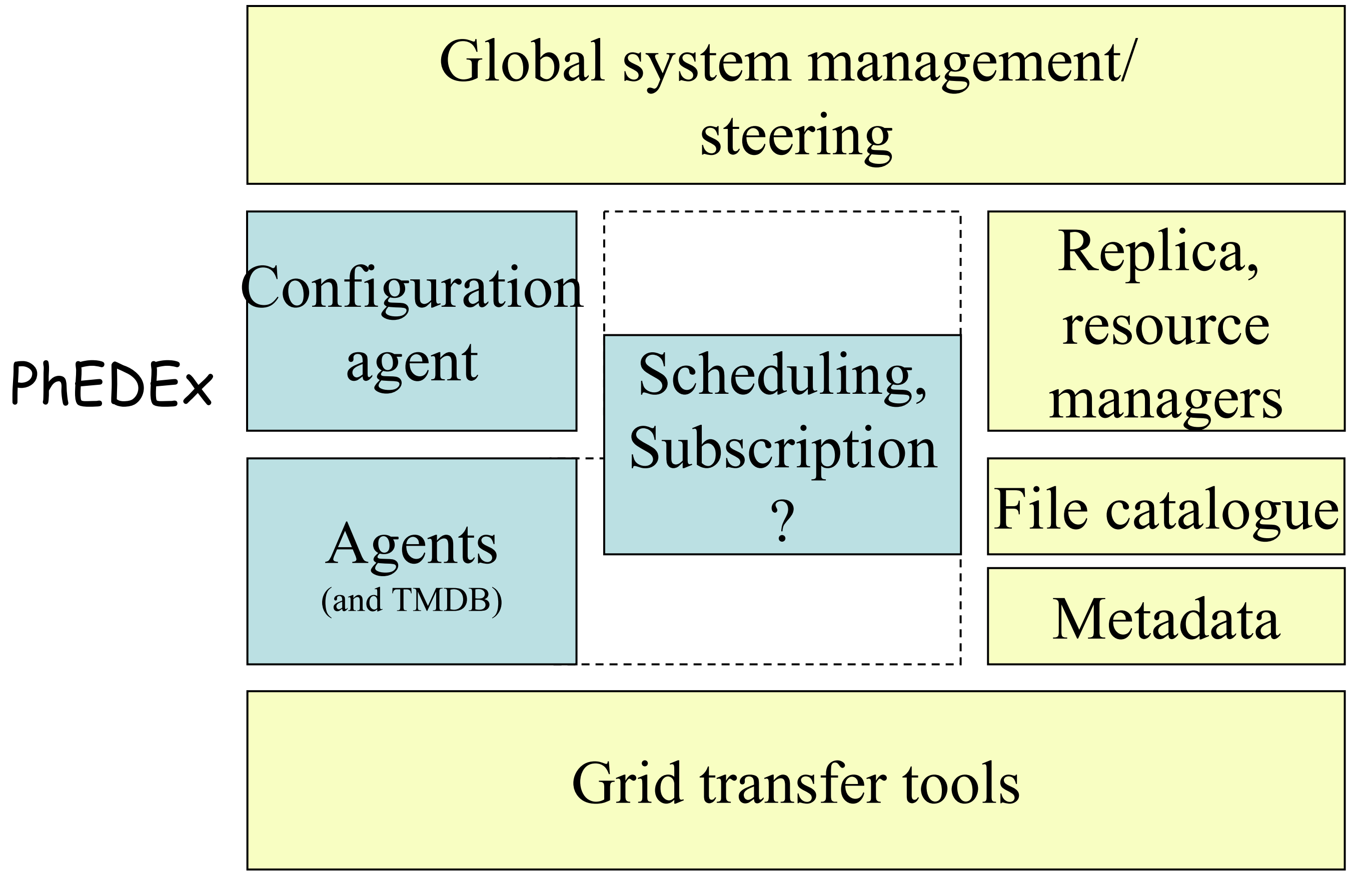
20 minutes median delay from T0 to analysis at T1







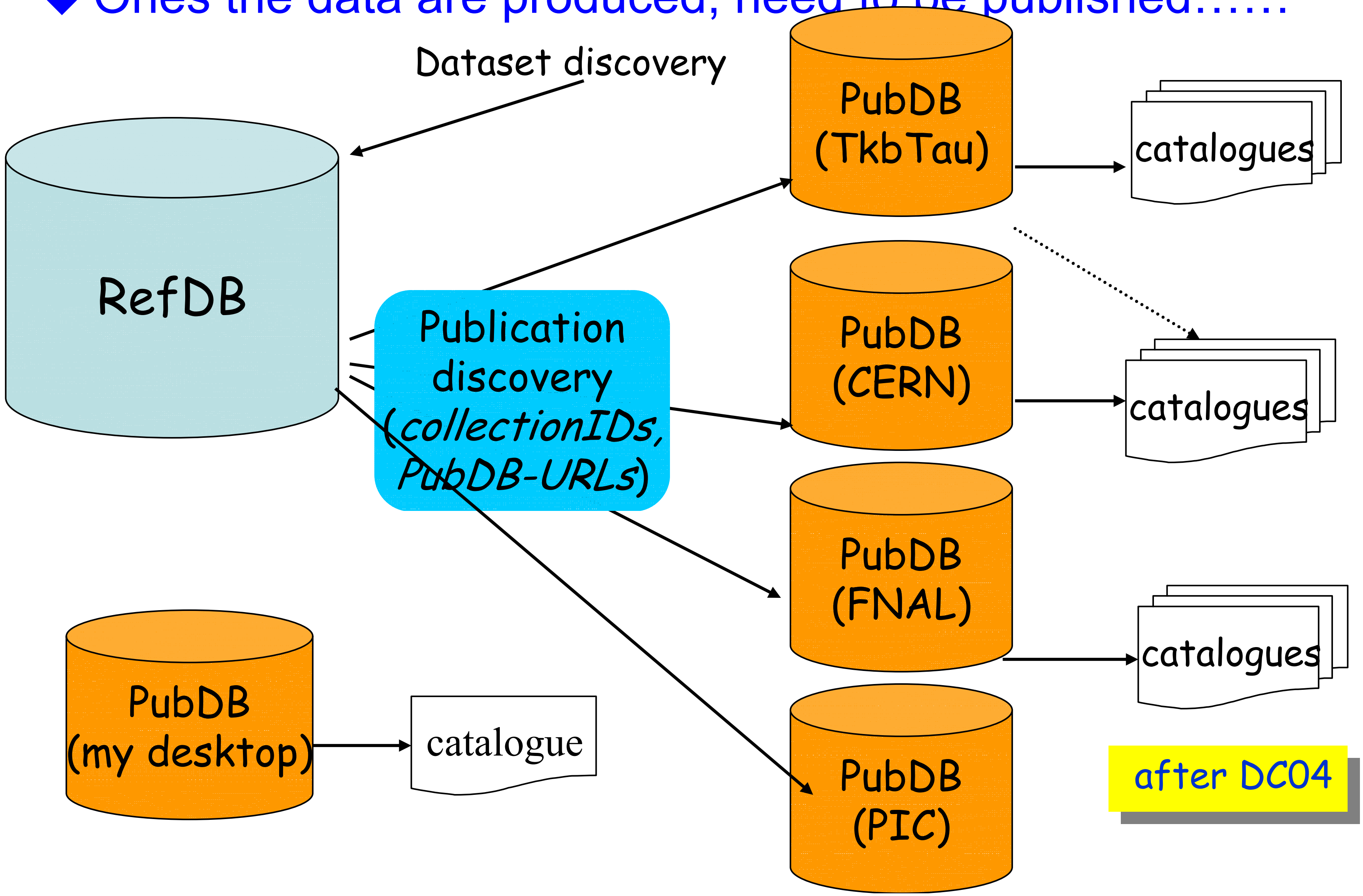
Data Movement - PhEDEx



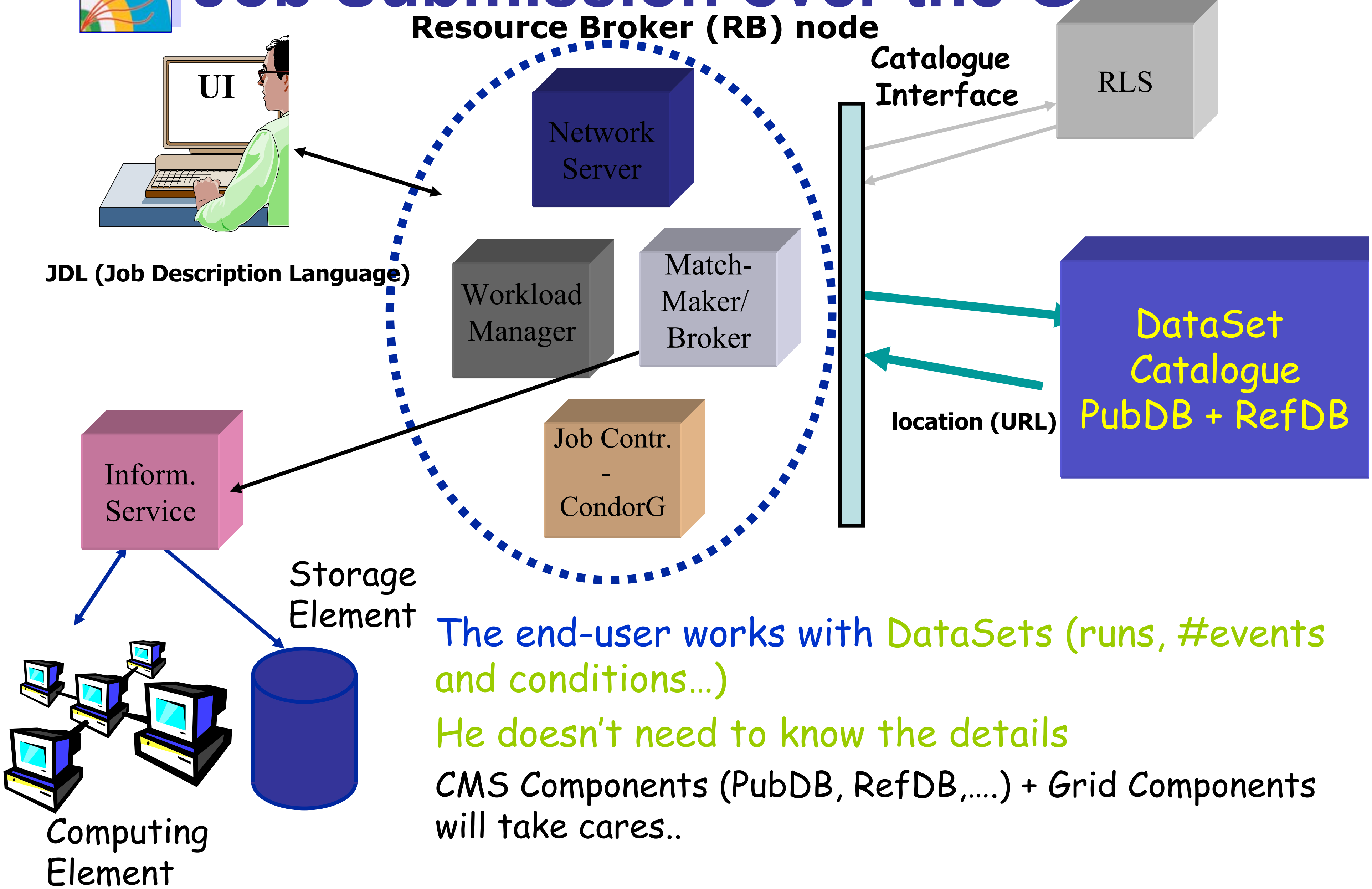
Includes a Management layer:
able to handle allocation based on subscription (polices)

Data Access - PubDB

- Ones the data are produced, need to be published……



Job Submission over the Grid

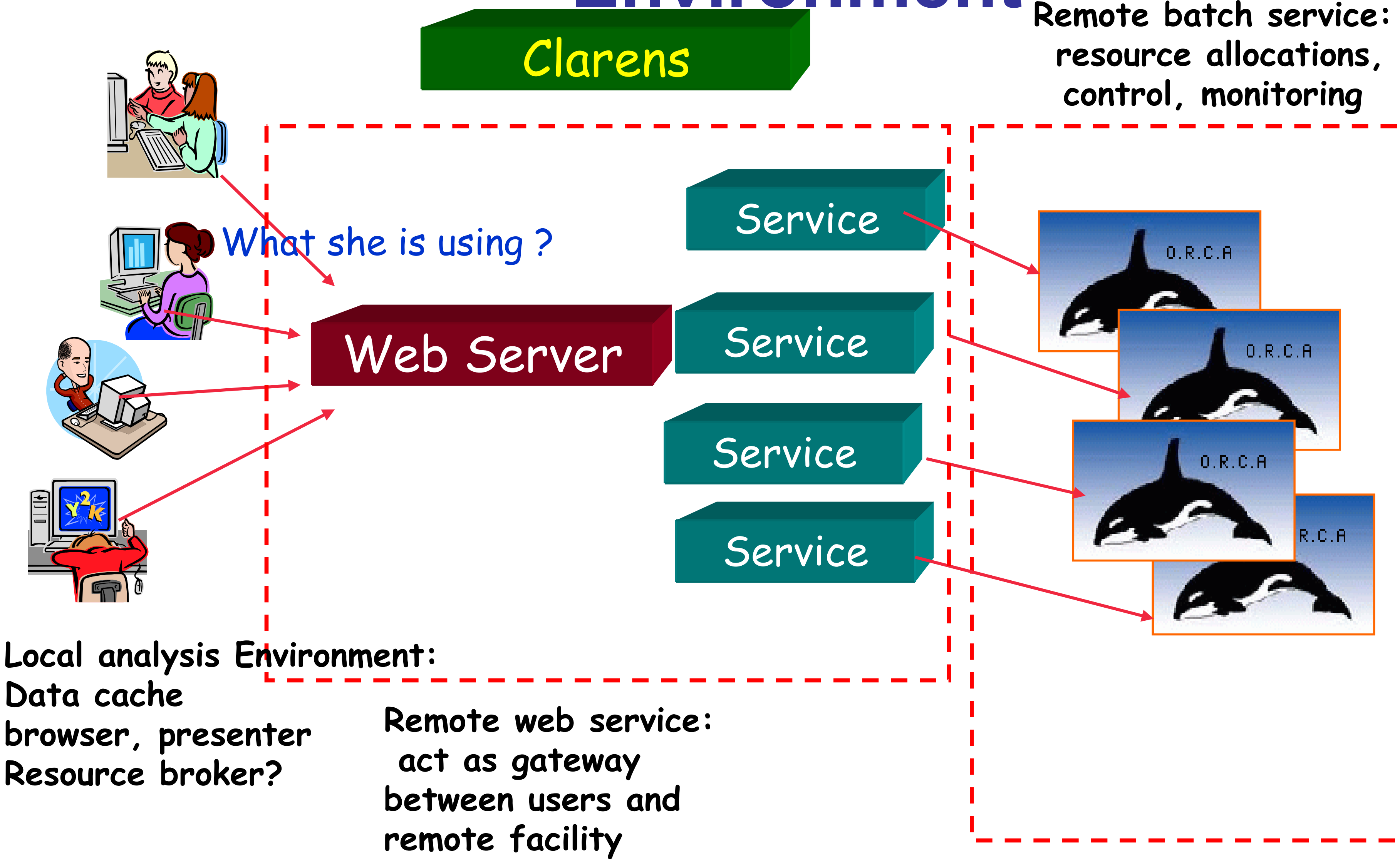


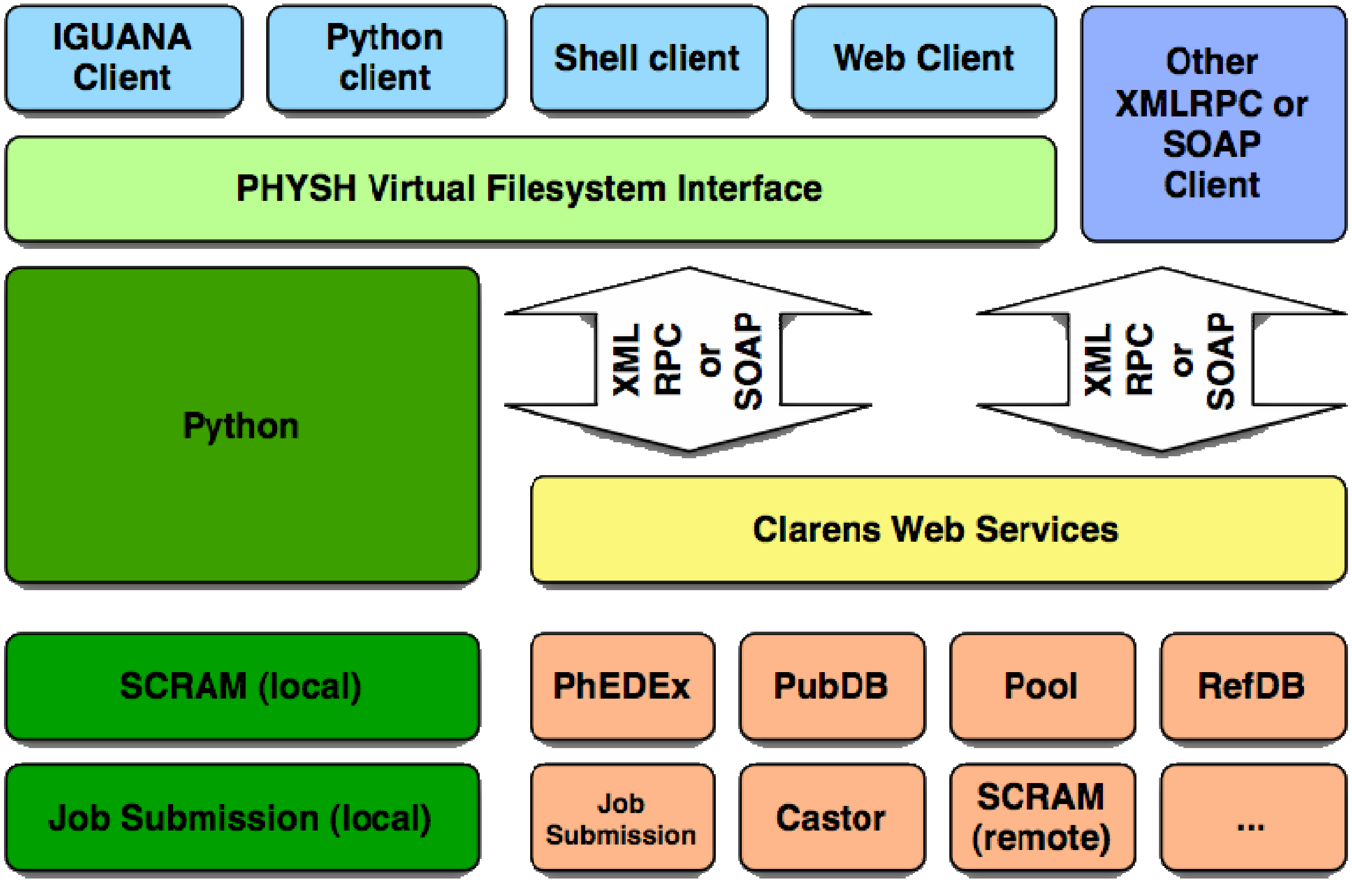
The end-user works with DataSets (runs, #events and conditions…)

He doesn't need to know the details

CMS Components (PubDB, RefDB,….) + Grid Components will take cares..

Analysis on a distributed Environment





PhySh: "end user" analysis environment.

A "glue" interface among different services already present (or to be coded).

The user's interface is modeled as a virtual file system.