cherenkov telescope array

Asterics
Astronomy ESFRI & Research Infrastructure Cluster

Big Data en Física de Astropartículas

José Luis Contreras

Grupo de Altas Energías (GAE) UCM

Indice

Objetivo: Explicar las actividades de *Big Data* en el GAE con vistas a posibles colaboraciones.

- Introducción:
  - El grupo
  - Por qué generamos muchos datos
  - Aprendizaje automático
- MAGIC
  - Qué es: volumen de datos
  - Qué hacemos
- CTA:
  - Qué es: volumen de datos
  - Qué hacemos

- Proyectos.

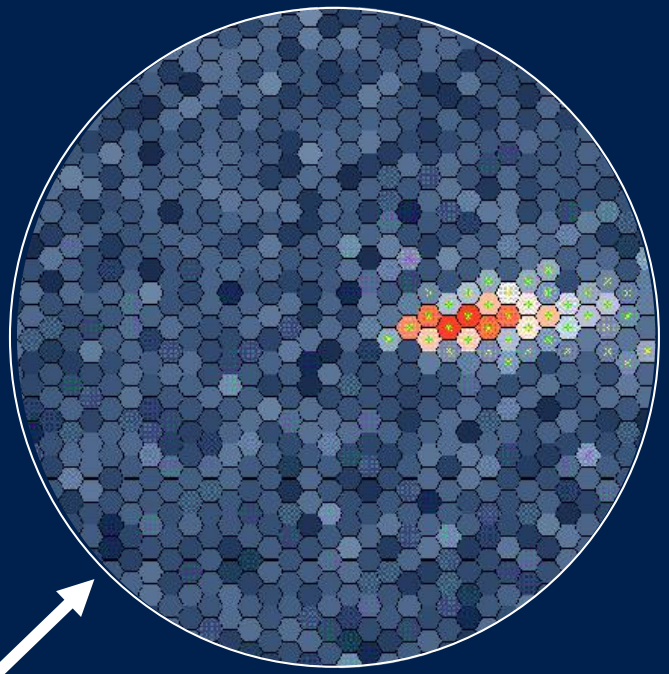
El GAE

- CC Físicas. 10 doctores (6 permanentes) + 4 estudiantes
- Investigamos en Física de *Astropartículas*.
- Observacionales. Buscamos fuentes que emitan rayos gamma de muy altas energías.
- Primera fuente (nebulosa remanente de una supernova) detectada en 1989. Actualmente más de 160..

Las bases del método

Partícula de Alta Energía

Cascada atmosférica



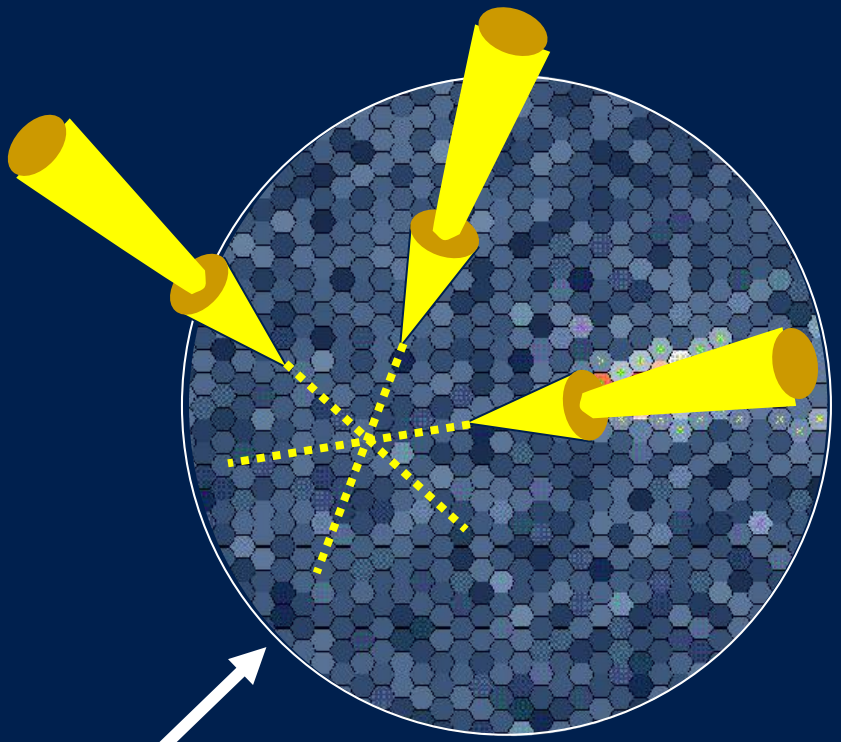
Cono de luz Cherenkov

Imagenes de baja resolución y muy alta velocidad

Telescopio

W. Hofmann

¿ Por qué tantos datos ?



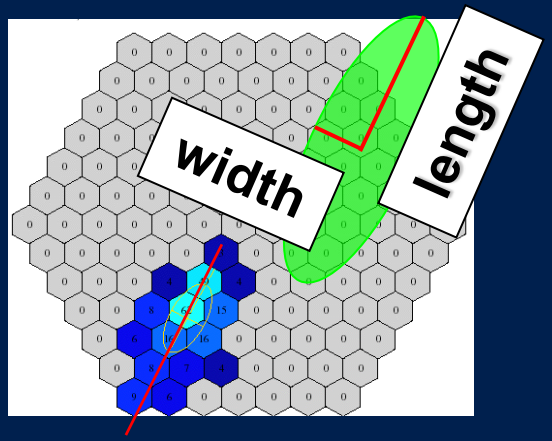
Una particula =
2-20 “microvídeos”
de 30-100 imágenes
de 1000-2000 pixels

W. Hofmann

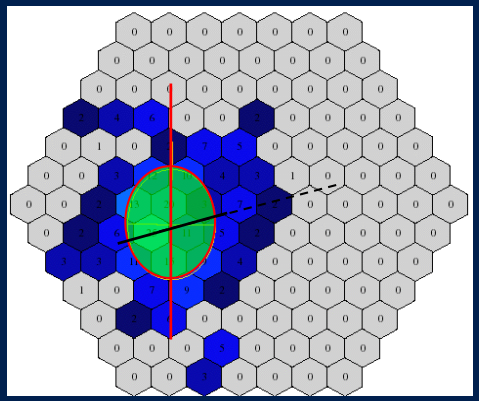
Separar la señal del fondo 1/1000

100 GeV proton

Rayos gamma
(Estrecha apunta a la fuente)



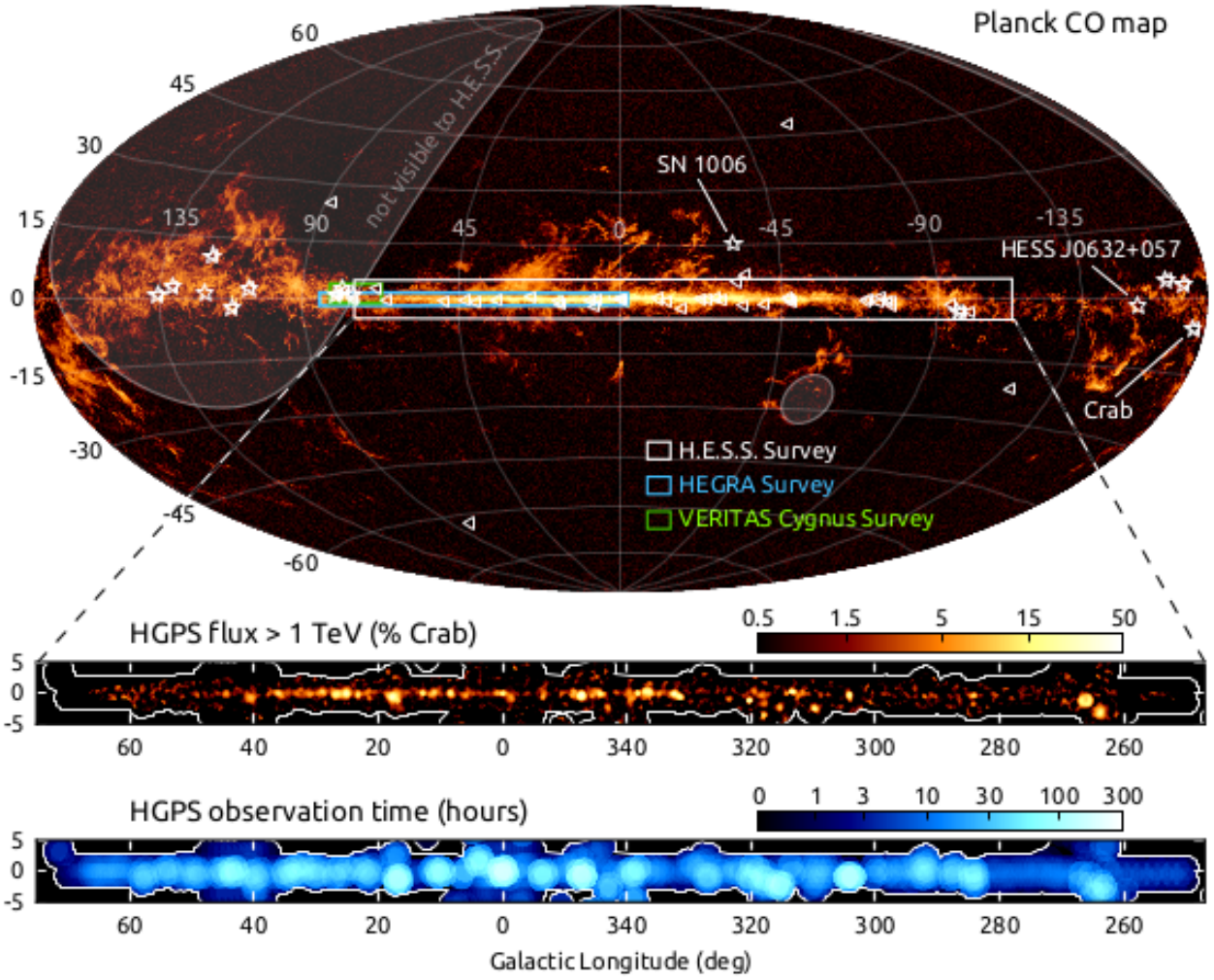
Protones…
(Ancha, orientada al azar)



Influyen otras variables → Complejo → Métodos de aprendizaje automático basados en simulaciones:

Actualmente: Random Forest → Hemos aplicado aprendizaje automático a otros campos

La Galaxia en gammas de muy alta energía



Smart Data

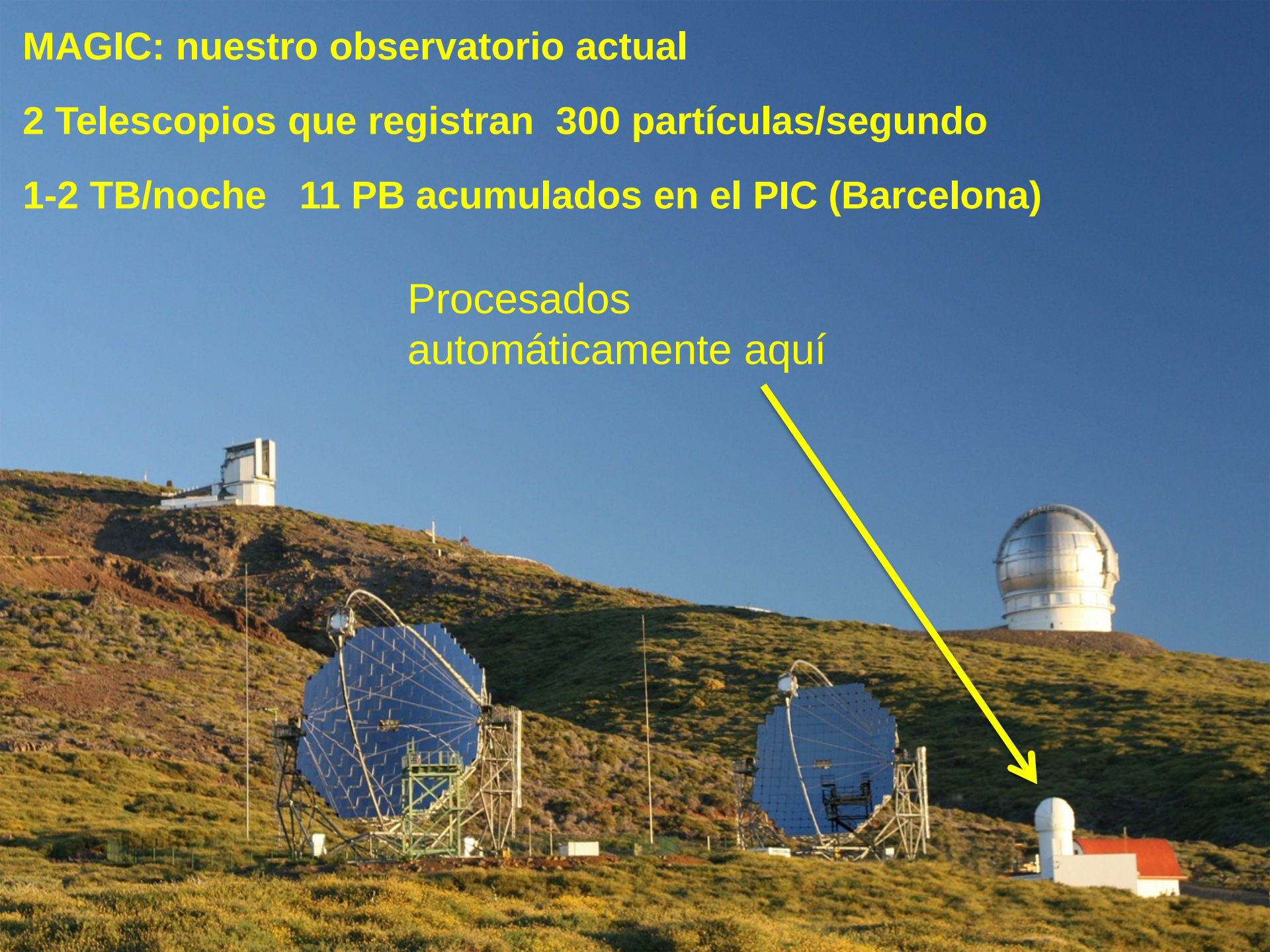
El resultado de una observación, tras filtrar y reconstruir: unos pocos miles de rayos gamma

Este es el resultado de 12 años de observaciones de HESS, la competencia

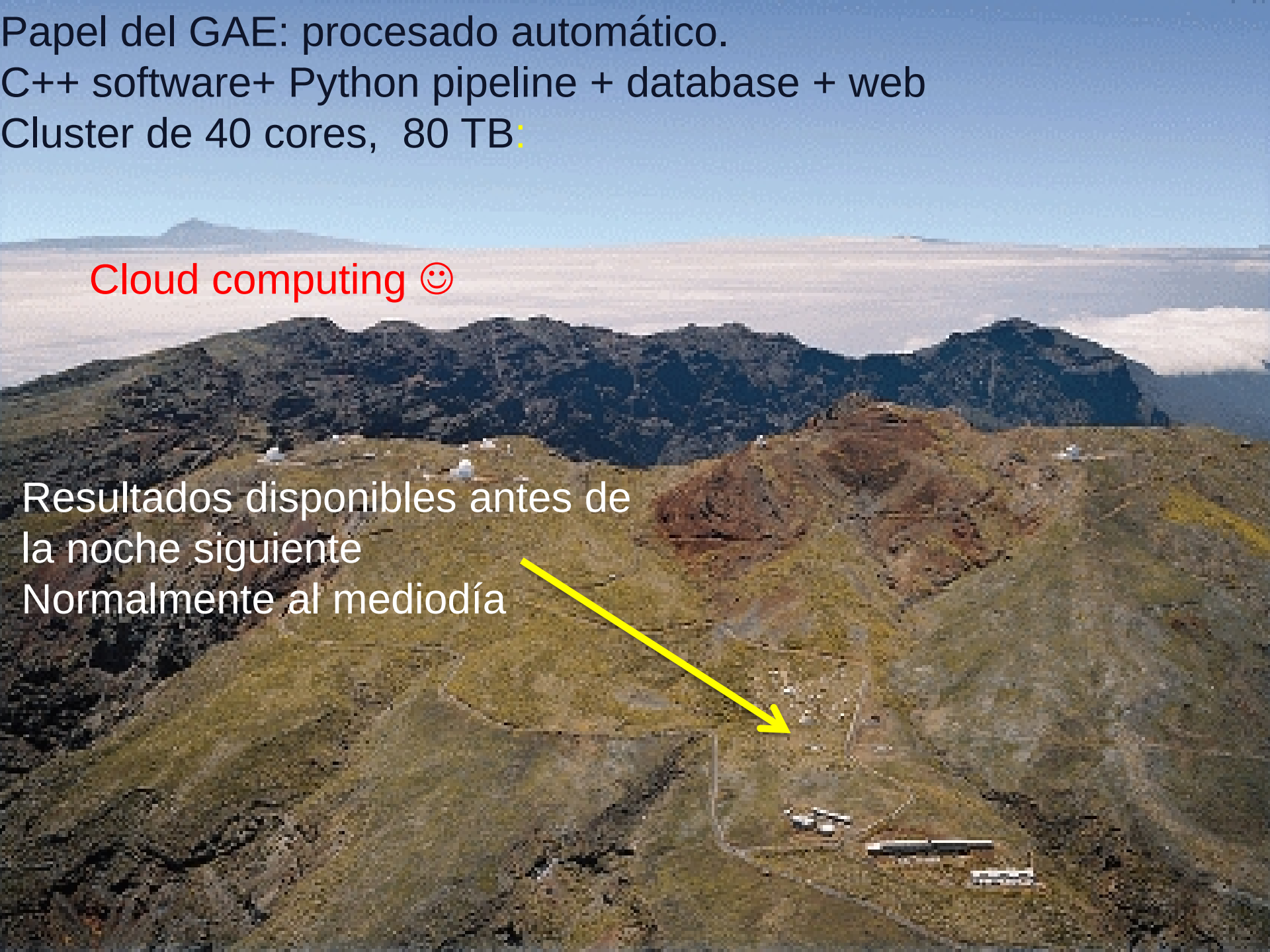
MAGIC: nuestro observatorio actual

2 Telescopios que registran 300 partículas/segundo

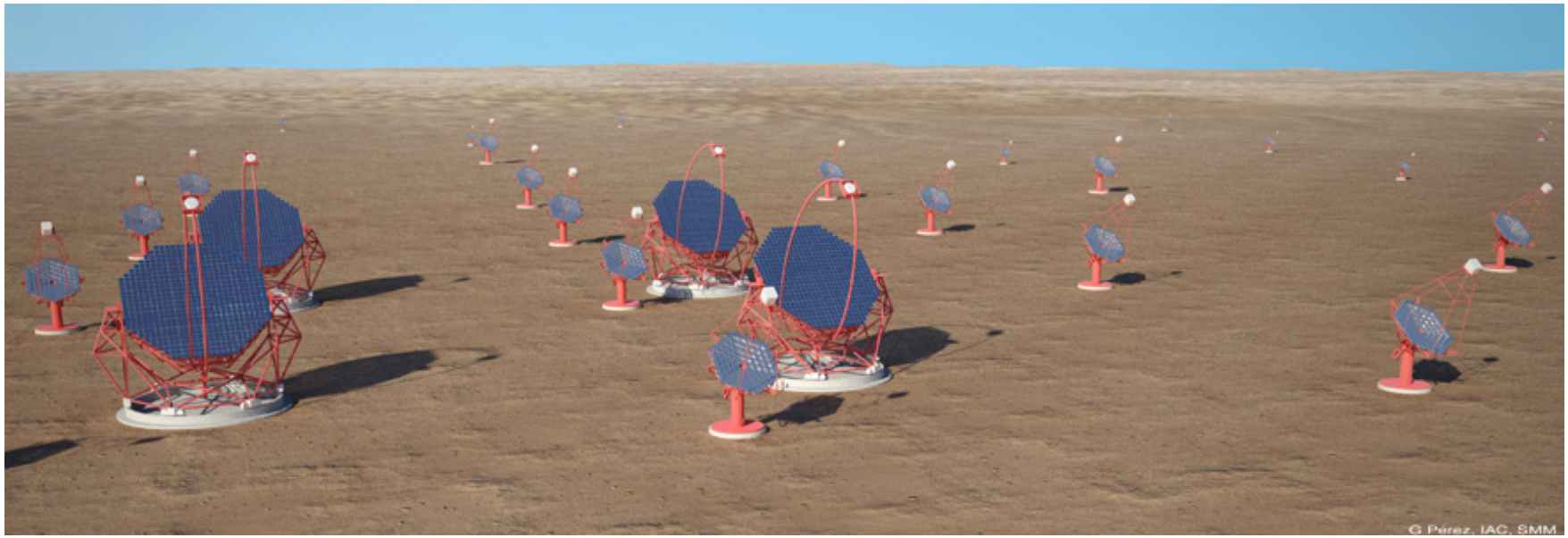
1-2 TB/noche 11 PB acumulados en el PIC (Barcelona)



Papel del GAE: procesado automático.
C++ software+ Python pipeline + database + web
Cluster de 40 cores, 80 TB:



El futuro: el proyecto CTA



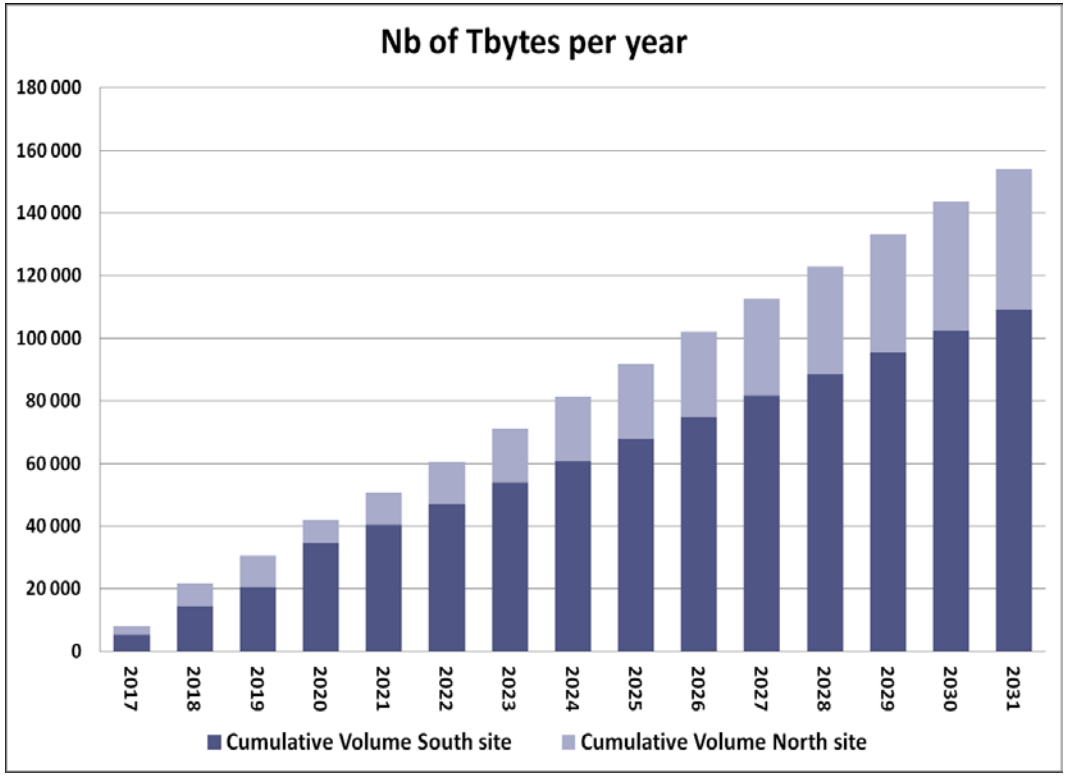
- Un observatorio en cada hemisferio para cubrir todo el cielo
  - Atacama, Chile en el Sur 50-100 Telescopios
  - La Palma, en el Norte 10-20 Telescopios
- 10 veces más sensible
- La construcción del observatorio Norte ya ha empezado en La Palma
- Observatorio abierto, datos y software ´publicos

cherenkov telescope array

Los futuros telescopios grandes de CTA

Volumen de datos en CTA

- 10000-40000 partículas/s
- Ritmo de toma de datos (Periodos activos)
  2-5 GB/s
- Volumen total en torno a 2-30 PB /año
- Actualmente 2 PB datos simulados: grid



Nuestro papel en software: Coordinar el modelo de datos:
Eg: Sistema de ficheros para datos de bajo nivel
Contenido datos de alto nivel

Proyectos

Actividad principal: Participación en MAGIC + CTA (proyectos MINECO) +

- H2020: ASTERICs: 2015-2019
  - Cluster de instalaciones científicas “reconocidas” en ESFRI del campo de la Astronomía + Física de Astropartículas:
  - Financia el **Observatorio Virtual** en Europa
  - Incluir Astropartículas: rayos gamma, neutrino, ondas gravitacionales
    - Formatos eficientes para datos de bajo nivel
    - Formato abierto de datos para alto nivel

- Deep Learning:
  - Proyecto “Jóvenes investigadores” 2017-2020
  - Idea: aplicar técnicas de deep learning a la identificación de rayos gamma en CTA , mejorar sobre el random forest

Gracias !