COSMOHUB

A web portal to interactively analyze and distribute massive cosmological data

https://cosmohub.pic.es

J.Carretero, P.Tallada, J.Casals, M.Caubet, C.Neissner, N.Tonello, J.Delgado, F.Torradeflot, M.Delfino, S.Serrano, P.Fosalba

PIC
port d'informació científica

ICE CSIC-IEEC

IFAE
25 years 1991-2016
Institut de Física d'Altes Energies

EXCELENCIA SEVERO OCHOA

BIST
Barcelona Institute of Science and Technology

MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

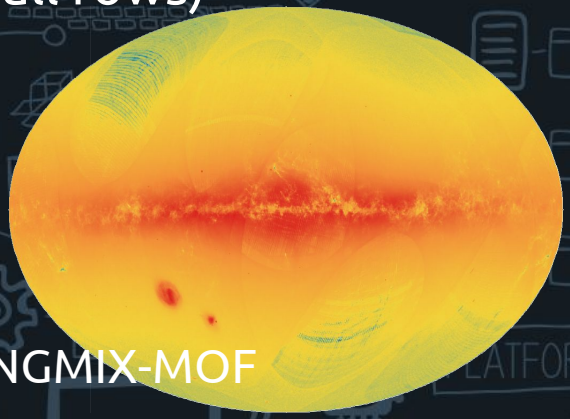
Ciemat
Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

EXCELENCIA MARÍA DE MAEZTU

PIC Big Data platform

- Based on Hadoop (Hortonworks HDP 2.5)
  - Open source
  - Distributed storage and processing
    - Runs on commodity computer clusters
    - Scalable from dozens up to thousands of nodes
      ✓ Performance scales with HW
    - Failure tolerance
      ✓ May use old/refurbished machines → cheaper

- PIC Hadoop platform
  - 30 nodes: 360 cores, 1080 GiB RAM, 120 TiB HDD

COSMO HUB

- Main features:
  - Generate and download custom catalogs
    - Guided process, no SQL knowledge needed
    - Expert mode available
  - Full dataset Interactive visualization (over all rows)
    - May use sampling
    - 1D histogram & 2D heatmap
  - Value-added-data ready to download
- Datasets:
  - DES:
    - SVA1, Y1A1 gold v1.0.3 + 3BPZ catalogs + NGMIX-MOF
    - Waiting for Y3 :p
  - Euclid, Gaia, MICE, PAU, SDSS, COSMOS, USNO, ...
- CosmoHub in numbers:
  - ULTRAFAST: 85% catalogs finish in < 3 min
  - ~ 400 users, > 1500 custom catalogs, > 5000 plots