Migration of the MAGIC Datacenter and Monte Carlo simulation to a Grid infrastructure

Presenter: Roger Firpo (PIC)

A.Vazquez (INSA), J.L. Contreras (UCM),
A. Moralejo (IFAE), A. Ibarra (INSA), E. Acción (PIC),
I. de la Calle (INSA), P. Casatella (UCM)

www.eu-egee.org

Summary

- MAGIC
- Grid opportunities for MAGIC
- MAGIC Monte Carlo production
- Migration of the data center to Grid

What is MAGIC?

MAGIC is a Cherenkov telescope system for g-ray astronomy in the very high energy range (VHE, E > 25 GeV)

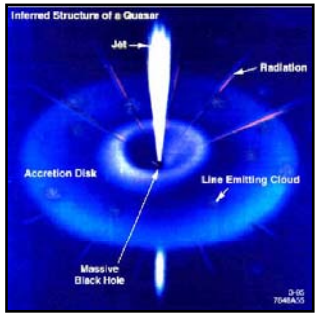
Scientific targets

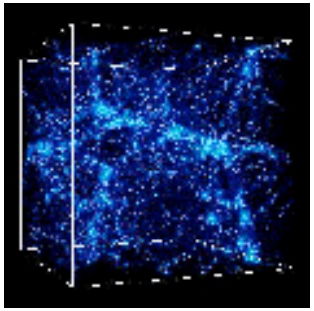
Cosmic Accelerators

- AGN, PWN, SNR, GRB …

Fundamental Questions

- Dark Matter, Cosmic Rays, Quantum Gravity, Cosmology …





The MAGIC Collaboration

The MAGIC Collaboration:

21 institutes (mostly in Europe)

~ 200 members

Telescope site in Canary Islands

Observatorio Roque de los Muchachos

- **MAGIC I operating since 2004**
- **MAGIC II in commissioning (2009)**

Future detector enhancements

Equip MAGIC I with same camera and readout as MAGIC II

Scientific Highlights

- Discovery of 10 new VHE **g**-ray sources
  - 7 extragalactic + 3 galactic
- New populations unveiled
  - Radio-quasar & Micro-quasar
- Detection of distant VHE **g**-rays
  - z = 0.54, farthest up to now
- Observation of GRB in prompt emission
  - No VHE g-ray detections so far
- Test on Lorentz Invariance (QG effects)
  - Using big emission flares
- More than 30 published papers
  - and many more are in the pipeline

MAGIC Data

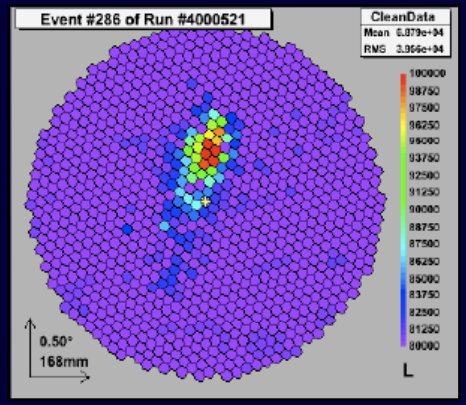
MAGIC records Cherenkov light flashes from g-ray induced atmospheric particle showers

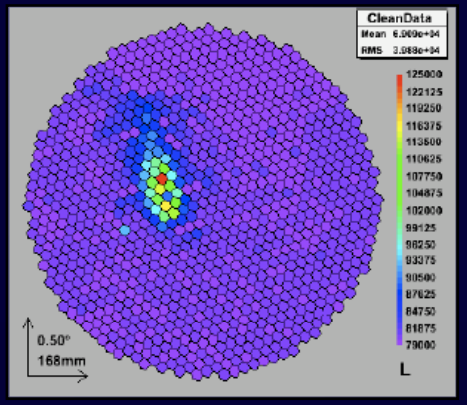
Major issue: Background rejection

- Separate g-rays from hadrons
- Based on image parameters

Monte Carlo simulations required

- No VHE “test beam” available





MAGIC Data Center @ PIC

- MAGIC produces **100 TB of raw data each year**
  - And up to 400 TB in the final configuration
- The **MAGIC data center at PIC** provides:
  - Data transfer from ORM and storage
  - Data reduction
  - User access and support
- PIC data center operating since 2006
  - Two telescope hardware upgrades
  - A second telescope in commissioning
- A data center **upgrade is needed!**

The MAGIC VO

MAGIC VO exists since 2004

- Initiative by H. Kornmayer et al.

Hiatus

- Main actors left the project
- Grid was no priority within the collaboration
- No manpower

2007-08: New crew taking over grid operations

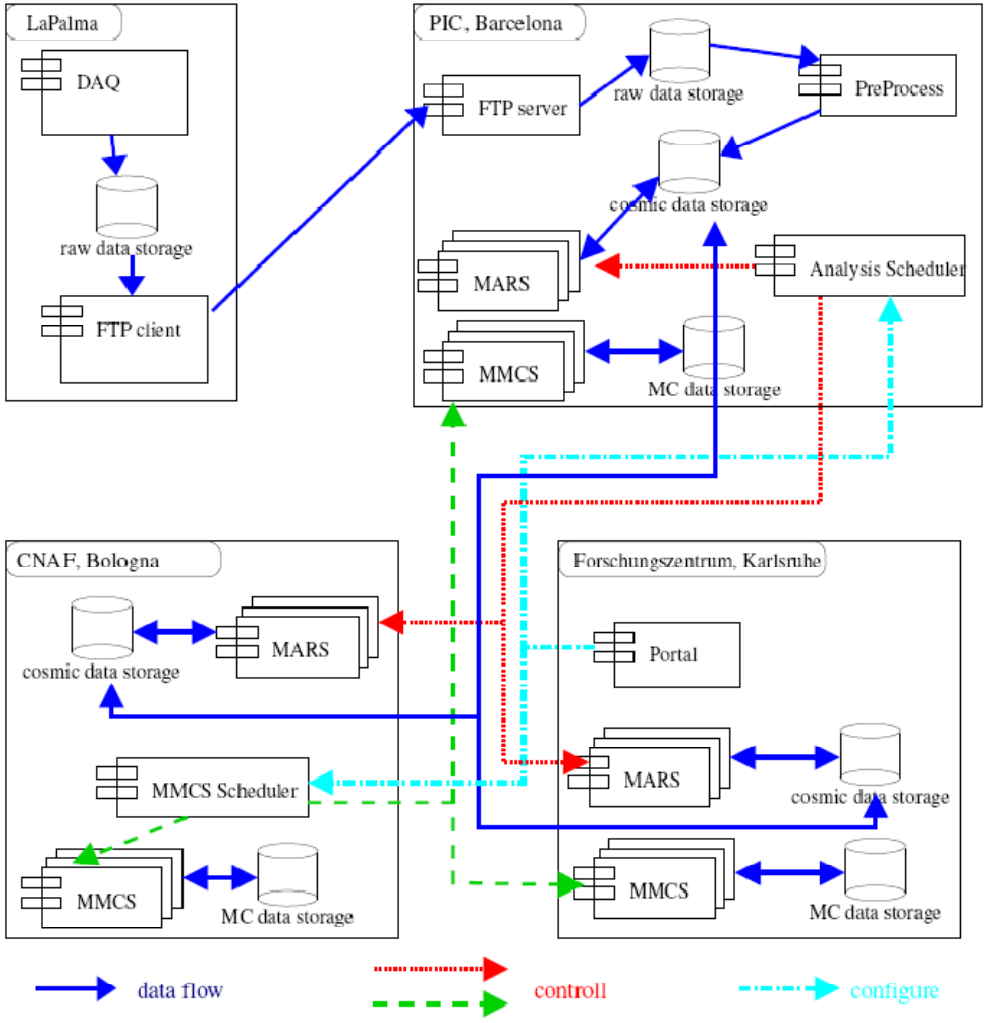
- UCM (Madrid) and Dortmund, in collaboration with INSA
- IFAE and PIC

Opportunities for MAGIC in GRID

- **Why GRID?**
  - Monte Carlo production and data reduction require lots of CPU
  - Data has to be distributed to all collaborators across Europe
  - Improved control over analysis & MC production control
  - User access to shared resources and standardized analysis tools
  - Better and easier data management
  - Increased technical support, benefit from LCG experience
- **How to proceed?**
  - Resume development of MC tools and start MC production
  - Migrate data to a grid-aware file system
  - Use grid tools for data transfer and distribution
  - Migrate existing analysis tools to grid & create new
  - Interfaces to access data, monitor jobs & transfers …
  - **BUT: Convince users to use this tools!** Training…

MAGIC in Grid (2004)

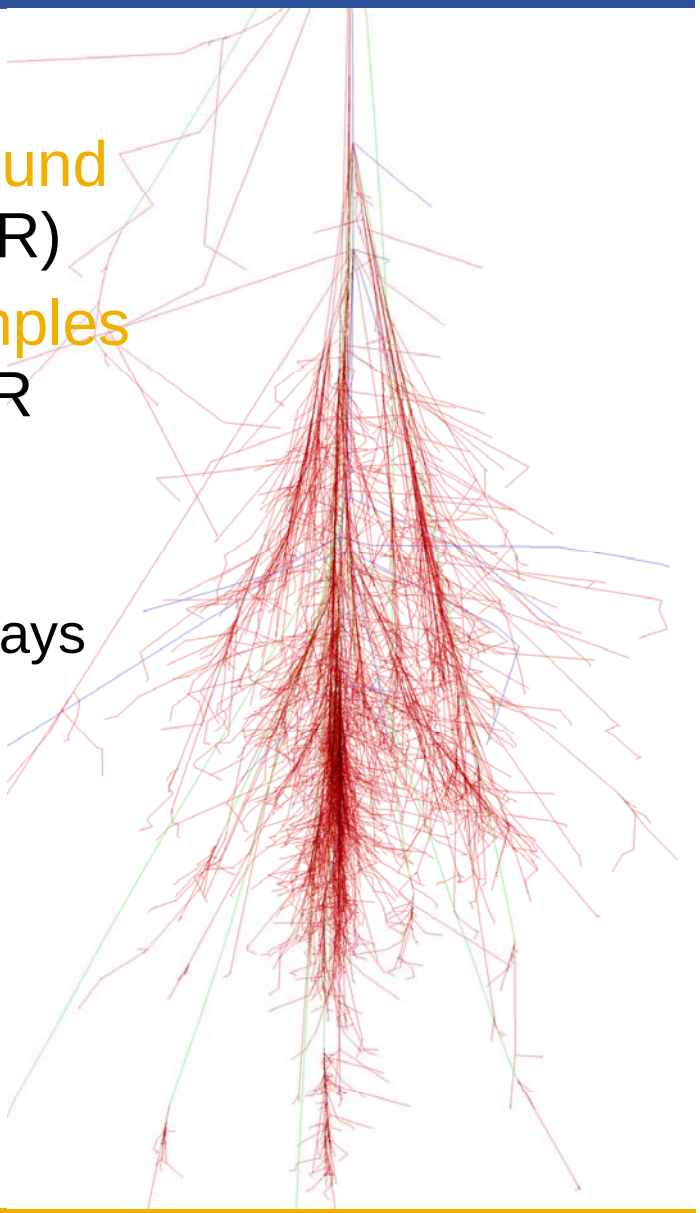
- H. Kornmayer proposed a workflow for MAGIC VO
- Involve 3 national centers
  - CNAF (Bologna)
  - PIC (Barcelona)
  - GridKA (Kalsruhe)
- Connect MAGIC resources
- 2 subsystems:
  - Monte Carlo
  - Analysis
- Start with MC first



MAGIC Monte Carlo production

Monte Carlo production

- The recorded data are mainly background events due to charged cosmic rays (CR)
- Background rejection needs large samples of Monte Carlo simulated **g**-ray and CR showers
- Very CPU consuming
  - 1 night of background > 10^6 computer days
- Access to simulated samples, MC production coordination, scalability (MAGIC II, …)

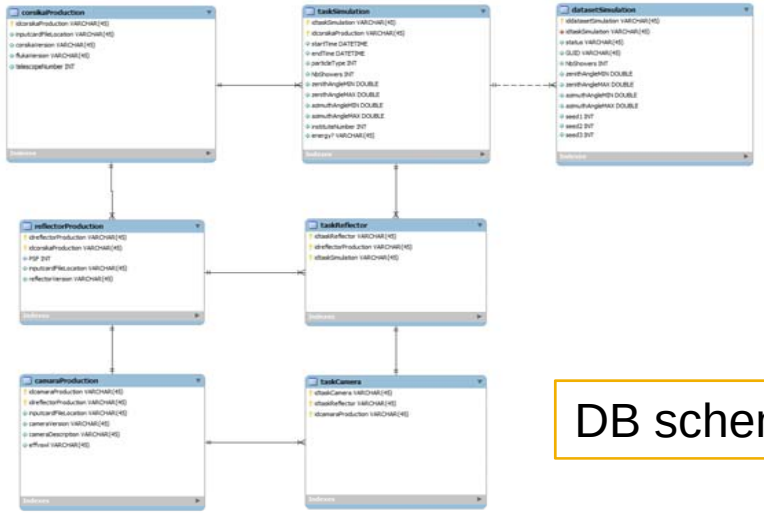


GRID can help with these issues

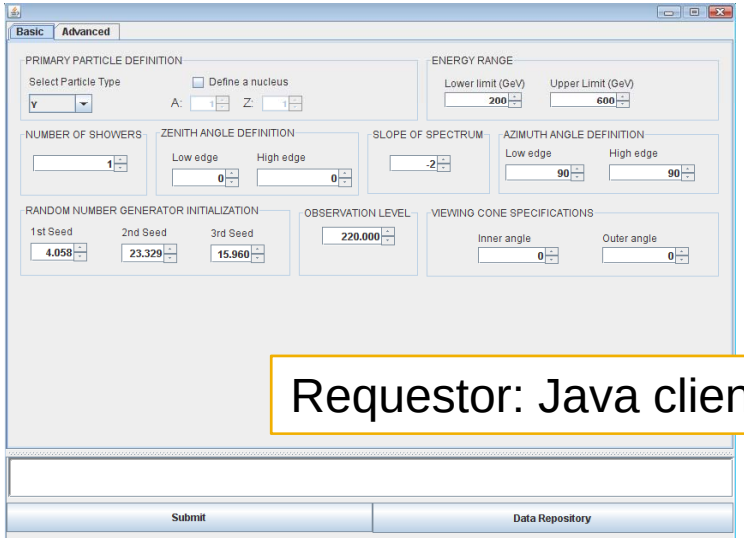
Implementation

3 main components

- Meta data base
  bookkeeping of requests, jobs and data
- Requestor
  Users insert requests to the meta data base with MC parameters
- Executor
  creates Grid jobs by checking the metadb and generating the input files

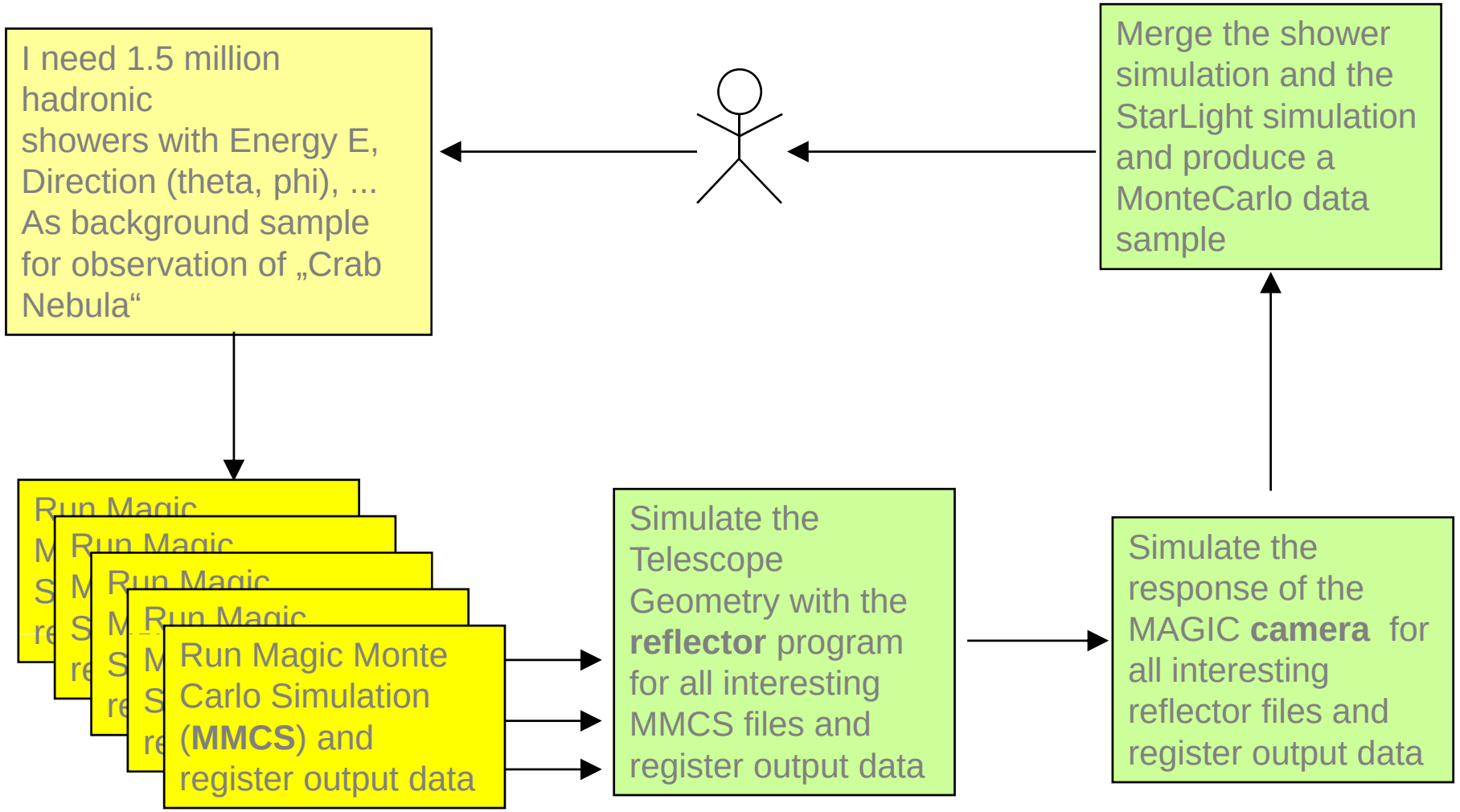


DB scheme



Requestor: Java client

MC workflow


I need 1.5 million hadronic showers with Energy E, Direction (theta, phi), ... As background sample for observation of „Crab Nebula“
Merge the shower simulation and the StarLight simulation and produce a MonteCarlo data sample
Run Magic Monte Carlo Simulation (**MMCS**) and register output data
Simulate the Telescope Geometry with the **reflector** program for all interesting MMCS files and register output data
Simulate the response of the MAGIC **camera** for all interesting reflector files and register output data

Monte Carlo production

- 2004: MC production workflow and first tool prototypes
- 2005: Production test

  …
- 2007-09
  - UCM + UT Dortmund + INSA retake the project (2008)
  - Development of a web interface to manage production
    - Java-based
    - Manages job configuration, submission and monitoring
    - GridWay used as a metascheduler
  - MC test production w/ reduced set of resources
    - PIC, CIEMAT & UT Dortmund clusters
    - Still some technical difficulties

- Plan to start producing MC for MAGIC-II soon

MAGIC data center@ P IC

MAGIC Data Center

The MAGIC data center is hosted by PIC in Barcelona

- PIC is the spanish Tier-1 for LHC

Data center services:

- Data transfers from ORM and storage
- Computing (internal data center)
- User data access and support

Challenges faced

- Second telescope and upgrades: Increase in data volume
- Scalability: Increase in complexity and maintenance time
- Storage and Computing: Increase and optimize resources
- Users: Improve data access, open computing resources

Data Center: Data volume

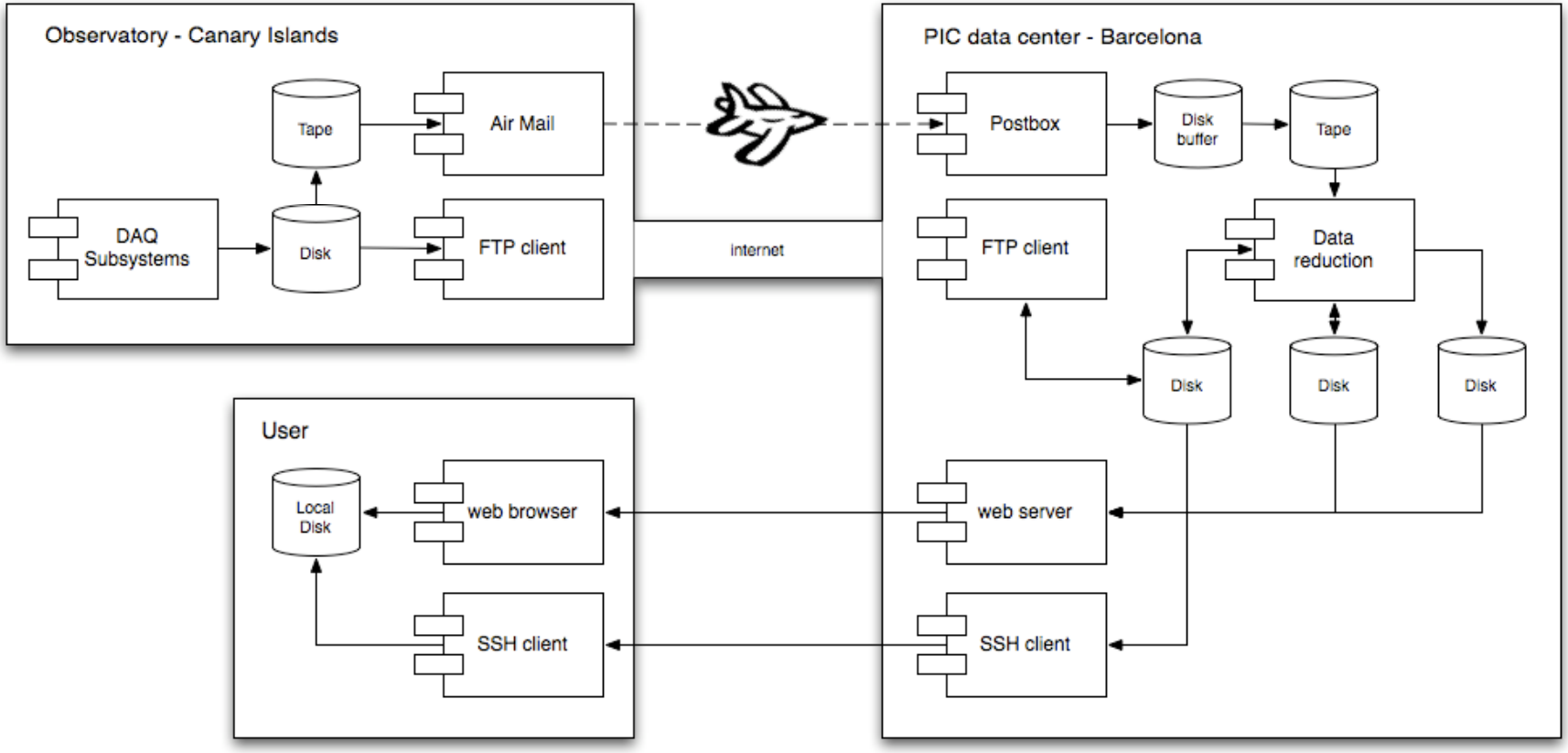
Increase in data volume foreseen in near future

Telescope system	MAGIC I	MAGIC II	MAGIC I + II	2 x MAGIC II
	2004 -	2009	2009	2011?
# of channels	577	1081	1658	2162
Event size (kB)	60.7	110.0	170.7	219.9
Event rate (Hz)	350			
Data Rate (GB/h)	73.0	132.1	205.1	264.3
Obs. Time (h/yr)	1500			
RAW data (TB/yr)	106.9	193.6	300	387.1
RAW.gz (TB/yr)	32.1	58.1	90.2	116.1
Reduced data (TB/yr)	4.1	7.1	12.3	15.3

In the next ~3 years data volume will increase 4-fold

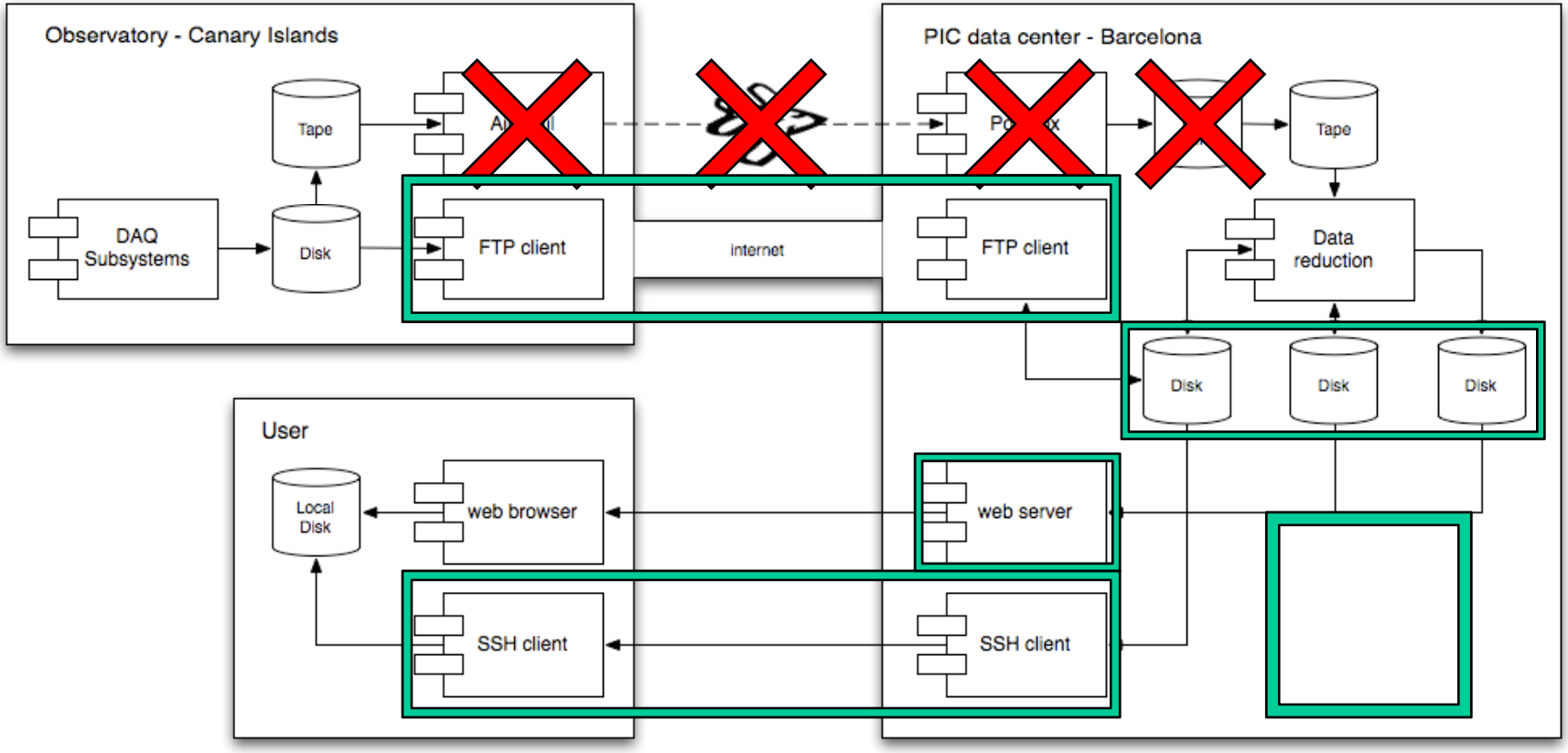
Data Center: Old data flow

- **Data flow**
  - Current scheme is obsolete: scalability problems
  - Maintenance has become a major time-eater



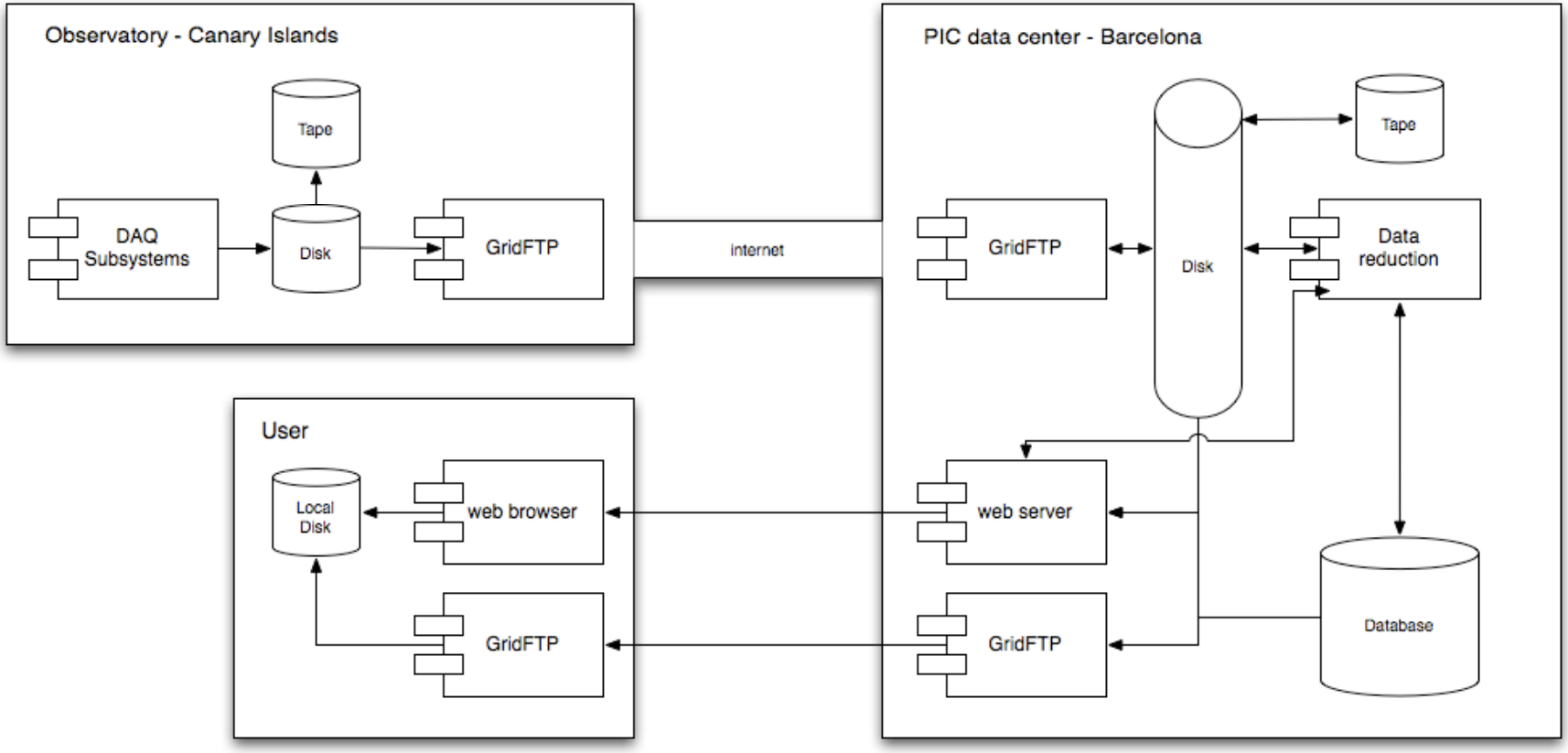
Data Center: New data flow

- **Data flow optimization**
  - Deprecate classic transfer methods in favor of Grid
  - Simplify flow and optimize resources



Data Center: New data flow

- **Goals**
  - Easy management and service monitoring
  - Better user experience



Data Center: Storage

- Current storage system requires too much maintenance
- Non-existent file catalog, requires custom tool development
- **Solution**: adopt Tier1-grade Grid-based storage system
  - Standard tools + supported service @ PIC
  - LFC: Easier data management and monitoring

	Old scheme	New scheme
Storage	NFS + CASTOR	dCache (w/ ENSTORE)
Tape access	Different protocol	Transparent to user
Maintenance	High, custom tools	Low, PIC service
Data catalog	-	LFC
Security	No access control	User certificates, VO roles, …

Transition to new scheme will be done while in production

Data Center: data access

- **Data access requirements:**
  - Access data anytime from anywhere
- **Two approaches:**
  - Data access using GridFTP or equivalent
    - Robust transfers, not easy file browsing
    - BUT: Not all institutes support Grid
  - Web access
    - Easy file browsing, not that easy transfers
- **Solution:**
  - Build web-based service to interface to GridFTP
  - Use httpdoors as backup solution & for “Grid-handicapped” users
  - Use LFC + project database as backend

Data Center: Computing

Computing at MAGIC now

- Each institute uses its own computing resources (CPU + Storage)
- Only few can access a computing farm
- Data center CPUs exclusive to “official” analysis

We go towards **opening the computing service to all users**

- Grid-based computing
- Universal access to data in the SE and use of the CE (and +CEs)
- Standard analysis tools
  - Job submission and management using a web UI
- PIC data center will still play a central role
  - Data management, manpower, …

+ resources & efficiency: **more and better scientific outcome**

Summary

- **MAGIC has resumed Grid activity**
- **Grid-based Monte Carlo production system developed**
  - Systematic production will start soon
- **MAGIC is migrating data and analysis to Grid**
  - Migration of the data center already started
  - Analysis tools will start being developed soon

Related experiment

- **Cherenkov Telescope Array (CTA)**
  - Next generation of IACT
  - Big step-up with respect to current IACT
  - CTA VO already exists and active
  - PIC supports the CTA VO