Overview of Computing and Data processing for "Event-Experiments"

Giovanni Lamanna

LAPP - Laboratoire d'Annecy-le-Vieux de Physique des Particules,

Université de Savoie, CNRS/IN2P3, Annecy-le-Vieux, France

ASPERA-Workshop,
Barcelona 30-31 May 2011

Outline

- Introduction about data/computing models …
- Experiments: Auger, Antares, HESS, Fermi, CTA, (XMM)
- Summary and conclusions

Many thanks to my "info-providers":

R. Dubois, T. Johnson (Fermi)
M.De Naurois, P. Espigat (HESS)
J.Brunner (ANTARES)
JN. Albert (Auger)
R. Lemrani (CCIN2P3)
A.Ibarra (XMM)
C. Arvistet (ESA)
N.Komin, J.Ponz (CTA)

Computing and Data Models: an introduction

Data management main sub-systems

- On-site system (close to antennas)
- Data transfer and archive system
- Data processing (close to data and to scientists)
- Data access

Computing model main specifications

- Data stream and scientific products
- Experiments or Observatory (# of users/observers)
- Analysis-system (“Client/server”, “Mainframe”)
- “Time-line” and/or “formal” constraints
  (e.g. ESO/NASA/ESA vs ASPERA)

ICT applied to four main sub-components

- HW: Computing Elements (CE), Storage Elements (SE)
- Service: Archive, Database and Meta-Database
- SW: Reconstruction/Simulation
- Middleware

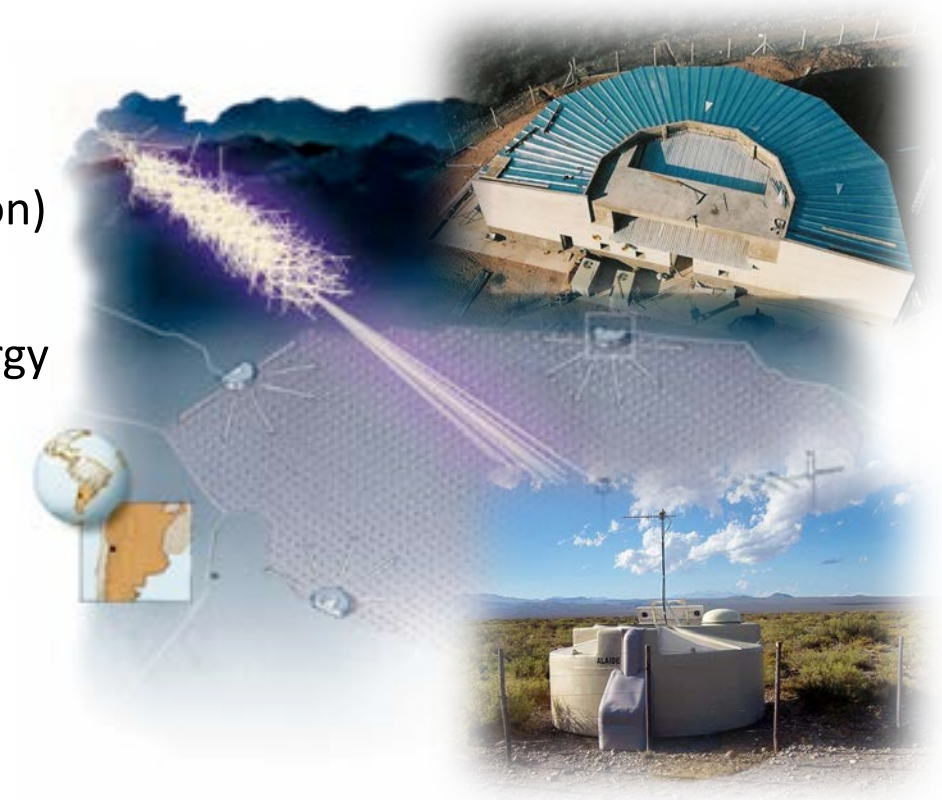
Final implementation into a Scientific Analysis System

SAS

The Auger Experiment

- The largest Cosmic Rays Observatory
  - 3 000 km^2 – in the Argentina Pampa
  - 1 600 Surface Detectors (Cherenkov)
  - 4 Fluorescence Telescopes
    - 6 x 4 Fluorescence Cameras
    - 10 % of the time (night – no moon)
  - Hybrid Events : SD + FD
    - Better determination of the energy
- High Energy Cosmic Ray
  - Energy 10^{20}
  - 1 / km^2 / century
  - Expected : 30 events/year
  - Observed UHECR : 50 (GZK effect)

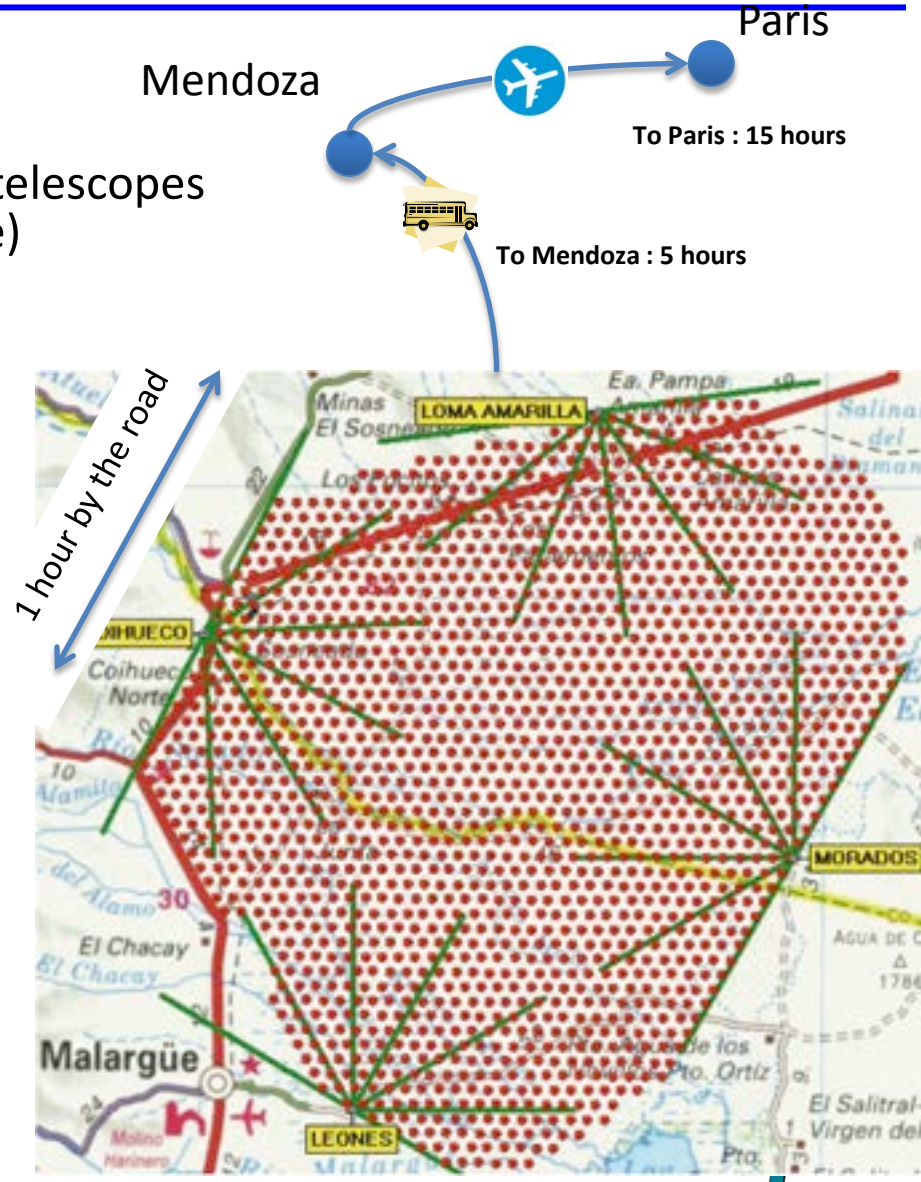
Fluorescence Telescope



Surface Detector

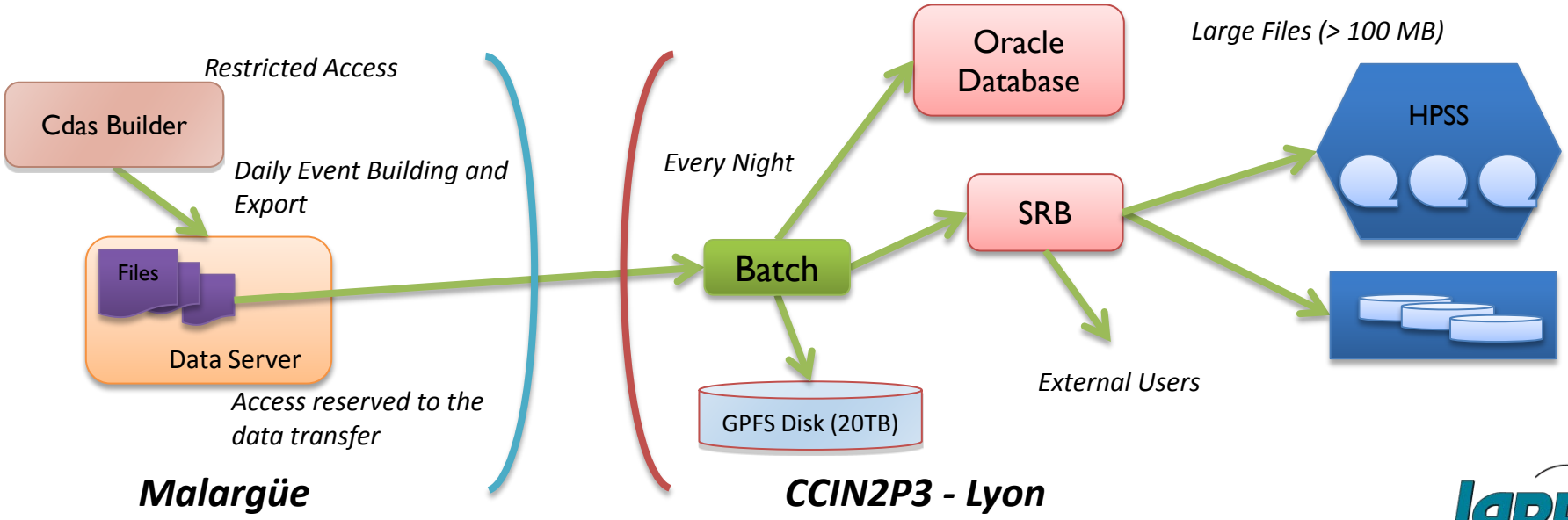
Data Taking

- Local Data Taking
  - Radio transfer from the thanks and telescopes to the Cdas main building (Malargüe)
    - Merged in Events
      - Root format
      - Immediate reconstruction of the event parameters
      - Copied to a data server for export
    - Daily transfer to Lyon
      - 1 – 3.5 GB/day (no calibration)
      - Poor Pampa BW ~ 50KB/s
      - Calibration data sent by disk every 2 months



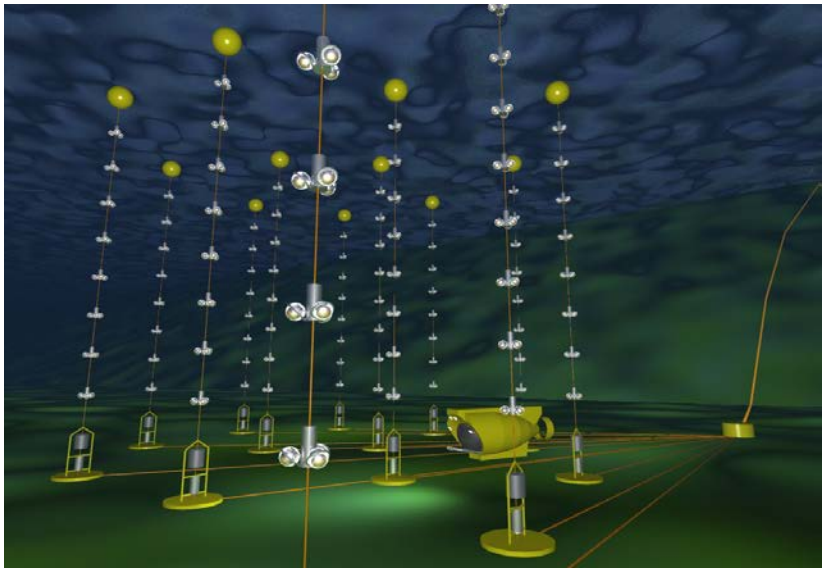
Data Transfer / Data process

- **IN2P3 Computer Center is the Main Repository for the Auger Data (Tier 0)**
  - Data files are copied every nights from the Malargüe server to CCIN2P3 by batches running at Lyon (on a GPFS large disk (20 TB), import manager status on Oracle DB)
  - The files are duplicated in SRB (Storage Resource Broker, a data Grid distributed files manager) for external access by the collaboration sites (Read-only access for the Auger user needing a CC account to get the authorization)
  - Data size: 8 TB since 2006 , 1.2 TB of reconstructed data.
    In average : Reconstructed data: 850 MB/day; Raw data + Calibration: 4.8 GB/day
  - MC production on GRID (showers + telescopes simulations)



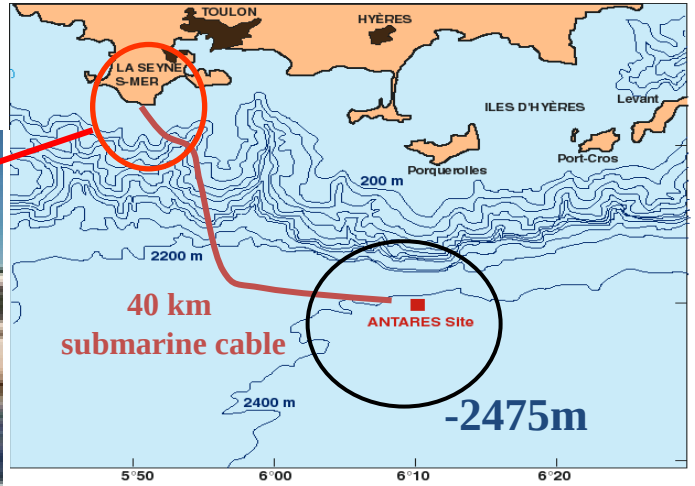
The submarine cosmic neutrino detector

- Cherenkov light from μ induced by ν interaction detected by 3D PMT array (2400 m depth)
- Time & position of hits allow the reconstruction of the μ (~ ν) trajectory



ANTARES produces ~ 10 TB raw data per year.

- Data are buffered at the shore station and transferred nightly to CC-Lyon.
- BW of the connection is 1 Gbit/sec.

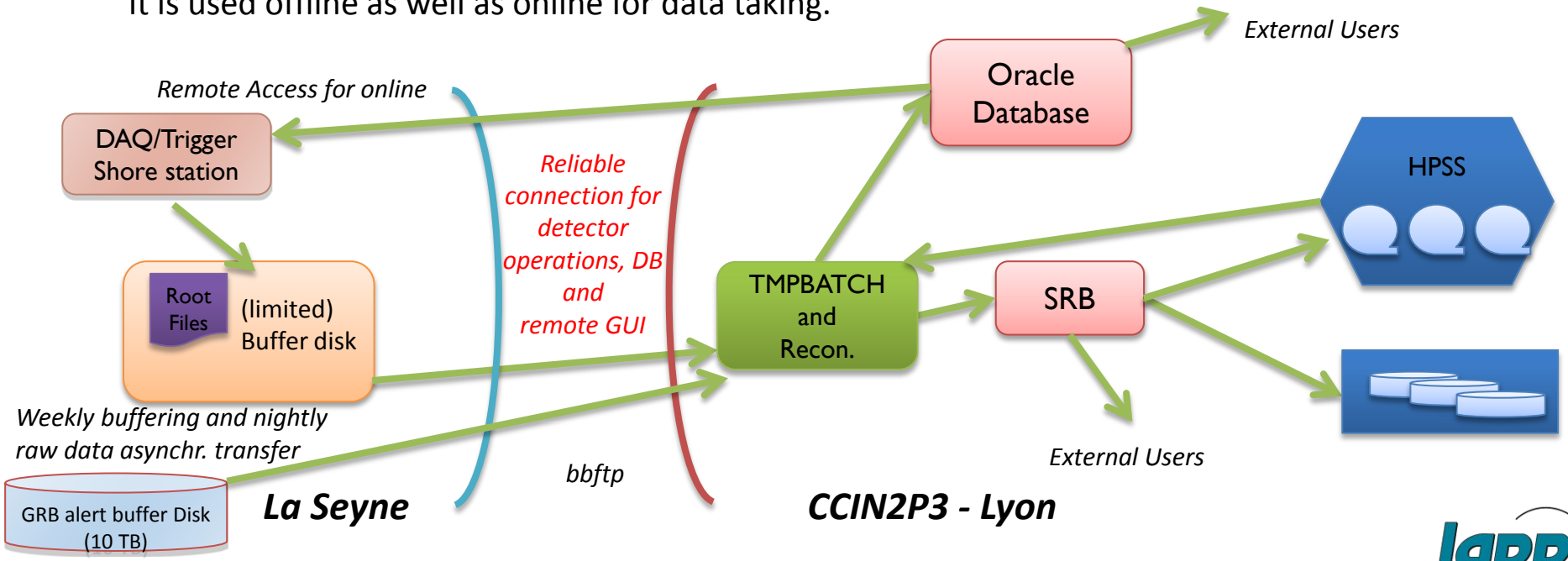


ANTARES shore station La Seyne/Mer

Data transfer & Data processing

- **IN2P3 Computer Center is the Main Repository for the ANTARES data (Tier 0)**

- The unfiltered data stream from the detector to the shore is ~1GB/sec. The trigger reduces this to 20 GB/day, i.e. a reduction of a factor ~4000.
- Data stream 50 GB/day (20 GB triggered + 30 GB untriggered data - external GRB alerts)
- Mass storage is done on HPSS.
- Main format is Root, read access to data for processing via XRootd.
- Access to external institutes is provided via SRB.
- Complex data (calibration, slow control, settings) are in an Oracle DB. Remote access is allowed. It is used offline as well as online for data taking.



The first 4-telescope IACT stereoscopic system

HESS is one of the leading observatories studying VHE gamma-ray astrophysics.

- Four 13m diameter telescopes in the Khomas highlands of Namibia (southern Africa)
- 100 GeV – 100 TeV, 15% energy resolution
- 5' angular resolution, 5° field of view (each tel. has 107 m^2 mirror area, 15 m focal)
- 960 photomultiplier tube pixels
- 5° FoV (1.4 m)
- 1 GHz sampling

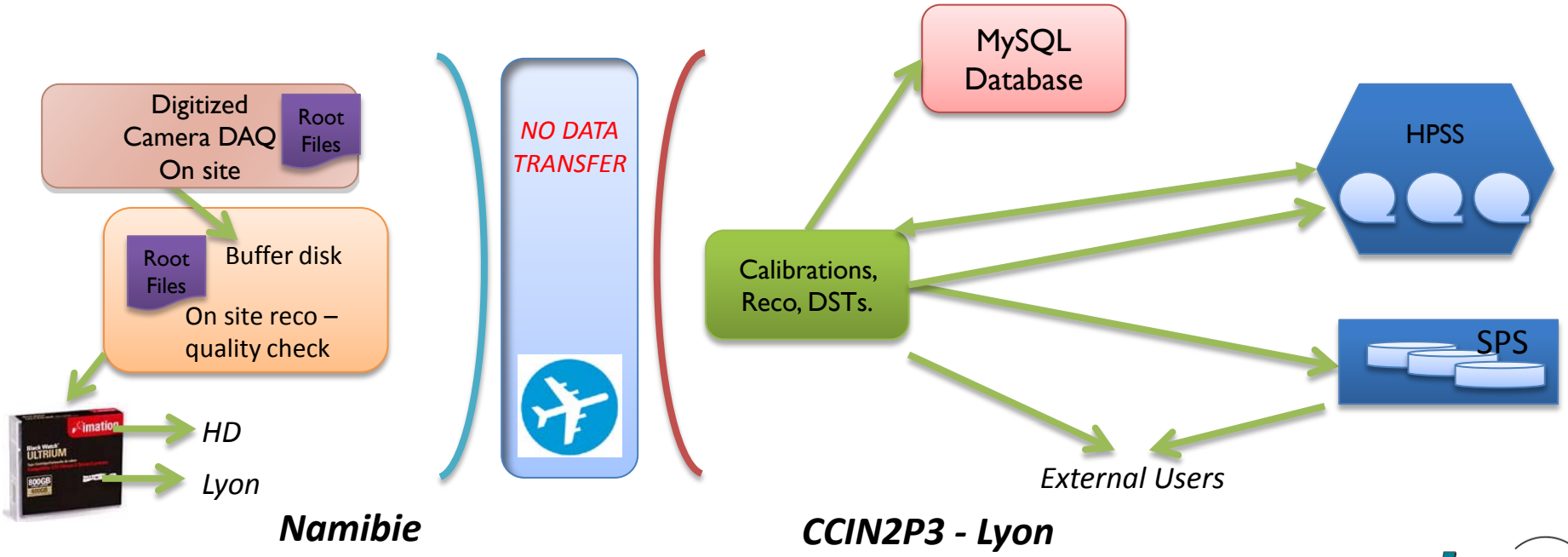
- Camera front-end digitization
- Total raw data stream ~ 8GB/hour
- Trigger rate 200 Hz
- On site control room (20 PC + some TB SE)
- All files in ROOT format
- Data are buffered and saved on tapes

then fly to Europe . Not enough BW for transfer

Data transfer & Data processing

- **IN2P3 Computer Center is the Main Repository for the HESS-FR data (Tier 0)**

- 10 TB raw data per year at CCIN2P3 (Today: 291 077 GB + 14 TB scratch disk + ~ 200 TB MC)
- MySQL DB for book keeping, jobs handling, calibration, ..
- DST (10 times smaller than raw data) are Root files (data are Root files at any level), Pedestals & Broken pixels are Root files as well.
- CPU: 1 day machine per run (28 min observation time) for calibration,
  1 day machine per run for DST production (~ 6% CCIN2P3 CPU to be doubled with HESS2)
  ==> ~ 5000 day machine per year (Analysis is fast compared to calibration and reconstruction)
- FR-HESS users ~20-30 accessing CC for analysis



Data access/analysis

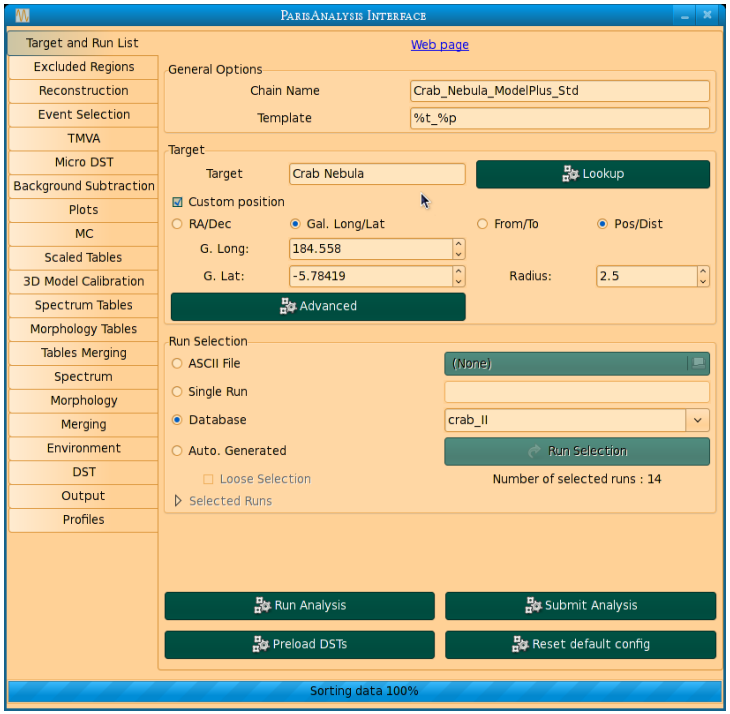
HESS-FR computing model for data production, access and analysis at CC-Lyon : Client /server paradigm:

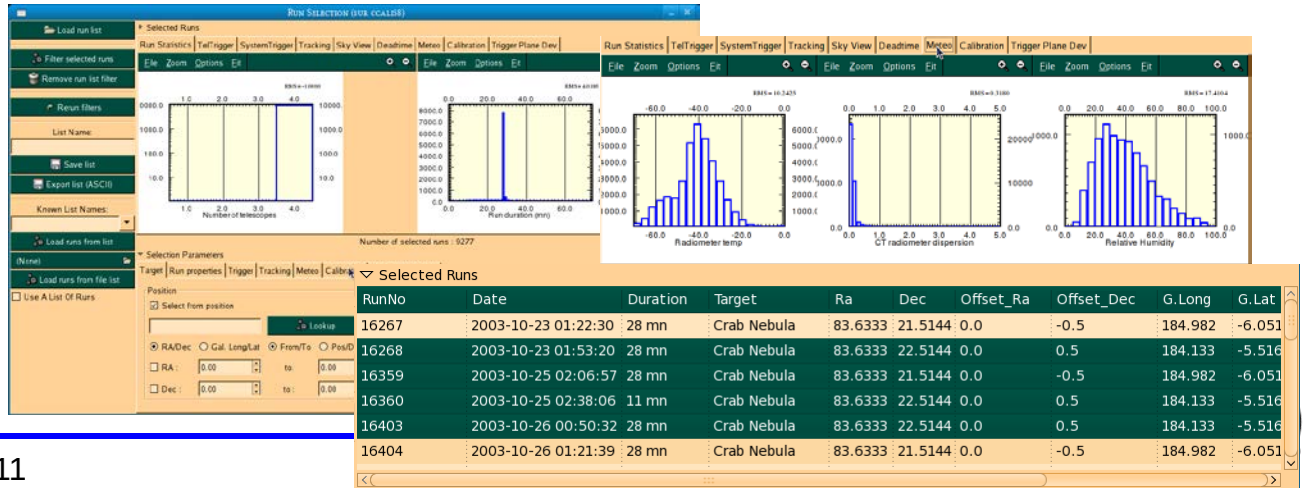
- Web services for remote data processing, calibrations control …
- Client for data selection (sky, source, run/time, pre-lists, other lambdas, pub. lists)
- Client interface (and VM) for DAQ sys.
- Client DB-DAQ slow-control parameters data selection (trigger, atmos. ,cameras status, telescope pointing…)
- GUI-Client service for job-analysis submission-> interfaced to GridEngine batch system.
- Analysis is not highly process demanding since DSTs have all reconstruction methods and param. .

  (… periodical re-proces.)
- Analysis in Root env.

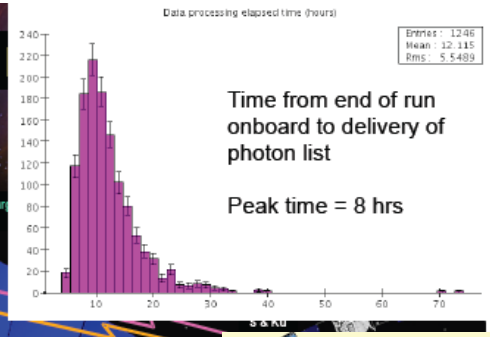
("ParisAnalysis" framework)

included output high-level files.





Fermi LAT offline computing



Fermi LAT

CTA Toulouse May 17, 2011

Principal Computing Resources

- **SLAC compute farm (Fermi allocation)**
  - 1600 CPUs in batch farm running LSF (+ peak loads of >3000)
  - 1100 TB disk - xroot – 32 TB raid 10 Sun thumpers
  - 1500 TB tape in robotic silo (HPSS)
- **Lyon compute farm (CCIN2P3, France)**
  - 1200 CPUs in batch farm running BQS
  - few TB disk allocated for all Fermi uses (transfer generated files to SLAC)
  - seamlessly used by pipeline from SLAC
- **Plans for using Italian GRID**
- **Extensive use of relational DB**
  - Two 'Niagara' redundant Oracle servers each with 1.5 TB raid disk
    - Science monitoring
    - LAT+Spacecraft 'housekeeping' trending
    - Bookkeeping (dataCatalog, processing config, etc.)
    - Science data (e.g., GRB catalog)
    - Resource trending

R.Dubois

6/29

Fermi LAT offline computing

Fermi LAT

CTA Toulouse May 17, 2011

Level 1 Processing Many-Ringed Circus

raw

digi

recon

xrootd used as starting and end points
(xrootd is a cluster filesystem)

Decompress \ Root

1 hr

DL

down link

1.5 hr

Fits

Root Fits

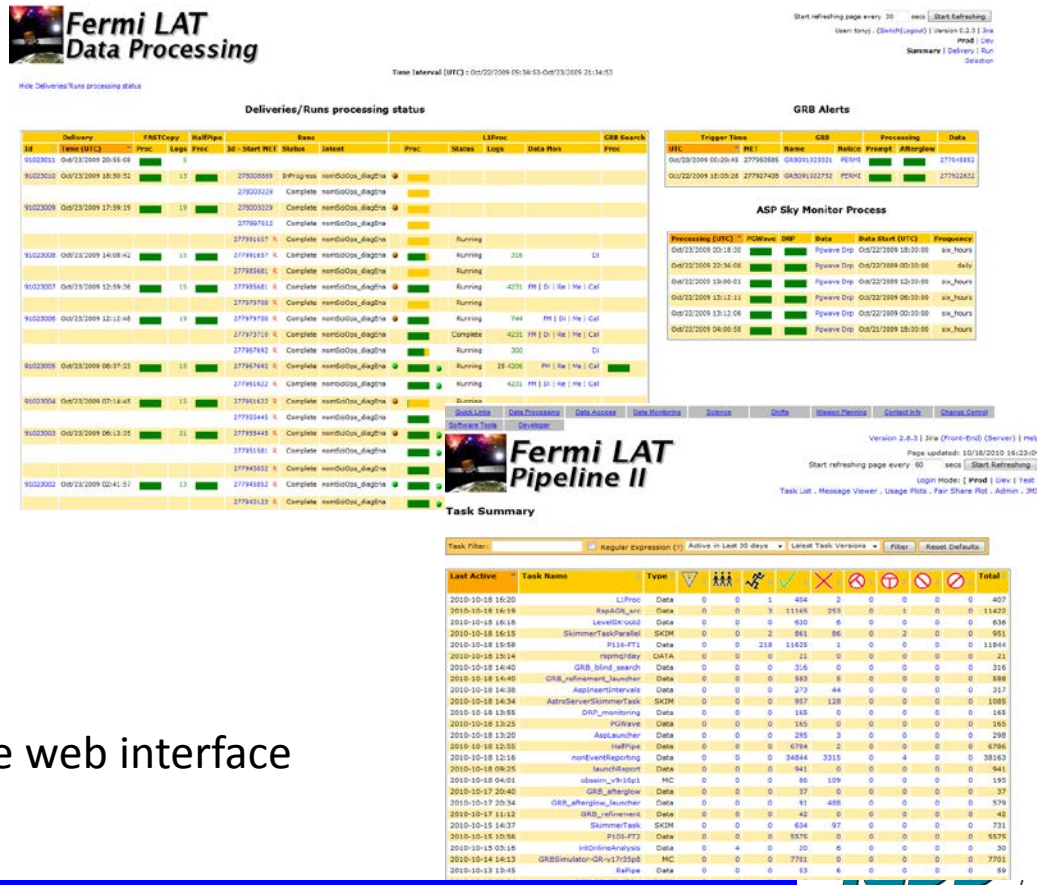
6 GB/day trending data into Oracle

25/29

Fermi LAT offline computing

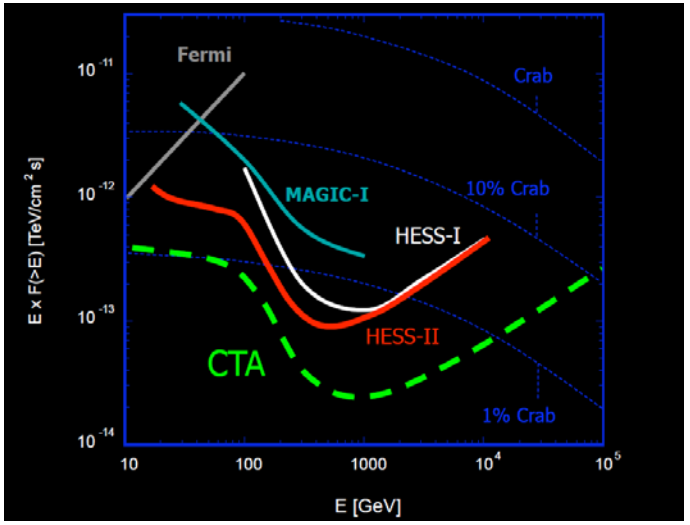
Fermi-approach:
"web-service" for quality check of data pipeline workflows (recon. jobs)

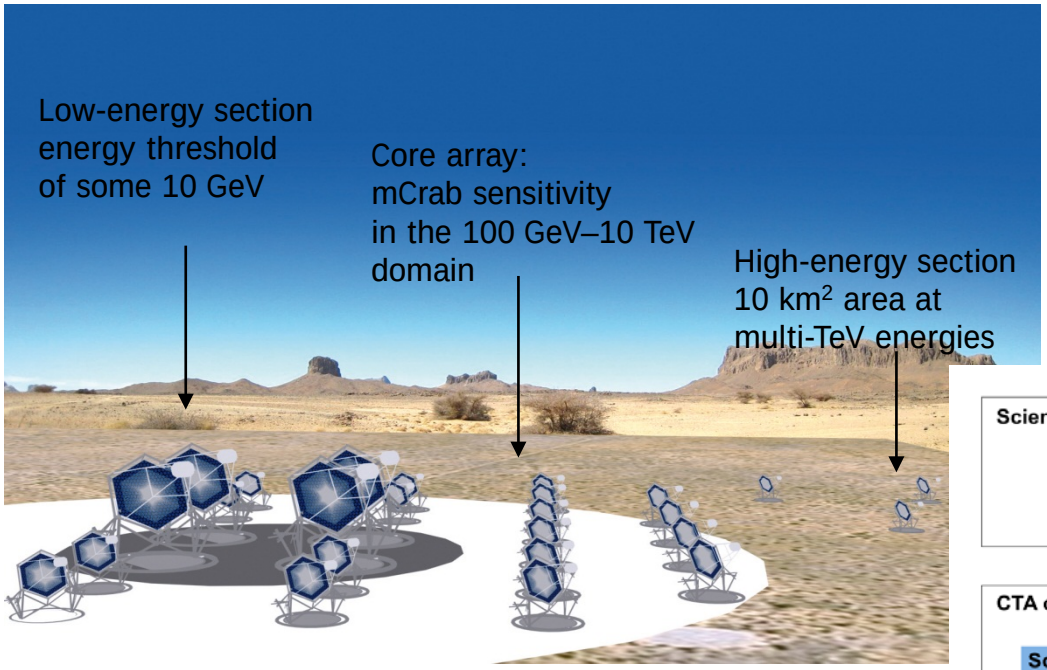
Web interfaces allows:
- Quick overview of data processing
- Flags runs requiring further attention
- Allows "drill-down" to isolate/identify problems



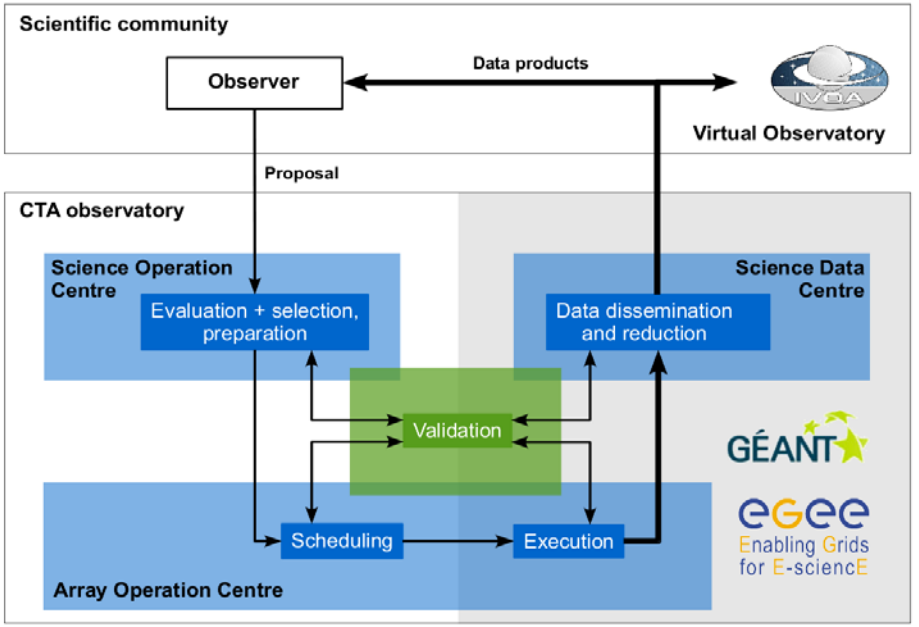
Pipeline web interface allows :
-Many views of data processing, down to log files of individual hobs
- Jobs submission
- Jobs failing, can be rolled back from the web interface

Cherenkov Telescope Array: the first gamma-ray Observatory for the next decade



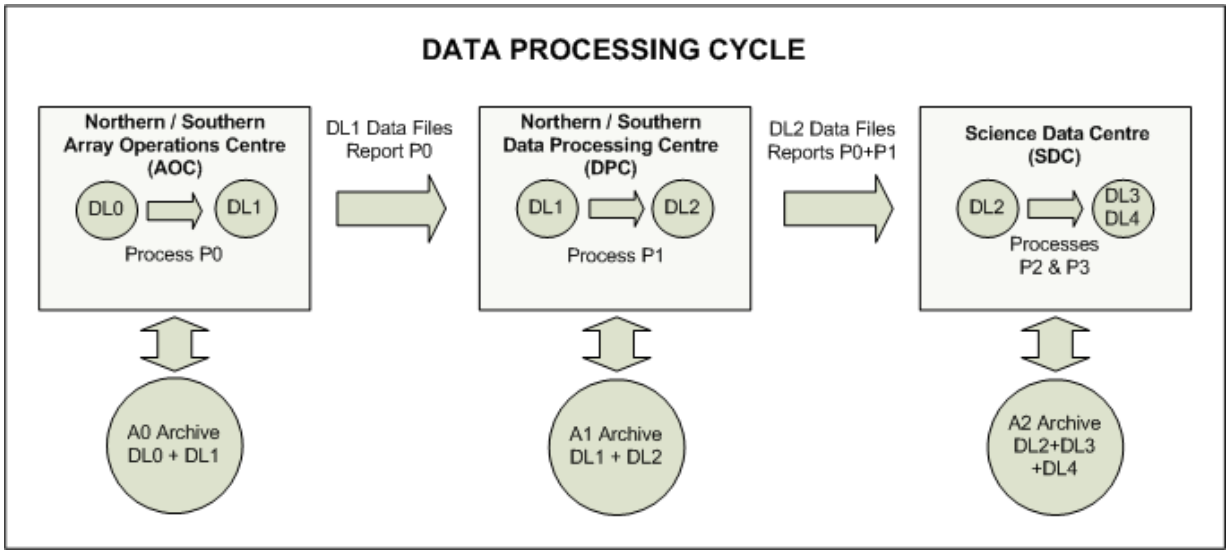


Cameras with 2000 pixels
Triggers at a rate of 10 kHz (array 3 kHz).
Main data stream of 1-10 GB/s.
Total data volume : some PBs per year.

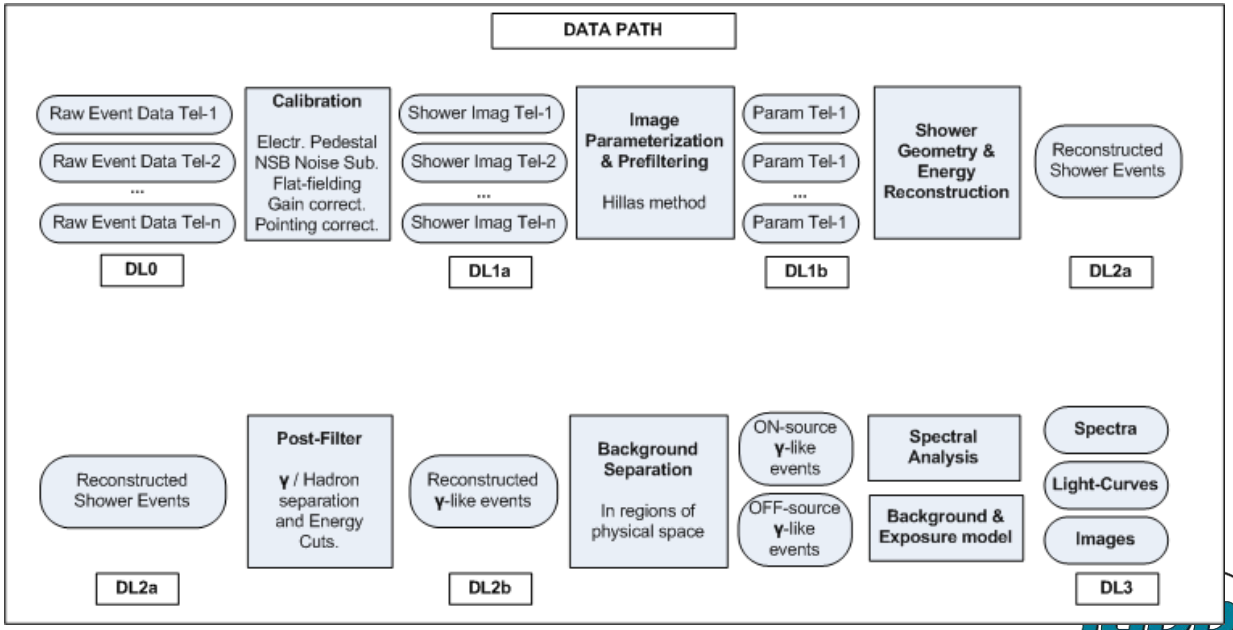


CTA model analysis

Data Processing Cycle: for the distributions of functions and products in Operations Centres



Data Path to be optimized: by examining the complete processing, with reference to the Data Levels



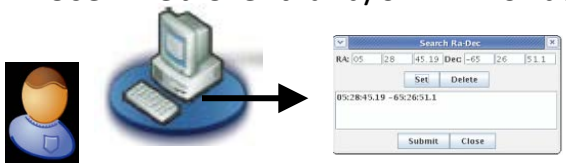
Preliminary (work in progress)

Preliminary
(work in progress)

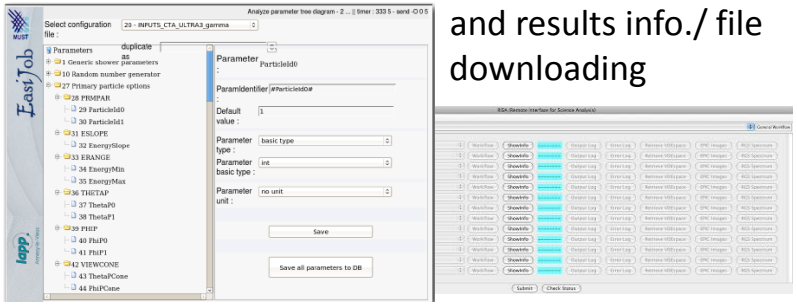
Then a model: e.g. Grid - SAS

Example: a possible Grid-SAS model based on the MC CTACG model and ..

User Web Client and/or VM for data searching



Client application for analysis job submission and results info./ file downloading



High-level data request

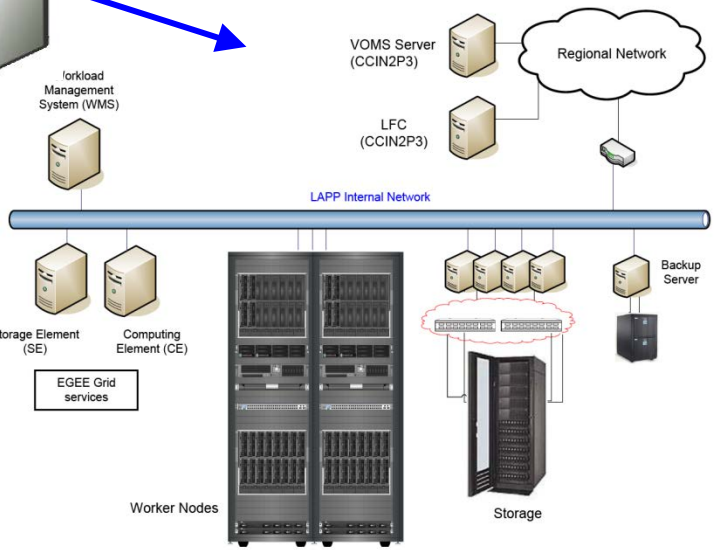
FITS and VObs data access

FITS and VO compliant results
VObs tools (VOSpec, Aladin)

Interface layer with archive DB

Web Server and Grid jobs broker

Grid jobs management



Science Archive,
DB and MetaDB systems,
notification and results uploading

Grid data operation system

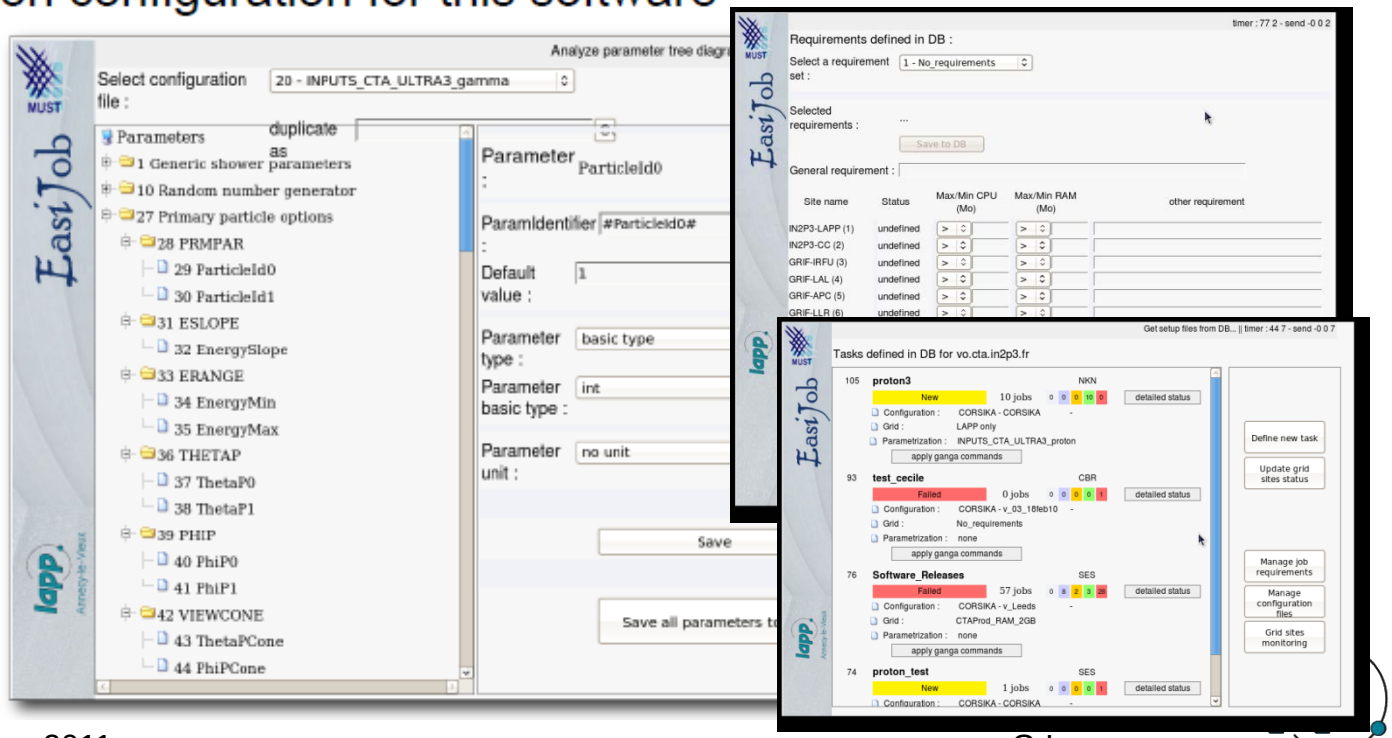
EasiJob

Easy Integrated Job Submission
(developed at LAPP and based on ATLAS-Ganga)
(DIRAC is also explored more recently)

… although aimed for CTA-MC is conceived to configure any workflow: MC, data reduction, calibration, reconstruction, analysis.

- description of parameters of a software
  - set of parameters, with default values, define if browsable
  - representation in data base
- create task based on configuration for this software
  - change parameters
  - define number of jobs
- web interface:

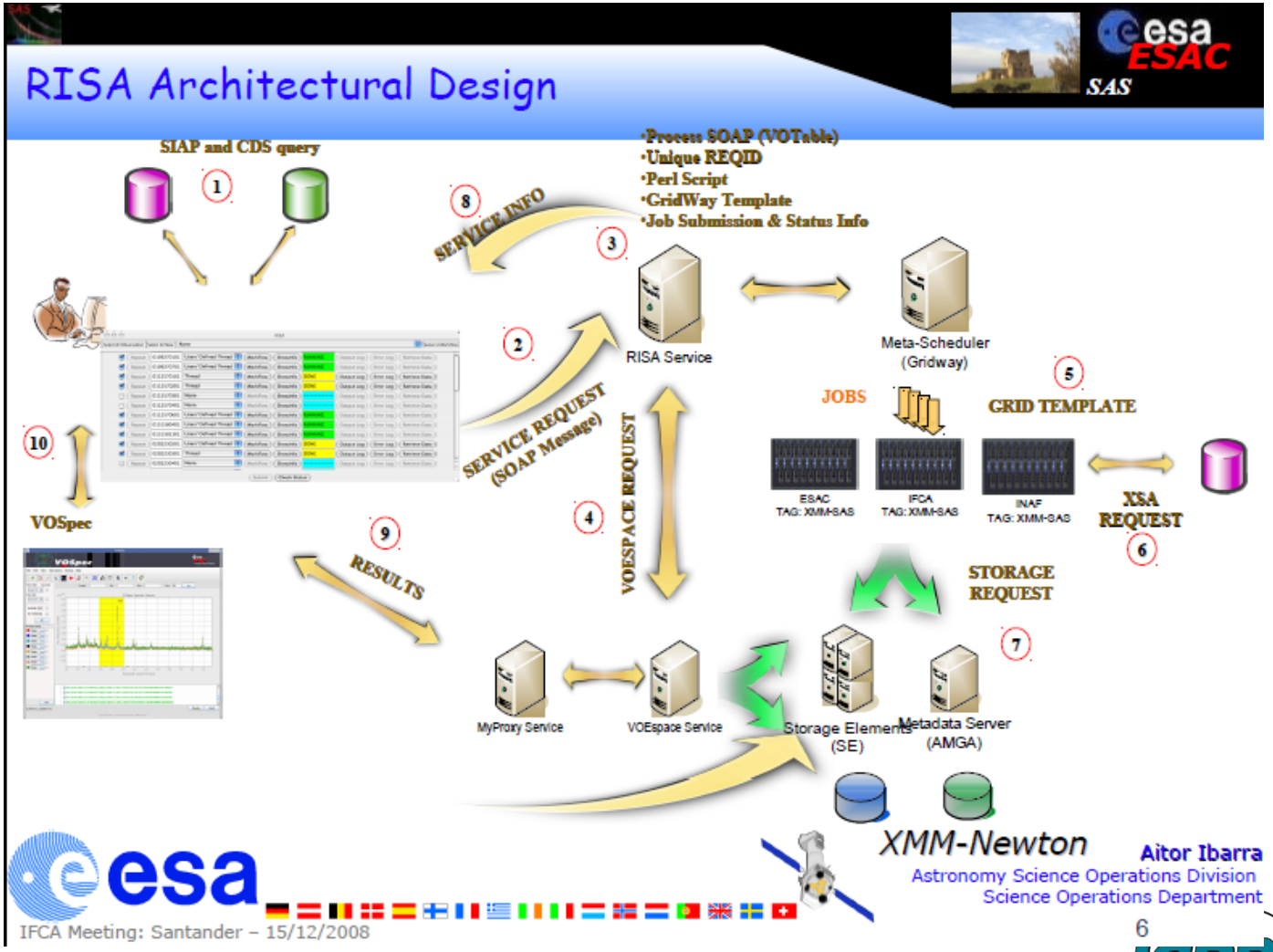
- Quality check of data pipelines Grid jobs
- Monitor workflow through Grid Worker nodes



Grid – SAS (XMM RISA)

RISA (Remote Interface for Science Analysis) within the XMM-Newton Science Operations Centre at ESAC / ESA = A Scientific Analysis System (SAS) through web services.

A SAS Remote Interface that allows users to reduce and analyze XMM Newton data, using a Client/Server application which runs the processes in a GRID architecture.



Summary

Exper.	Data rate (GB/day)	Data rec. (GB/day)	MC	Data transfer	Recon.	Access	Model
AUGER	5	1	Grid	50 kB/s + tapes	Tier 0 (CCIN2P3)	Member users	**Centralized-T0 MAINFRAME :** Standalone server + Domain controller. Database-centric architecture
ANTARES	80000 to shore unfi	20 filt. +30 unf.	Central/ (Grid)	1GB/s to sh. 1 Gb/s to CC	Tier 0 (CCIN2P3)	Member users	
FERMI	15	750 (L1)	Central	8x2 GB/day download	Distributed (SLAC+CC)	Users + Observers	Pseudo-P2P (2T0 ?); C/S jobs execution (logical multi-tier); DB-centric + high-lvl data c/s Obs. Access
HESS	<100	<10 (DST)	Central	Tapes	Two // (CC+HD)	Member users	**Centralized-T0 MAINFRAME** + c/s Obs.-like services
CTA	3-30 10^3	<3-30 10^3	Grid	?	?	Users + Observers	Client/server **multi-tier architecture** (e.g. CTACG-MC-GRID and XMM) **?**

Summary and conclusions

Astroparticle “event-experiments”:

- Recon./Analysis :Not specific “complicated’ algorithms (not really computing demanding)
- Relatively low rate, fast reco., calibrations more critical (Grid-MC and re-processing), database-centric.
- Computing architecture is more affected by data acquisition and data transfer issues. (services applied and available at CC: XRootd, SRB, HPSS, Oracle DB, dCache, …. and Grid middleware)
- Data access issue becomes critical for the Observatories (data type, services, users/observers) .
- ESA and NASA constrain (and support) the need for rapid processing and open access….
- HESS model is already “observatory-oriented” but still centrally manageable.
- ESA (e.g. XMM) applies GRID-paradigm for SAS. (GRID as a secure and mature distributed batch system moving to cloud = Grid + simplified porting non-batch applications)
- CTA Observatory….. will implement the first “ASPERA-Observatory” computing model for Astroparticle Observatory Data Center ?

-> A **multi-tier architecture /client–server architecture** “*in which the presentation, the application processing, and the data management are logically separate processes*” and modern ICT solutions

Thank you !