COSMO HUB Science Portal

Jorge Carretero on behalf of the PIC Team

Red de Infraestructuras de Astronomía

Promoviendo sinergias entre grandes observatorios españoles I

Temática: Gestión de datos científicos

22- Oct - 2023, La Palma, Canarias

Port d'Informació Científica

PIC
port d'informació
científica

Port d'Informació Científica

- An advanced and multidisciplinary computing center
- Founded in 2003: collaboration between IFAE and CIEMAT
- Team of 24 people (50% scientists - 50% engineers)
  - Contact person for each experiment
  - Agile teams that embed in scientific groups
- Run production data services for large collaborations
  - Collaborative environments, distributed infrastructures
- Flexibility to adapt to evolving needs
  - Integration of data processing tools/methods
  - Trans-disciplinar cross-pollination
  - Infrastructure - experimental can-do attitude

PIC
port d'informació
científica

PIC projects support

- Particle Physics:
  - LHC: Spanish primary center & ATLAS analysis facility
  - Neutrinos: T2K & DUNE
- Astrophysics:
  - Gamma-ray astronomy: MAGIC & CTA
  - Cosmology: DES, PAUS & Euclid
  - Gravitational waves: Virgo/LIGO
- Other data-intensive disciplines:
  - material sciences, bioimaging, genomics

MAPA DE INFRAESTRUCTURAS CIENTÍFICAS Y TÉCNICAS SINGULARES

GOBIERNO DE ESPAÑA · MINISTERIO DE CIENCIA E INNOVACIÓN

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional

Plan de Recuperación, Transformación y Resiliencia

Infraestructuras Científicas y Técnicas Singulares

EUSKADI
ReDIB - CIC biomaGUNE
NANBIOSIS - CIBER-BBN
MARHIS - BiMEP
R-LRB - LRE

GALICIA
RES - Finis Terrae (CESGA)
FLOTA- CSIC/IEO
NANBIOSIS – CIBER-BBN

CANTABRIA
RES - Altamira (UC)
MARHIS - GTIM-CCOB

NAVARRA
RES-Urederra (Nasertic)

ARAGÓN
Laboratorio Subterráneo de Canfranc (LSC)
Observatorio Astrofísico de Javalambre (OAJ)
RES - Caesaraugusta (UNIZAR)
ELECMI - LMA
NANBIOSIS - CIBER-BBN

CATALUÑA
Sincrotrón ALBA
RES - MareNostrum y MineTauro (BSC-CNS)
RES- PIC (CIEMAT-IFAE)
MICRONANOFABS - SBCNM
MARHIS - ICIEM
OmicsTech - CNAG - CRG
OmicsTech - COS
RLASB - CReSA
NANBIOSIS - CIBER-BBN
ELECMI - UMEAP
FLOTA - CSIC

Centro Nacional de Investigación sobre la Evolución Humana (CENIEH)
Centro de Láseres Pulsados (CLPU)
RES - Caléndula (SCAYLE)

CASTILLA Y LEÓN

COMUNIDAD DE MADRID
Laboratorio Nacional de Fusión (LNF)
Red Académica y de Investigación Española (RedIRIS)
RES-Cibeles (UAM)
RES-Xula (CIEMAT)
NANBIOSIS - CIBER-BBN
ReDIB - BIOIMAC (UCM)
MARHIS - CEHIPAR
R-LRB - LMR
MICRONANOFABS -CT-ISOM
ELECMI - CNME
RLASB - CISA
ReDIB - TRIMA-CNIC
IABA-Centro de Microanálisis de Materiales (CMAM)

EXTREMADURA
RES-Lusitania (COMPUTAEX)
RES- Turgalium (CIEMAT)
NANBIOSIS - CCMIJU
NANBIOSIS - CIBER-BBN

ILLES BALEARS
Sistema de Observación Costero de las Illes Baleares (SOCIB)
FLOTA-CSIC/IEO

COMUNITAT VALENCIANA
RES - Tirant (UV)
MICRONANOFABS - NF-CTN
NANBIOSIS - CIBER-BBN
ReDIB - Imaging La Fe

CANARIAS
Plataforma Oceánica de Canarias (PLOCAN)
Observatorios de Canarias (OOCC)
Gran Telescopio CANARIAS (GTC)
RES - La Palma (IAC)
MARHIS - PLOCAN-TS

ANDALUCÍA
Reserva Biológica de Doñana (RBD)
Plataforma Solar de Almería (PSA)
Observatorio Astronómico de Calar Alto (CAHA)
Radiotelescopio IRAM 30M
RES - Picasso (UMA)
IABA- Centro Nacional de Aceleradores (CNA)
ELECMI - DME-UCA
NANBIOSIS- BIONAND

REGIÓN DE MURCIA
Infraestructura para el Cultivo del Atún Rojo (ICAR)
FLOTA - BIO Hespérides

CASTILLA - LA MANCHA
Observatorio de Yebes

ANTÁRTIDA
Base Antártica Española Juan Carlos I (BAE-JCI)
Base Antártica Española Gabriel de Castilla (BAE-GdC)

"Una manera de hacer Europa"

ICTS CON LOCALIZACIÓN ÚNICA | ICTS DISTRIBUIDAS | RED DE ICTS

e-NIPO: 8312201X

Red Española de Supercomputación

- ICTS distribuida por toda la geografía española
  - 14 nodos interconectados
- Ofrece recursos y servicios de supercomputación, gestión de datos e inteligencia artificial a proyectos científicos y tecnológicos innovadores y de alta calidad
  - Convocatorias competitivas

What do we do at PIC?

- Data transfer
- User support so they can process and analyze the data
- Store data and/or results
- Data distribution

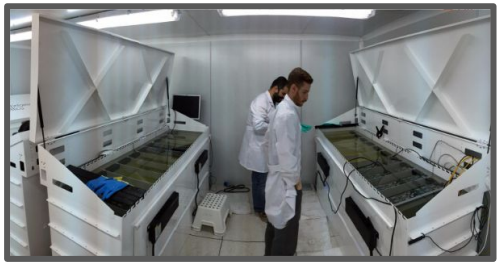
- Provide tools to enable science
- Promote the dissemination of the science we work with

PIC data center

- Facilities, ~150 kW IT
  - ~120 kW in 150 m^2 air-cooled room
    - high efficiency, PUE 1.44
  - ~30 kW in 25 m^2 liquid immersion cooling system
    - PUE 1.1

IBM TS4500

- Data processing services
  - Disk - dCache: **20 PB**
  - Tape - Enstore: **63 PB**
  - Computing - HTCondor: **12000 cores**, 18 GPUs
  - Computing - Hadoop: 720 cores, 2.5 PB net storage



- Connectivity
  - 2x100 Gbps to Academic Network
  - **Largest data mover in Spanish academic network**: **100PB** in+out per year

PIC
port d'informació
científica

PIC user services

- Computing, data processing and analysis
  - Batch processing through HTCondor
  - Notebooks ecosystem: Jupyter (Dask / *Spark*)
  - CosmoHub
- Mass storage
  - Tape
  - Disk
- Web Services (Gitlab, Wiki, Redmine, Webdav, Monitoring, etc.)
- Consulting support

What do we do at PIC?

- Data transfer
- User support so they can process and analyze the data
- Store data and/or results
- Data distribution

- Provide tools to enable science
- Promote the dissemination of the science we work with

What do we do at PIC?

- Data transfer
- User support so they can process and analyze the data
- Store data and/or results
- Data distribution

- Provide tools to enable science
- Promote the dissemination of the science we work with

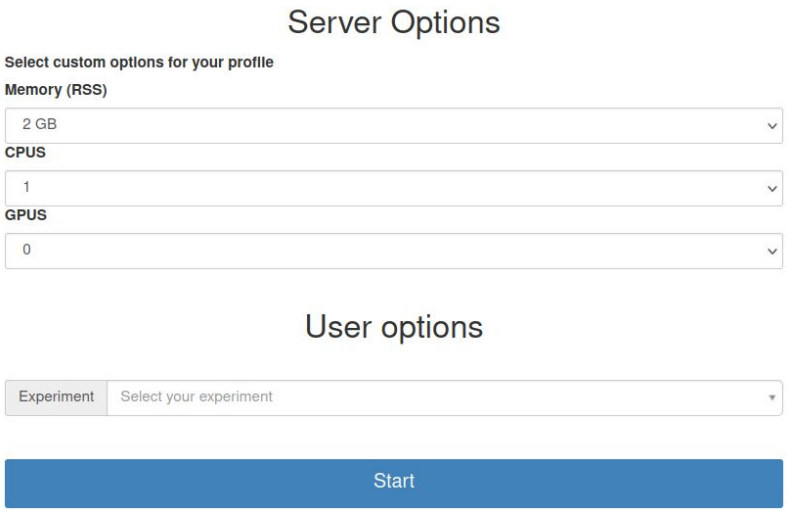
PIC
port d'informació
científica

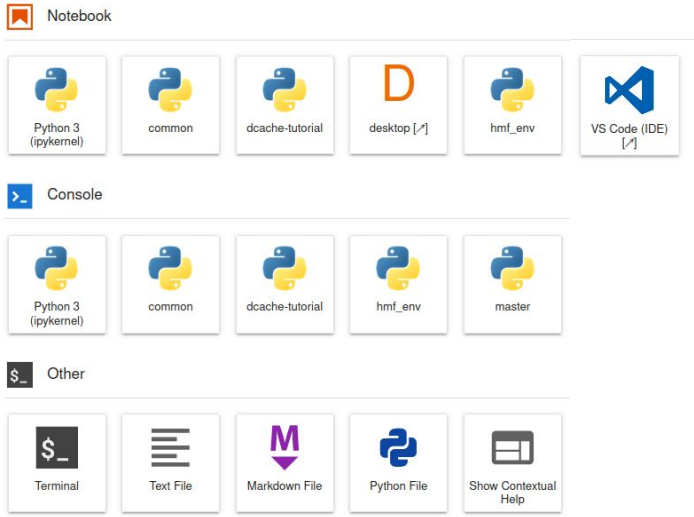
Computing key recommendations

- Define and plan the data lifecycle
- Favor distributed architectures
- Ecological awareness
- **Bring user to the data**
  - **open & collaborative environments**
- People are very valuable assets

PIC
port d'informació
científica

Jupyter notebook

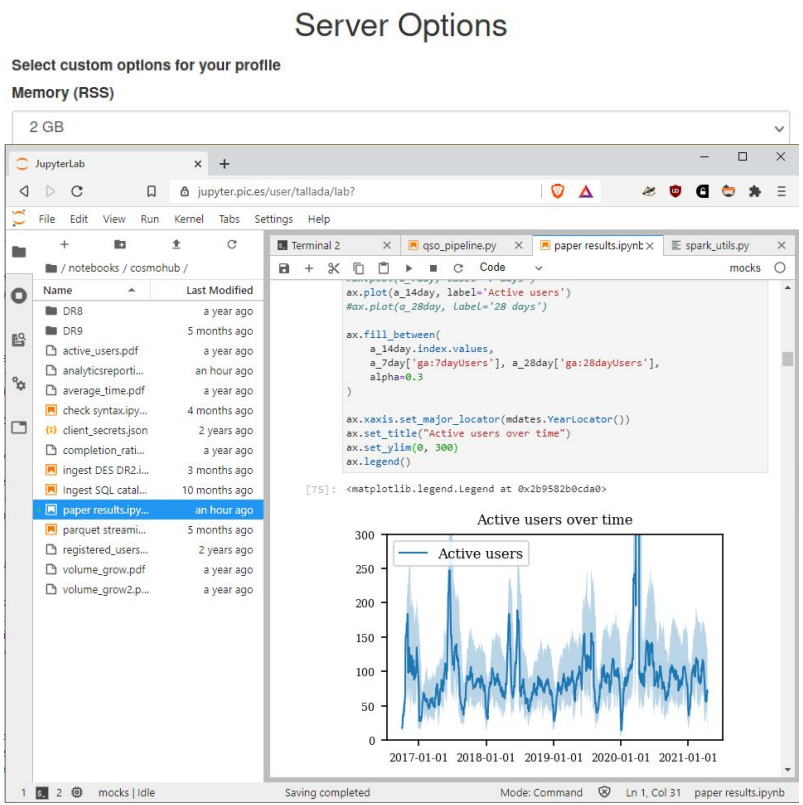


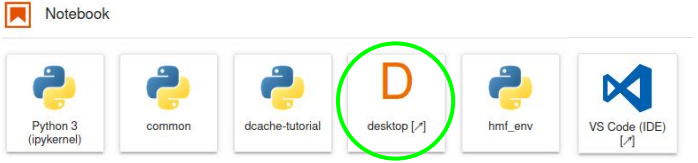


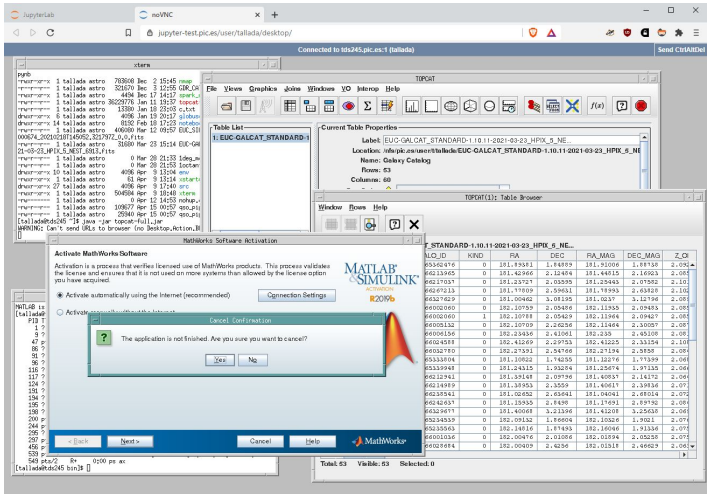
https://jupyter.pic.es/

PIC
port d'informació
científica

Jupyter notebook







https://jupyter.pic.es/

- **Interactive exploration** (visualization)
  - **Very fast (85% plots < 30s)**
  - **Full dataset plots** (over all rows)
    - May use sampling
    - Cone search tool
    - 1D histogram & 2D heatmap
  - **Guided process (no SQL knowledge required)**
    - Expert mode
    - Custom UDFs: healpix, geometric...
- Distribution
  - **Parquet, CSV, FITS format** (custom SerDe)
  - Email with a link to download dataset
- Based on Python / Flask + Hadoop / Hive

- Data
  - 40 TiB catalogued data
  - >100 catalogs (simulated and observed)
  - Supporting multiple projects
    - DES, PAU, Euclid, MICE and Gaia
- Users
  - >1500 usuarios registrados
    - ~150 active users worldwide
  - >13.5K custom catalogs generated
  - >20k interactive queries
- Performance
  - **>75% custom catalogs finish in <3 min**
  - Resource queues with reservation
  - Preemption to keep interactive response times

PIC
port d'informació
científica

Science Portal: service improvements

- Multi-instrument, -frequency, -messenger
  - GW, optical, Gammas, Neutrinos, etc.
- Integrate with Jupyter notebooks
- Guided data products:
  - Spatial / time cross-match
- Additional UDFs functions:
  - geometric, array, footprint, MOCs
- Federated identity with EduGAIN
  - Per-user storage, quotas & accounting
  - Each user has its own schema (MyDB)

- VO protocols:
  - TAP (ADQL, VOTable, ...)
  - UWS, VOSpace, ...
- New frontend
  - more plot types (healpix maps)
  - improved density/performance
- Social features
  - Catalogs and plots can be linked/shared
  - Public permissions to edit metadata (with manual moderation)

PIC
port d'informació
científica

Challenge 1: How to handle massive data

- Data volume evolution: from a laptop, to a data center, to the cloud?
- Goals
  - Scalable storage
  - Short iteration cycles
    - Fast access
    - Efficient analysis
  - Reproducibility
  - Traceability
- Our approach
  - Hadoop, Hive, Spark (code to the data!)
  - Multidisciplinary teams
    - so that scientists and software engineers understand each other
    - stand behind developers / Tutoring / Teaching (Bootcamps, MOOCs)
  - Jupyter+Spark notebooks, jupytext, git+LFS
  - Optimize algorithms
    - Custom algorithms (spark + treecorr)

PIC
port d'informació
científica

Challenge 2: Gaining (more) independent users

- Goal:
  - Reach a broader community
  - User should not notice the transition from laptop, to cluster, to grid
  - Simple, usable interfaces
    - Guided processes
    - Avoid configuration files and terminals
    - While still providing access to advanced features for expert users
- Our approach
  - Interfaces
    - CosmoHub
    - JupyterHub (+ extensions: i.e. VNC)
    - HTTP/Webdav
    - VO protocols
  - Training:
    - Documentation / Quick start guides (e.g. HTCondor / Spark)
    - Bootcamps / MOOC

PIC
port d'informació
científica

Thanks for your attention
Questions?

Credits to: E. Acción, V. Acín, C. Acosta, A. Alou, A. Bruzzese, J. Carretero, J. Casals, R. Cruz, M. Delfino, J. Delgado, M. Eriksen, D. Graña, J. Flix, E. Johana, G. Merino, C. Neissner, A. Pacheco, C. Pérez, A. Pérez-Calero, E. Planas, M.C. Porto, J. Priego, P. Tallada, F. Torradeflot

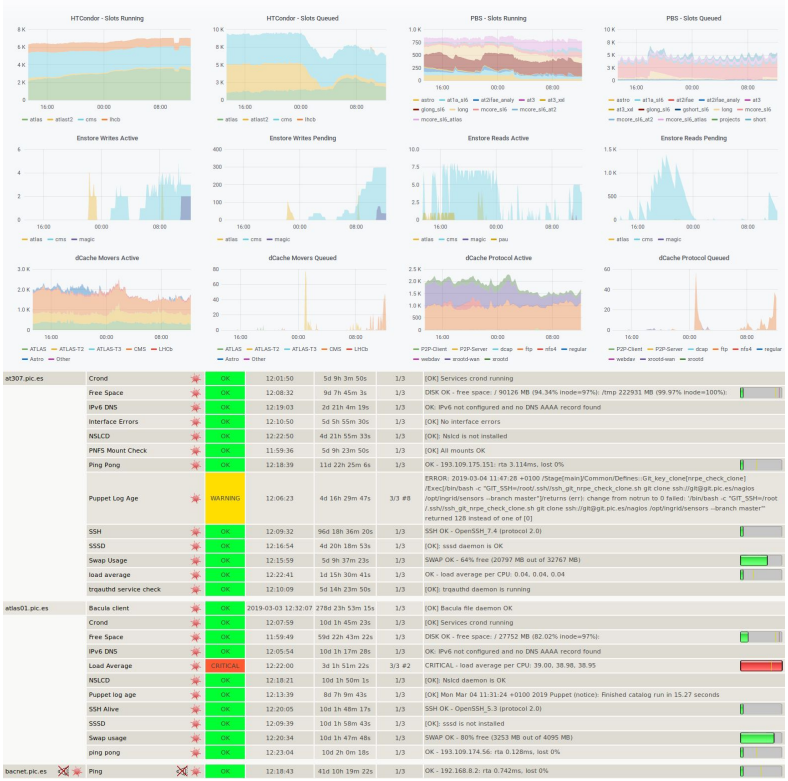
www.pic.es

Questions

- How many users prefer the query builder to the expert SQL mode?
  - We don't know. We did not take the time to measure it.
- How do you see what the users are doing and get their feedback?
  - We can see the queries done
  - Batch historic
    - Interactives are a bit more difficult to get (Tez)
  - Many users directly know us and they send emails
  - Contact form (a bit hidden)
- Are sub-queries allowed?
  - Yes
- What happens if a uset tries a very-long running query, or one generates huge output?
  - We don't have a limit. Up to know nothing horrible happened, except very large custom catalogs that they cannot be used. We remove them after one month.
  - We could add a time limit
  - Disk quota for the future
- What is the database backend and why did you choose it?
  - Hive
    - Free software
    - SQL interface (transparent migration from Postgres)
    - Workflow that perfectly fits
    - Storage and processing scale linearly, parallel queries, great performance (low latency "not allowed")
    - Expandable
- Does CosmoHub sill run when the database is being backup?
  - Yes. File system snapshot
- What is the codebase for the online plot generator
  - Custom SQL wrapper to aggregate plot values in Hive
  - Backed transfers values in csv+json via websockets to the browser
  - Plots in frontend use plotly

PIC support services

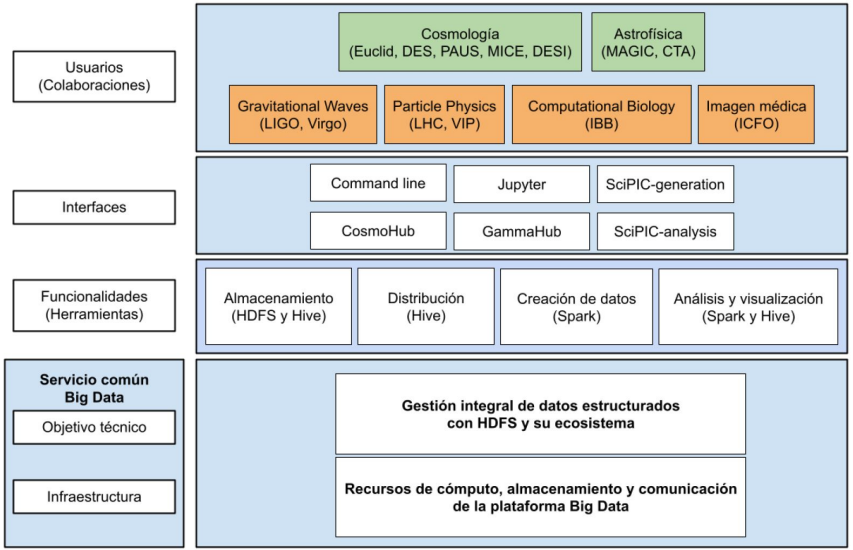
- Administration
  - Puppet
  - LDAP
  - Wiki
  - Git
- Monitoring
  - Nagios + Thruk
  - Elasticsearch + Logstash + Kibana
  - Collectd + Graphite + Grafana
  - Pakiti
  - Cacti
- Manager on Duty (MoD) (24/7)



PIC Big Data common service

- Based on hadoop
  - Open source Big Data Platform
  - Distributed storage and processing
    - Runs on commodity computer clusters
    - Scalable from dozens up to thousands of nodes
      - *Performance scales with HW*
  - Fault tolerant
    - Simple machines working together
      - *no single point of failure*

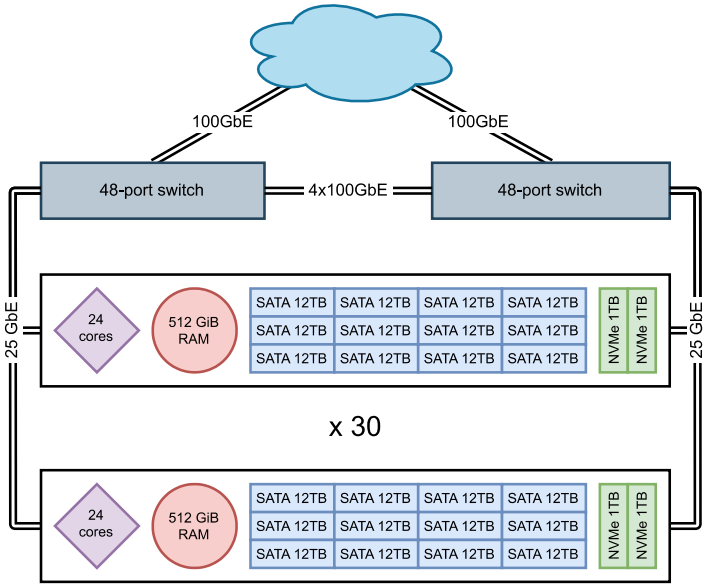
- Last update (summer 2020):
  - Custom DIY nodes
    - 12 nodes AMD Threadripper 1920X
    - 128 GB RAM, 12 x 3 TB SATA HDD hot-swap
    - 2x1 TB NVMe SSD i 2x10-GBASE-T LAN
  - Hortonworks HDP 3.1.4
    - Hadoop 3.1.0
    - Hive 3.1.0
    - Spark 2.3.2



Organización servicio común Big Data. Las disciplinas que hacen uso habitual de la plataforma aparecen en color verde mientras que los potenciales usuarios aparecen en color naranja.

Hadoop platform expansion

- “Convocatoria de ayudas para la adquisición de equipamiento científico-técnico” (EQC2021-007479-P) (2021-2024)
- Near future update (Dec 2023):
  - **30 nodes, 720 cores**
  - **15 TiB RAM, 2.0 PB, 60TB NVMe**
  - Custom Hadoop distribution



PIC
port d'informació
científica

Catalog data volume

Project	Date	volume / night	Number of objects (catalog)
SDSS	**2000 - now**	variable	2×10^6
MICE GC (sims)	**2013**	NA	5×10^8
DES	**2013 - 2018**	2.5 TiB	4×10^8
GAIA*	**2014 - 2019**	40 GiB	1.8×10^9
Euclid Flagship (sims)	**2016 - now**	NA	8.7×10^9
Euclid	**2023 - 2029**	100 GiB	1.5×10^9
LSST	**2024 - 2034**	15 TiB	1×10^{10}

* DR3 full sky star catalog

PIC
port d'informació
científica

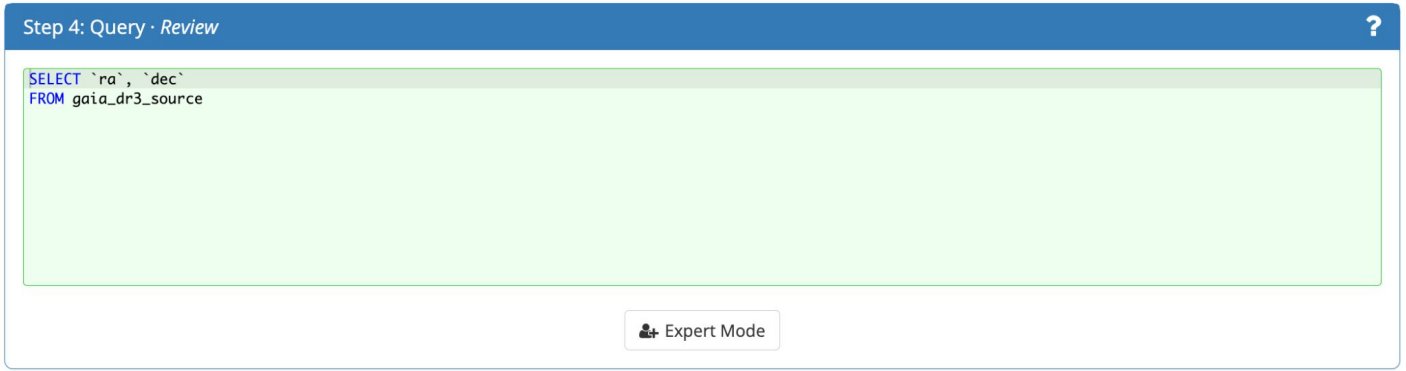
Guided process

- Catalog description
- Value Added Data · Directly download useful or necessary files to analyse the catalog
- Step 1: Columns · Select the fields you need
- Step 2: Sampling · Select a subset and get faster results
- Step 3: Filters · Add conditions to refine your search
- Step 3a: Cone Filter · Restrict results around a sky position
- Step 4: Query · Review

PIC
port d'informació
científica

Guided process

- Catalog description
- Value Added Data · Directly download useful or necessary files to analyse the catalog
- Step 1: Columns · Select the fields you need
- Step 2: Sampling · Select a subset and get faster results
- Step 3: Filters · Add conditions to refine your search
- Step 3a: Cone Filter · Restrict results around a sky position
- Step 4: Query · Review



PIC
port d'informació
científica

Guided process

- Catalog description
- Value Added Data · Directly download useful or necessary files to analyse the catalog
- Step 1: Columns · Select the fields you need
- Step 2: Sampling · Select a subset and get faster results
- Step 3: Filters · Add conditions to refine your search
- Step 3a: Cone Filter · Restrict results around a sky position
- Step 4: Query · Review
- Step 5: Analysis · Explore the selected data (Table, Scatter, Histogram, **Heatmap**)

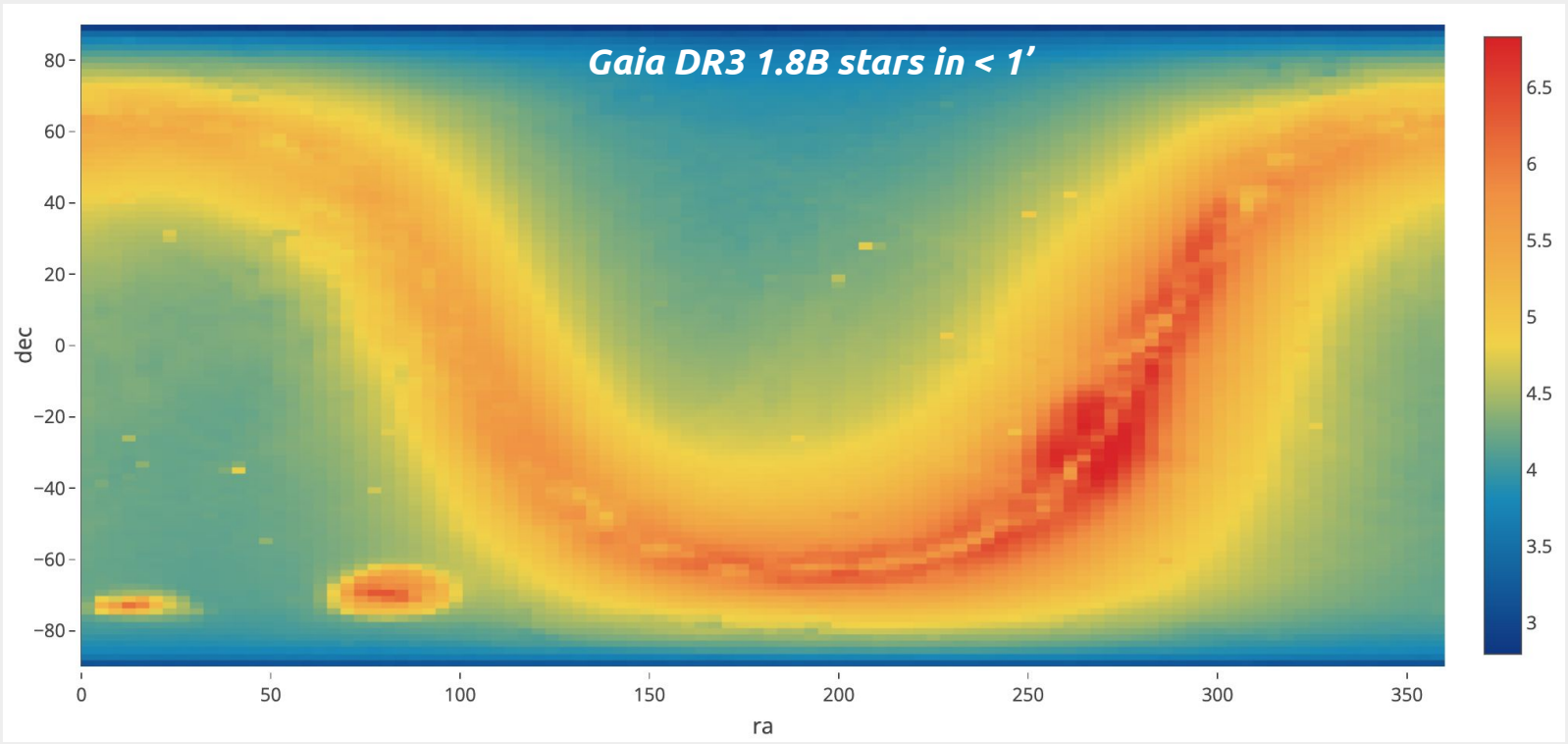
PIC
port d'informació
científica

Guided process

- Catalog description
- Value Added Data · Directly download useful or necessary files to analyse the catalog
- Step 1: Columns · Select the fields you need
- Step 2: Sampling · Select a subset and get faster results
- Step 3: Filters · Add conditions to refine your search
- Step 3a: Cone Filter · Restrict results around a sky position
- Step 4: Query · Review
- Step 5: Analysis · Explore the selected data (Table, Scatter, Histogram, Heatmap)
- Step 6: Format · Select a file type (csv.bz2, FITS, parquet, ASDF)
- Step 7: Request · Review citation guides

PIC
port d'informació
científica

Guided process



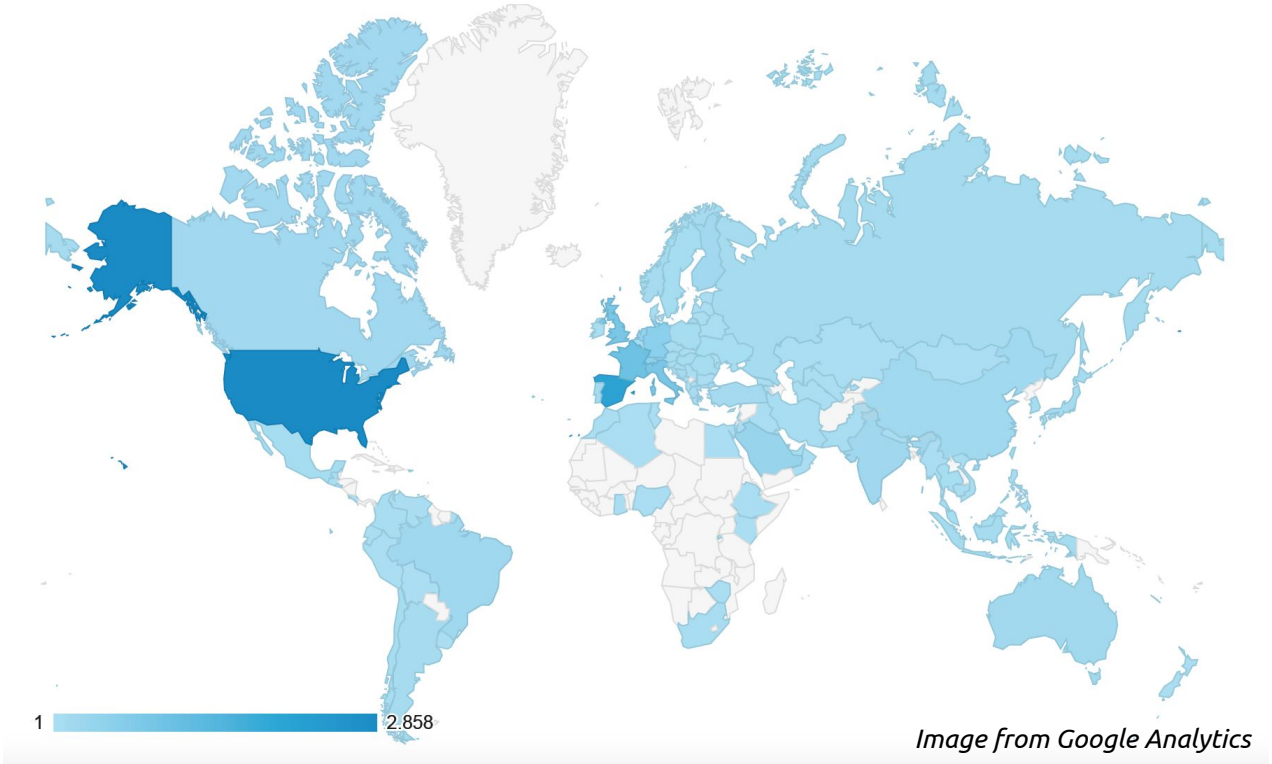
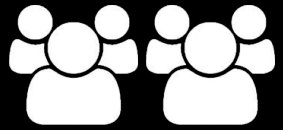


Image from Google Analytics

Ubicación física de los usuarios de CosmoHub desde su creación

COSMO HUB


~ 150 active users


~ 12K custom catalogs


~ 40 TiB hosted data


> 100 catalogs

Public catalogs

- Gaia (DR3, Mean Spectrum, EDR3, DR2 & DR1)
- DESI Legacy Survey DR8 PZ
- DESI Legacy Survey with Photoz (DR8)
- COSMOS 2020 (Classic | Farmer)
- COSMOS 2015 Laigle (v2.1)
- LSST DESC DC2 (Truth-match | Object table)
- DES DR2
- DES Y1A1 Morphological catalog (v1.0)
- DES Y1A1 Gold Data (v1.0)
- GLADE (v2.3, v2.4) & GLADE +
- VIPERS photometry and spectroscopy (PDR2)
- KiDS (DR4)
- CANDELS Bulge-Disk decomposition (2018)
- CFHTLenS (good fields) (v1.2)
- Alhambra photometric redshifts (v1.0)
- ALHAMBRA S/G CLASSIFIED (v1.0)
- PAUS+COSMOS photo-z catalog (v0.4)
- PAUS-COSMOS Early Data Release (v1.0)
- PAU.MillGas Lightcone (2016-07-18)
- DEEP2 Redshift catalog (DR4)
- MICE halo properties
- MICECAT (v2.0, v1.0)

PIC
port d'informació
científica

Hive tuning

- We have set the platform so that queries over large tables are really fast:
  - Apache Tez execution engine instead of the venerable Map-reduce engine
  - ORCfile: a new table (column based) storage format
  - Vectorized query technique: batches of 1024 rows at once

PIC
port d'informació
científica

Load balancing

- Set up two different queues given the two different profiles:
- ‘Interactive’: real-time analysis (low latency)
- ‘Batch’: custom catalogs (high latency)
- Configure queue shares and preemption:
- batch jobs take idle resources to maximize efficiency (10-90)
- interactive jobs can take resources from batch queue (90-100)

PIC
port d'informació
científica

Backend

- ReST API powered by Flask:
- flask-restful - ReST framework
- sqlalchemy - database ORM
- websockets - bidirectional communications
- gevent - asynchronous framework
- pyhive - hive connection library
- pyhdfs - hdfs bindings

PIC
port d'informació
científica

Frontend

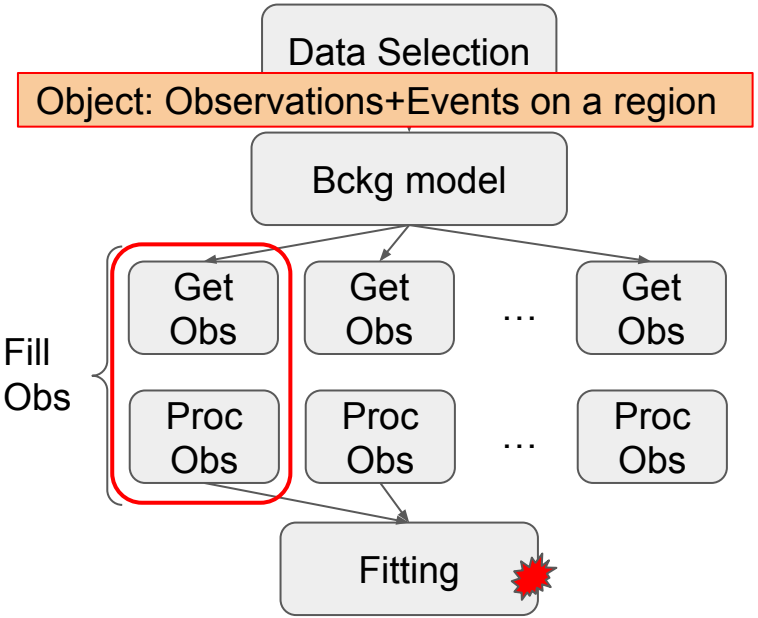
- Responsive Web interface powered by:
- Angular JS - web app oriented HTML framework
- Bootstrap - responsive frontend framework
- Plot.ly for plotting
- Wordpress as backend to edit "static" content

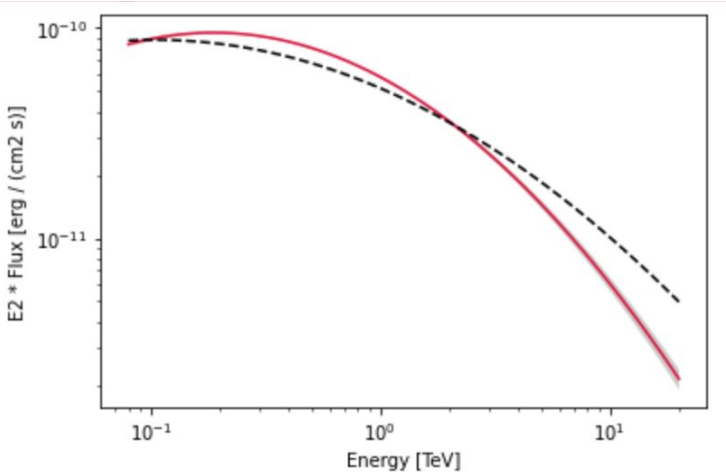
CosmoHub Science Portal

- Plan Complementario: Astrofísica y Física de Altas Energías (2022-2025)
- Línea 8: Computación, big data e inteligencia artificial:
  - "Dar un salto cualitativo en la participación española en la siguiente generación de proyectos internacionales líderes en el área de la Astrofísica y Física de Altas Energías, con un énfasis particular en sus aspectos más tecnológicos" (MICINN)
  - "Aprovechar al máximo las capacidades actuales de las infraestructuras existentes de análisis de big data, expandiendo su capacidad y su ambición, hasta tener en España un hub de datos de astronomía multi-mensajero único en Europa" (BOE 2023)

GammaHub Data Products: introducing parallelism with Spark

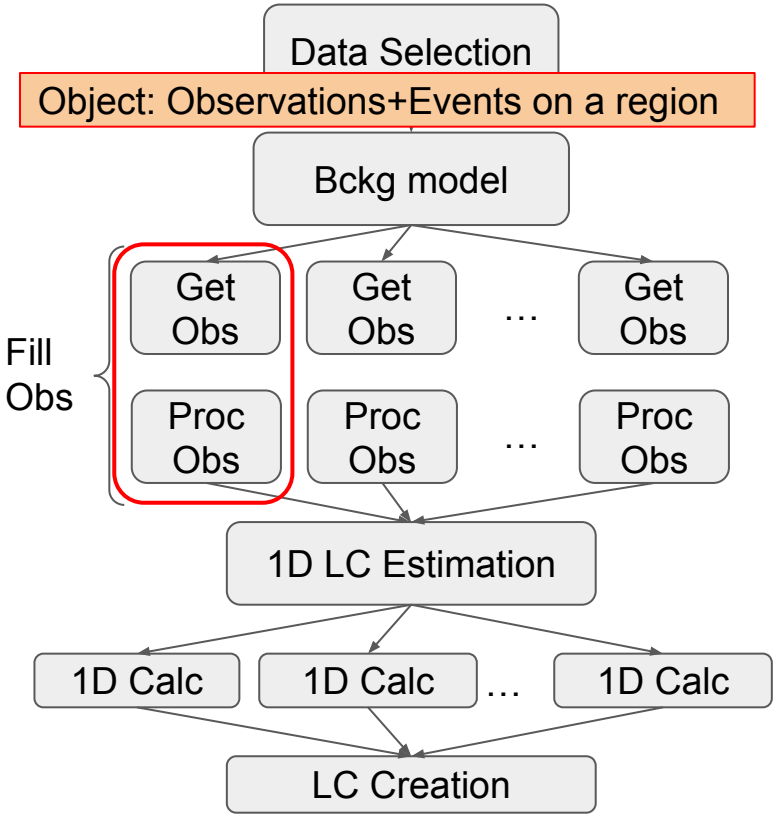
Gammapy low level SED workflow + parallelism (v.1)

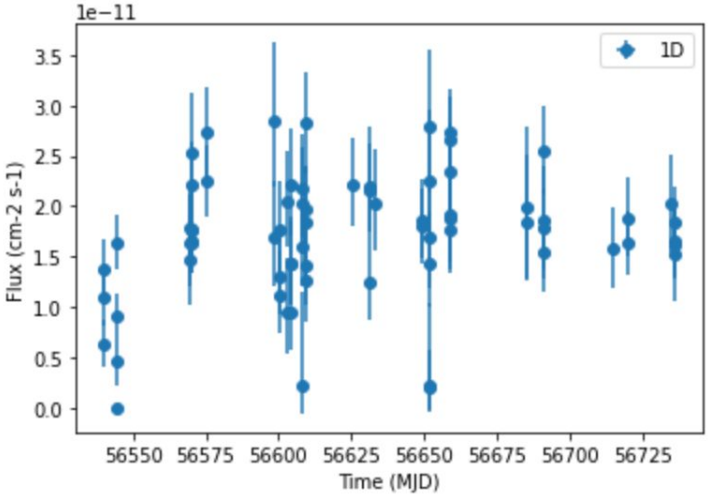




GammaHub Data Products: introducing parallelism with Spark

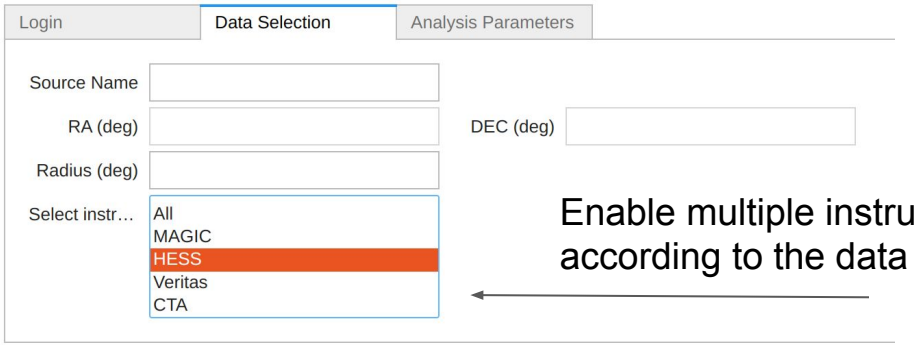
Gammapy low level LC workflow + parallelism (v.1)



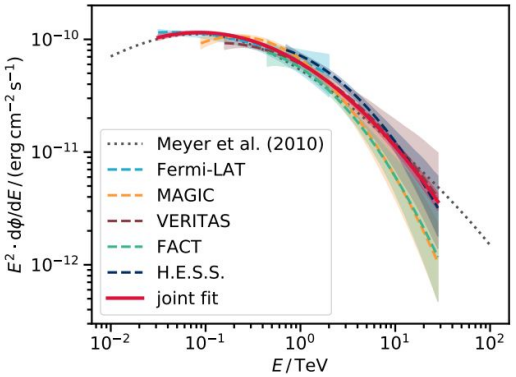


MultiInstrument GammaHub Data Products

Objective: allow to perform multi-instrument data selection and analysis



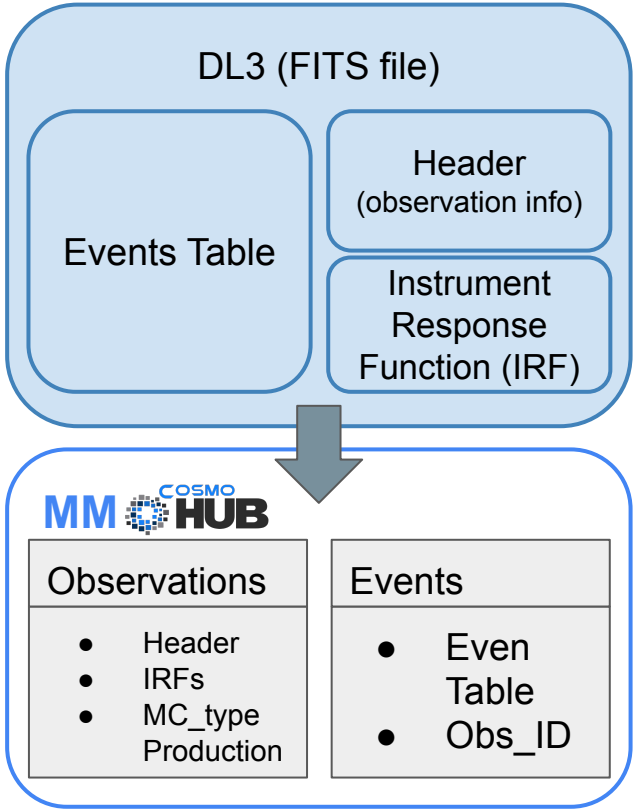
Enable multiple instruments selection according to the data stored in Hive



Example: joint crab paper. Scale the calculation using parallel calculations to obtain each curve and then combine all together on a single plot

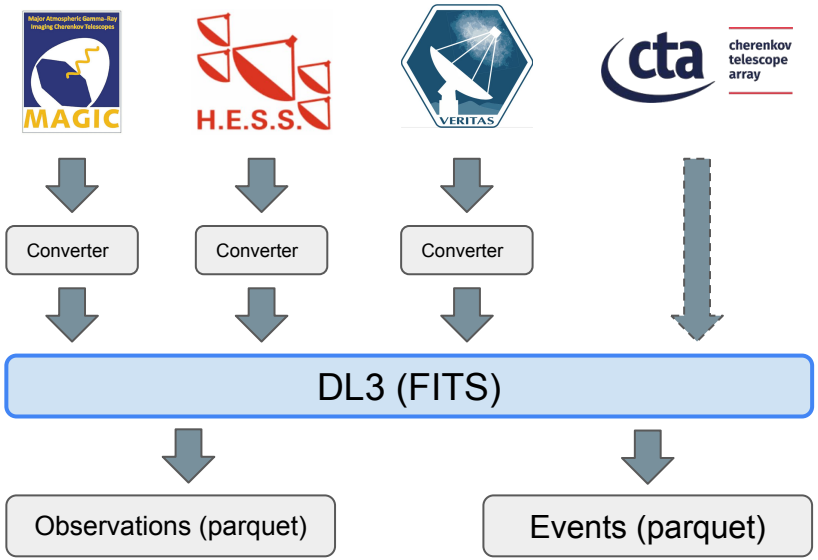
Science Portal: service improvements

- Multi-instrument, -frequency, -messenger
  - GW, optical, Gammas, Neutrinos, etc.
- Guided data products:
  - **Spatial / time cross-match**
  - ad-Hoc Value Added Data
- Integrate with **Jupyter notebooks**
- Additional UDFs functions:
  - geometric, array, footprint, MOCs



Science Portal: service improvements

- Multi-instrument, -frequency, -messenger
  - GW, optical, Gammas, Neutrinos, etc.
- Guided data products:
  - **Spatial / time cross-match**
  - ad-Hoc Value Added Data
- Integrate with **Jupyter notebooks**
- Additional UDFs functions:
  - geometric, array, footprint, MOCs



parquet file format: columnar data format + useful metadata for data ingestion